KAM theory without action-angle variables

R. de la Llave [*] A. González[*] À. Jorba[†] J. Villanueva [‡]

15th September 2003

Abstract

Given a torus approximately invariant under a symplectic map, we find sufficient conditions for the existence of a true invariant torus close to the given one. We do not assume either that we are working on action-angle variables nor that the map is close to integrable. The proposed method is also useful to compute numerically invariant tori for symplectic maps. Analogous results are also true for vector fields.

Contents

---

[*] Dept. Mathematics, University of Texas at Austin

[†] Dept. de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona

[‡] Dept. de Matemàtica Aplicada I, Universitat Politècnica de Catalunya

1 Introduction

The goal of this paper is to present a proof of a KAM theorem on existence of invariant tori with a Diophantine rotation vector.

The results for an exact symplectic map f of a $2n$-dimensional manifold $\mathbf{U}$ are based on the study of the equation

$$(f \circ K)(\theta) = K(\theta + \omega) \tag{1}$$

where $K : \mathbb{T}^n = (\mathbb{R}/\mathbb{Z})^n \to \mathbf{U}$ is the function to be determined and $\omega \in \mathbb{R}^n$ satisfies a Diophantine condition.

We will assume that $\mathbf{U}$ is either $\mathbb{T}^n \times U$ with $U \subset \mathbb{R}^n$ or $B \subset \mathbb{R}^{2n}$, so that we can use a system of coordinates. In the case that $\mathbf{U} = \mathbb{T}^n \times U$, we note that the embedding K could be non-trivial.

Notice that (1) implies that the range of K is invariant under f. The map K gives a parameterization of the invariant torus which makes the dynamics of f restricted to the torus into a rigid rotation.

What we show is that, given a K which solves approximately equation (1) and which is not too degenerate (see Definition 2), then, there is a true solution nearby. We will refer to such proofs as 'polishing'.

Remark 1. *It is a general fact that the results on invariant tori for maps imply the results for vector fields if we have local uniqueness.*

If we denote by S_t the flow, if $S_1 \circ K(\theta) = K(\theta + \omega)$, then, for all t

$$S_1 \circ (S_t \circ K)(\theta) = S_t \circ S_1 \circ K(\theta) = (S_t \circ K)(\theta + \omega)$$

By the local uniqueness, we conclude that, for small t, we have $S_t \circ K(\theta) = K(\theta + \varphi_t)$, where φ_t is a differentiable function. From $S_{t+s} = S_t \circ S_s$, we conclude that $\varphi_{t+s} = \varphi_t + \varphi_s$ and, since φ is differentiable, that $\varphi_t = \eta t$ for some vector η. Since $\varphi_1 = \omega$, we conclude that $\eta = \omega$. Hence, we obtain from the equation of invariance for the time one map and the uniqueness that

$$S_t \circ K(\theta) = K(\theta + t\omega)$$

for all t.

A different argument to show that invariant tori theorems for maps imply results for flows (and viceversa) can be found in [Dou82]. Note that the argument presented above applies to other types of invariant objects.

Nevertheless, in Section Section 8, we will present a direct proof for flows along the same lines as the proof for maps. This has a certain interest since the direct proofs presented here can be converted in algorithms. Moreover, we also formulate the results for vector fields precisely in Section 2.3.

The customary statements of KAM for nearly integrable systems follows since one can take the tori for the integrable system as approximate tori for our system.

We emphasize, however, that the result we prove (see Theorem 1) does not require that the map f is close to integrable.

Presenting the KAM theory in a polishing way -as it is done in [Mos66b, Mos66a, Rüs76a, Zeh75, Zeh76, SZ89, Ran87, CC97]- has certain advantages, we mention just a few, which we hope to develop in future papers.

- In many practical applications, one has to consider systems that are not close to integrable but nevertheless one has – e.g. by applying a non-rigorous numerical method, some asymptotic expansion etc. – some approximate solutions of (1). Then, the result Theorem 1 implies that close to these approximate solutions there is a true solution. That is, Theorem 1 can validate the approximate solutions. In numerical analysis, theorems of this type are often called *"A posteriori estimates"*.
- From a more theoretical point of view, it is possible to take as approximate solutions smoothing of other solutions or other approximations and a theorem of the type indicated above, more or less automatically leads to results for finitely differentiable maps.
- By taking as approximate solutions the results of formal expansions, one obtains, differentiability with respect to parameters or with respect to the frequency. (One obtains that the family of tori are Whitney differentiable). (A general discussion of this can be found in [Van02, dlLV00].)

  By a similar method, one can also obtain results on the abundance of tori near a torus (sometimes this is called *"condensation"*).
- The smooth dependence on parameters leads to several other results such as estimates on the measure occupied by the torus and the persistence of tori in near-integrable systems with weak conditions of non-degeneracy.

A novelty of our approach compared with other approaches to KAM theory is that the method does not require at all action-angle variables. This seems to be useful in several situations. We just mention:

- In several occasions the action-angle variables are much more singular than the original system (a recent remarkable example of very smooth integrable system with complicated action-angle variables is [RC95, RC97]).
- In practical – numerical – applications, it is often not easy to work with action-angle variables. Hence, from the point of view of using the result as a validation tool for numerics the requirement of using action-angle variables is unduly restrictive.
- There are also some theoretical applications where the action-angle variables are much more singular than the map. Two important examples of this situation are the study of invariant tori near an elliptic fixed point or near a

separatrix. (Studies using action-angle variables for invariant tori near an elliptic fixed point can be found in [DG96, Rüs98, Sev96] and for tori near separatrices in [Neĭ84, Her83, DdlLS03]. The paper [DdlLS03] also shows that these tori near separatrices allow to overcome the large gap problem in diffusion and play a prominent role in a mechanism for topological instability).

Reducibility of invariant curves

The method we present is based on a very simple geometric observation which has further applications. We use the geometry of the map f to show that an invariant torus has the remarkable property that the equations of infinitesimal perturbations can be transformed into upper triangular linear difference equations, with diagonal elements equal to one. A perturbative argument will show that for an approximately invariant torus, the linearized equations can be approximately reduced to upper triangular linear difference equations, with diagonal elements equal to one. This is enough for a quasi-Newton method – see [Mos66b, Mos66a, Zeh75, Zeh76].

The geometry of this phenomenon, can be understood pictorially in dimension 2. We hope that this will give the flavor of the ideas.

Note that if we take derivatives of (1), we obtain that the vector field $u(\theta) = \partial_\theta K(\theta)$ based at the point $K(\theta)$ satisfies

$$Df(K(\theta))u(\theta) = u(\theta + \omega) \tag{2}$$

Since $K(\theta + \omega) = f(K(\theta)$, we see that (2) implies that the vector field $u(\theta)$ is invariant under the map.

We now consider the vector field $v(\theta)$ based also at $K(\theta)$, perpendicular to $u(\theta)$ but normalized so that the area of the parallelogram based on $u(\theta), v(\theta)$ is 1 (see figure bellow).

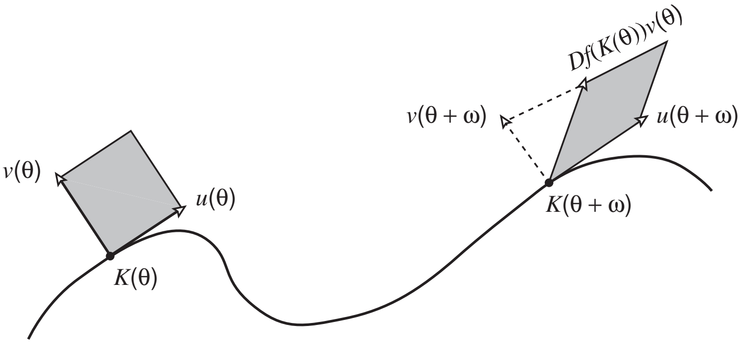

Figure 1: Geometric reason for automatic reducibility.

The fact that the map f preserves area, together with the fact that u is preserved implies that

$$Df(K(\theta))v(\theta) = v(\theta + \omega) + c(\theta)u(\theta) \tag{3}$$

Together equations (2) and (3) imply that if we consider the basis given by $u(\theta), v(\theta)$ – at the point $K(\theta)$ – the derivative $Df(K(\theta))$ becomes a particularly simple matrix

$$M(\theta + \omega)^{-1} Df(K(\theta))M(\theta) = \begin{pmatrix} 1 & c(\theta) \\ 0 & 1 \end{pmatrix} . \tag{4}$$

where M is the matrix of the change of variables to $u(\theta), v(\theta)$.

This means geometrically that the infinitesimal perturbations just get sheared along invariant the torus.

The infinitesimal equations involved in a Newton method to solve (1), are given by

$$Df(K(\theta))\Delta(\theta) - \Delta(\theta + \omega) = -\eta(\theta) . \tag{5}$$

Using the equation (4) they become

$$\begin{pmatrix} 1 & c(\theta) \\ 0 & 1 \end{pmatrix} M(\theta)\Delta(\theta) - (M\Delta)(\theta + \omega) = -M(\theta + \omega)^{-1}\eta(\theta) . \tag{6}$$

The equation (6) can be studied using the usual techniques for difference equations. It is well known that one can solve equations of the form (6), provided that certain averages vanish. As it turns out, the vanishing of the averages can be guaranteed by a geometric condition and a non-degeneracy condition on $c(\theta)$. As a possible alternative, we will also propose the use of the method of the translated curve theorem and solve problems with extra parameters which makes the averages zero.

Of course, there is little point in applying linearized equation to something that is already a solution, but it turns out that if the K is not an exact solution but an approximate solution of (1), it is possible to construct a matrix which satisfies approximately (4). This is enough to get a quasi-Newton method.

Note that equation (2) is a feature of all conjugacy problems (It plays a prominent role in the discussions of [Mos66b]). On the other hand, (3) uses the geometric properties of the map f.

The automatic *'approximate reducibility'* – an analogue of equation (3) – is guaranteed by quite a number of geometric properties. For example, in this paper we will discuss the situation for symplectic mappings and Hamiltonian vector fields. In such cases the fact that invariant tori are Lagrangian plays an important role.

There are, however, quite a number of other contexts. We just mention that the same phenomenon happens for conformally symplectic mappings in the sense of [WL98] – those are mappings in which the symplectic form gets multiplied by a constant – as well as volume preserving maps.

We point out that automatic reducibility has been considered in the case of Lagrangian systems. For example, it appears in [Koz83] and [Mos88]. The Lagrangian automatic reducibility has the advantage that it applies also to systems with more independent variables. On the other hand, there seems to be no straightforward variational context for several situations – such as conformally symplectic or volume preserving.

Besides Lagrangian context, reducibility has been also used in the Hamiltonian context in [Bos86], where the author uses reducibility to prove the theorem of persistence of invariant tori in Hamiltonian systems close to integrable. In [Rüs76b] the author presents a geometric method, using reducibility, to prove existence of invariant tori for twist mappings. In [CC97], the authors use the automatic reducibility for tori which are graphs in mechanical systems (action angle coordinates and Hamiltonians of the form kinetic plus potential energy) to obtain the existence of invariant tori in rotations of the three body problem with parameters close to those of the Ceres asteroid.

This paper is organized as follows: In Section 2 we formulate the results of this paper. In Section 3 we explain briefly the general idea of the proofs. Section 4 contains the main step on the construction of the polishing method for symplectic maps, we prove that the linearized equation of (1) can be transformed into an upper triangular linear equation which can be solved approximately. In Section 5 we use the results from Section 4 to construct a quadratic convergent Newton-like method for solving equation (1). In Section 6 we prove that solutions of equation (1) are locally unique. In Section 7 we will see how with the use of parameters, the non-degenerated condition N2 (see Definition 2) can be avoided. More concretely, we will describe a polishing method to prove the existence of an invariant torus for an element of a given parametric family of symplectic maps f_λ. Finally, in Section 8 we sketch the the procedure form Hamiltonian vector fields.

2 Statement of results

2.1 Notation and preliminaries

Before we formulate the results of this paper, we introduce some notation. Let $\Omega = d\alpha$ be an exact symplectic structure on $\mathbf{U}$, and let $a : \mathbf{U} \to \mathbb{R}^{2n}$ be defined by

$$\alpha_z = a(z)dz \qquad \forall z \in \mathbf{U}. \tag{7}$$

For each $z \in \mathbf{U}$ let $J(z) : T_z\,\mathbf{U} \to T_z\,\mathbf{U}$ be linear isomorphism satisfying

$$\Omega_z(\xi, \eta) = <\xi, J(z)\eta>\,, \tag{8}$$

where $< \cdot, \cdot >$ is the Euclidean scalar product on $\mathbb{R}^{2n}$. It is well known that J satisfies $J(z)^\top = -J(z)$.

Remark 2. *Although it is possible to assume (by changing the metric on* $\mathbf{U}$*, c.f [Wei77]), that* J *is an almost complex structure i.e.* $J^2 = -Id$*, we will not make this assumption in this paper. Indeed, we have strived to deal with the problem in whatever coordinates it is presented.*

Definition 1. *Given* $\gamma > 0$ *and* $\sigma \geq n$*, we define* $D(\gamma, \sigma)$ *as the set of frequency vectors* $\omega \in \mathbb{R}^n$ *satisfying the Diophantine condition:*

$$|\ell \cdot \omega - m| \geq \gamma \, |\ell|_1^{-\sigma} \qquad \forall \ell \in \mathbb{Z}^n \setminus \{0\}, m \in \mathbb{Z},$$

where $|\ell|_1 = |\ell_1| + \cdots + |\ell_n|$*.*

Similarly, given $\gamma > 0$ *and* $\sigma > n - 1$*. we define* $D_h(\gamma, \sigma)$ *as the set of frequency vectors* $\omega \in \mathbb{R}^n$ *satisfying*

$$|k \cdot \omega| \geq \frac{\gamma}{|k|_1^\sigma}, \qquad \forall \quad k \in \mathbb{Z}^n - \{0\}, \tag{9}$$

Let U_ρ denote the complex strip of width $\rho > 0$: $U_\rho = \{\theta \in \mathbb{C}^n : |\mathrm{Im}\ \theta| \leq \rho\}$. $\left(\mathcal{P}_\rho, \| \|_\rho\right)$ will denote the Banach space set of functions $K : U_\rho \to \mathbf{U}$ which are 1-periodic in all its variables, real analytic on the interior of U_ρ and continuous on the boundary of U_ρ, with the norm

$$\|v\|_\rho \overset{\text{def}}{=} \sup_{\theta \in U_\rho} |v(\theta)| \, ,$$

where $|\cdot|$ represents the maximum norm on the spaces $\mathbb{R}^m$ and $\mathbb{C}^m$, i.e. if $x = (x_1, \ldots, x_m) \in \mathbb{C}^m$, then

$$|x| \overset{\text{def}}{=} \max_{j=1,\ldots,m} |x_j| \, .$$

Similar notation for the norm will be also used for real or complex matrices of arbitrary dimension, and it will refer to the matrix norm induced by the vectorial one.

Similarly, we will consider the set $\tilde{\mathcal{P}}_\rho$ of functions $K : U_\rho \to \mathbf{U}$ real analytic on the interior of U_ρ and continuous on the boundary of U_ρ, and satisfying

$$K(\theta + k) = K(\theta) + (k, 0) \qquad k \in \mathbb{Z}^n. \tag{10}$$

Notice that it is equivalent to say that K satisfies (10) than to say that $K(\theta) - (\theta, 0)$ is periodic. Hence, we can consider $\tilde{\mathcal{P}}_\rho$ as an affine space modelled on $\mathcal{P}_\rho$.

Given a function g, analytic on a complex set $\mathcal{B}$. For $m \in \mathbb{Z}_+ \overset{\text{def}}{=} \mathbb{N} \cup \{0\}$ denote the C^m-norm of g on $\mathcal{B}$ by $|g|_{C^m, \mathcal{B}}$, i.e.

$$|g|_{C^m, \mathcal{B}} \overset{\text{def}}{=} \sup_{0 \leq |k| \leq m} \sup_{z \in \mathcal{B}} \left| D^k g(z) \right| \, .$$

Given a function $g : B \subset \mathbb{R}^d \to \mathbb{R}^m$, $Dg(z)$ will denote the $m \times d$ matrix with i, j coordinate $\frac{\partial g_i}{\partial x_j}$. For $h : \mathbf{U} \to \mathbb{R}$, ∇h will represent the gradient of h.

We denote the Fourier expansion of a periodic mapping $K : \mathbb{T}^n \to \mathbf{U}$ by

$$K(\theta) = \sum_{\ell \in \mathbb{Z}^n} K^{(\ell)} \exp(2\pi i\, \ell \cdot \theta),$$

where $\cdot$ is the Euclidean scalar product in $\mathbb{R}^n$, and the Fourier coefficients $K^{(\ell)}$ can be computed

$$K^{(\ell)} \stackrel{\text{def}}{=} \int_{\mathbb{T}^n} K(\theta) \exp(-2\pi i\, \ell \cdot \theta)\, d\theta\,.$$

The average of K is the 0-Fourier coefficient, we denote the average of K on $\mathbb{T}^n$ by

$$\operatorname{avg}\{K\}_\theta \stackrel{\text{def}}{=} \int_{\mathbb{T}^n} K(\theta)\, d\theta = K^{(0)}\,.$$

Definition 2. *Given a symplectic map f and $\omega \in D(\gamma,\sigma)$. A mapping $K \in \mathcal{P}_\rho$ is said to be non-degenerate if it satisfies the following conditions*

N1 *There exists an $n \times n$ matrix-valued function $N(\theta)$, such that*

$$N(\theta)\ \left(DK(\theta)^\top DK(\theta)\right) = I_n,$$

where I_n is the n-dimensional identity matrix.

N2 *The average of the matrix-valued function*

$$S(\theta) \stackrel{\text{def}}{=} P(\theta+\omega)^\top \left[Df(K(\theta))J(K(\theta))^{-1}P(\theta)\ - J(K(\theta+\omega))^{-1}P(\theta+\omega) \right]$$

with

$$P(\theta) \stackrel{\text{def}}{=} DK(\theta)\, N(\theta)\,, \tag{11}$$

is non-singular.

We will denote the set of functions in $\mathcal{P}_\rho$ satisfying conditions **N1-N2** *by $\mathcal{ND}(\rho)$.*

By the Rank Theorem, Condition **N1** guarantees that $\dim K(\mathbb{T}^n) = n$. For the KAM theorem, the main non-degeneracy condition is **N2**, which is a twist condition. Its role will become clear in Section 4.4. Note that **N1** depends only on K whereas **N2** depends on K and f.

We will say that f_λ is a d-parametric family of symplectic maps if there is a function $f : \mathbf{U} \times B \to \mathbf{U}$, with $B \subset \mathbb{R}^d$, such that for each $x \in \mathbf{U}$ the map $f(x,\cdot)$ is C^2, and such that for each $\lambda \in B$, the map $f_\lambda \stackrel{\text{def}}{=} f(\cdot,\lambda) : \mathbf{U} \to \mathbf{U}$ is symplectic and real analytic.

Definition 3. *Given a $2n$-parametric family of symplectic maps f_λ and $\omega \in D(\gamma,\sigma)$. The pair $f_{\hat\lambda}$, $K \in \tilde{\mathcal{P}}_\rho$ is said to be non-degenerate if it satisfies the following conditions*

T1 *There exists an* $n \times n$ *matrix-valued function* $N(\theta)$*, such that*

$$N(\theta)\ \left(DK(\theta)^\top DK(\theta)\right) = I_n,$$

T2 *Let* P *be as in (11) and*

$$T(\theta) = P(\theta)^\top \left[I_n - J(K(\theta))^{-1} P(\theta) DK(\theta)^\top J(K(\theta)) \right] ,$$

the $2n \times 2n$ *matrix*

$$\Lambda(\theta) = \begin{pmatrix} T(\theta+\omega) \\ DK(\theta+\omega)^\top J(K(\theta+\omega)) \end{pmatrix} \left(\left.\frac{\partial f_\lambda}{\partial \lambda}\right|_{\lambda=\hat\lambda} (K(\theta)) \right) ,$$

is such that $\operatorname{avg}\{\Lambda\}_\theta$ *has range of dimension* $2n$*.*

2.1.1 Small divisors equations

Here we enunciate the known results on the solvability of small divisors equations. For a proof see [Rüs76b, Rüs75, Rüs76c] (See also [dlL01]).

Proposition 1. *Let* $\omega \in D(\gamma, \sigma)$*, assume that the mapping* $h : \mathbb{T}^n \to \mathbb{R}^{2n}$ *is analytic on* U_ρ*, and has zero average,* $\operatorname{avg}\{h\}_\theta = 0$*. Then for any* $0 < \delta < \rho$*, the difference equation*

$$v(\theta) - v(\theta + \omega) = h(\theta) \tag{12}$$

has a unique zero average solution $v : \mathbb{T}^n \to \mathbb{R}^{2n}$ *real analytic on* $U_{\rho-\delta}$*, for any* $0 < \delta < \rho$*. Moreover the following estimate holds*

$$\|v\|_{\rho-\delta} \le c_0\, \gamma^{-1}\, \delta^{-\sigma}\, \|h\|_\rho ,$$

where c_0 *is a constant depending on* n *and* σ*.*

Proposition 2. *Assume that* ω *satisfies the Diophantine condition*

$$|k \cdot \omega| \ge \frac{\gamma}{|k|_1^\sigma}, \qquad \forall \quad k \in \mathbb{Z}^n - \{0\},$$

for some constants $\gamma > 0$ *and* $\sigma > n - 1$*.*

Let h *be a real analytic function with zero average. There exists a unique analytic solution* v *with zero average, of the linear equation*

$$\sum_{j=1}^n \omega_j \frac{\partial v}{\partial \theta_i} = h.$$

Moreover if $h \in \mathcal{P}_\rho$*, the solution* v *satisfies following estimate*

$$\|v\|_{\rho-\delta} \le c\, \delta^{-\sigma}\, \gamma^{-1} \|v\|_\rho ,$$

where c *is a constant depending on* n *and* σ*.*

2.2 Maps

Existence of an invariant torus

Now we are ready to state the sufficient conditions to guarantee the existence of a true invariant torus with Diophantine frequency vector ω.

Theorem 1. *Let* $f : \mathbf{U} \to \mathbf{U}$ *be an exact symplectic map symplectic map, and* $\omega \in D(\gamma, \sigma)$, *for some* $\gamma > 0$ *and* $\sigma > n$. *Assume that the following hypotheses hold*

H1. $K_0 \in \mathcal{ND}(\rho_0)$ *(i.e.* K_0 *satisfies Definition 2).*

H2. The map f *is real analytic and it can be holomorphically extended to some complex neighborhood of the image under* K_0 *of* U_{ρ_0}*:*

$$\mathcal{B}_r = \{z \in \mathbb{C}^{2n} : \sup_{|\mathrm{Im}\ \theta| < \rho_0} |z - K_0(\theta)| < r\},$$

for some $r > 0$, *satisfying* $|f|_{C^2, \mathcal{B}_r} < \infty$.

H3. If a *and* J *are as in (7) and (8), respectively,*

$$|a|_{C^2, \mathcal{B}_r} < \infty\,, \qquad |J|_{C^1, \mathcal{B}_r} < \infty\,, \qquad \left|J^{-1}\right|_{C^1, \mathcal{B}_r} < \infty\,.$$

Define the error function e_0 *by*

$$e_0 \overset{\mathrm{def}}{=} f \circ K_0 - K_0 \circ T_\omega. \tag{13}$$

There exists a constant $c > 0$ *depending on* σ, n, r, ρ_0, $|f|_{C^2, \mathcal{B}_r}$, $|a|_{C^2, \mathcal{B}_r}$, $|J|_{C^1, \mathcal{B}_r}$, $\left|J^{-1}\right|_{C^1, \mathcal{B}_r}$, $\|DK_0\|_{\rho_0}$, $\|N_0\|_{\rho_0}$, $\left|(\mathrm{avg}\,\{S_0\}_\theta)^{-1}\right|$ *(where* N_0 *and* S_0 *are as in Definition 2, replacing* K *by* K_0*) such that, if* $\|e_0\|_{\rho_0}$, *defined in (13), verifies the following inequalities*

$$c\,\gamma^{-4}\,\delta_0^{-4\sigma}\,\|e_0\|_{\rho_0} < 1, \tag{14}$$

and

$$c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} < r\,,$$

where $0 < \delta_0 \le \min(1, \rho_0/12)$ *is fixed, then there exists* $K_\infty \in \mathcal{ND}(\rho_0 - 6\delta_0)$ *such that*

$$f \circ K_\infty - K_\infty \circ T_\omega = 0\,.$$

Moreover,

$$\|K_\infty - K_0\|_{\rho_\infty} \le c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0}\,.$$

Remark 3. *We emphasize that Theorem 1 does not assume neither that the system is given in action-angle variables nor that the maps is close to integrable.*

The only smallness conditions needed are on the size of e_0.

Remark 4. *The geometric method we present here enables us to construct a quadratic convergent sequence of approximate solutions of (1) (see Lemma 13).*

The quadratic estimates (see inequality (66)) contain the term γ^{-4}*, this is the reason for the coefficient* γ^{-4} *in (14) which is the same as what is obtained in some classical papers on KAM theory (e.g. [Mos66b] [Mos66a], [Zeh76]), but it is worse than the* γ^{-2} *obtained in some classical KAM theorems. For example in [Pös01], [Sal86]. This estimation results in worse estimation of the measure of the invariant tori than what is possible.*

Similarly, the exponent of δ_0 *in the conditions is similar to that of [Mos66b], [Mos66a], [Zeh76] since it is proportional to* 4σ *but it is different that that of [Pös01], [Sal86] that contains only* $2\sigma + A$*. This leads to different losses in differentiability.*

We hope to come back to these problems in future research.

Remark 5. *Notice that given a mapping* K *satisfying the equation (1), for any* $\varphi \in \mathbb{T}^n$ *the mapping* $K(\theta+\varphi)$ *is also a solution of (1). We will adopt the criterion that two parameterizations,* K *and* $\hat{K}$*, of an embedded torus such that* $K(\theta) = \hat{K}(\theta+\varphi)$*, for some* $\varphi \in \mathbb{T}^n$*, are equivalent since they only differ in the arbitrary choice of the origin of the phases.*

The following Theorem 2 states that the solutions obtained in Theorem 1 are locally unique up to the irrelevant choice of origin of the phase. As we mentioned, in Remark 1 is a general argument that allows to obtain results for flows from results for maps with local uniqueness.

Theorem 2. *Assume that* $\omega \in D(\gamma,\sigma)$*. Let* $K_1, K_2 \in \mathcal{ND}(\rho)$ *(see Definition 2) such that:* $K_1(U_\rho) \subset \mathcal{B}_r$*, and* $K_2(U_\rho) \subset \mathcal{B}_r$*.*

There exists a constant $c > 0$ *depending on* n*,* σ*,* γ*,* ρ*,* ρ^{-1}*,* $|f|_{C^2,\mathcal{B}_r}$*,* $|J|_{\mathcal{B}_r}$*,* $|J^{-1}|_{\mathcal{B}_r}$ $\|K_2\|_{C^2,\rho}$*,* $\|N\|_\rho$*,* $|(\mathrm{avg}\,\{S\}_\theta)^{-1}|$ *(with* N*, and* S *as in Definition 2, replacing* K *by* K_2*), such that if* $\|K_1 - K_2\|_\rho$ *satisfies*

$$c\,\gamma^{-2}\,\delta^{-2\sigma}\,\|K_1 - K_2\|_\rho < 1\,,$$

where $\delta = \rho/8$*, then there exists an initial phase* $\tau \in \mathbb{R}^n$*, such that* $K_1 \circ T_\tau = K_2$ *in* $U_{\rho/2}$*.*

Translated torus theorem

Let us assume that $\mathbf{U} = \mathbb{T}^n \times U$ where $U \subset \mathbb{R}^n$ is an open set, and assume that there is an exact symplectic structure Ω on $\mathbf{U}$. Let $f_\lambda : \mathbf{U} \to \mathbf{U}$ be a $2n$-parametric family of symplectic maps, and let K_0 be an approximate solution of (1), with $f = f_0$. We will assume that K is a non-trivial embedding of the torus in the annulus rather than just a periodic function. That is $K(\theta + k) = K(\theta) + (k, 0)$ for $k \in \mathbb{Z}^d$. Assume that the pair f_0, K_0 satisfy Definition 3, then we prove that there exists a solution of

$$f_\lambda \circ K = K \circ T_\omega\ , \tag{15}$$

where $\omega \in D(\gamma, \sigma)$ is fixed.

In the simplest case – which gives the name to the theorem – the family f_λ consists just in $f_\lambda(x, y) = f_0(x, y) + (0, \lambda)$.

This translated curve theorem is a technical tool in the study of invariant circles when the non-degeneracy assumptions are not met.

The main result of this section is the following:

Theorem 3. *Let $\omega \in D(\gamma, \sigma)$ be a frequency vector. Let $\omega \in D(\gamma, \sigma)$ be a frequency vector and let f_λ be an $2n$-parametric family of symplectic maps. Assume that the following hypotheses hold*

H1'. *f_0 and $K_0 \in \tilde{P}_{\rho_0}$ satisfy conditions* **T1** *and* **T2** *(see Definition 3).*

H2'. *f_λ is a family of real analytic symplectic maps that can be holomorphically extended to some complex neighborhood*

$$\mathcal{B}_r = \{z \in \mathbb{C}^{2n} : \sup_{\theta \in U_{\rho_0}} |z - K_0(\theta)| < r\},$$

for some $r > 0$, such that $|f_\lambda|_{C^2, \mathcal{B}_r} < \infty$.

Define the error function

$$e_0(\theta) = f_0(K_0(\theta)) - K_0(\theta + \omega).$$

There exists a constant $c > 0$, depending on σ, n, ρ_0, r, $|f_0|_{C^2, \mathcal{B}_r}$, $\|DK_0\|_{\rho_0}$, $\left\| \frac{\partial f_\lambda}{\partial \lambda}\Big|_{\lambda=0} K_0 \right\|_\rho$, $\|N_0\|_{\rho_0}$, and $\left| (\operatorname{avg}\{\Lambda_0(\theta)\}_\theta)^{-1} \right|$ (where N_0, and Λ_0 are as in Definition 3, replacing K by K_0) such that, if $\|e_0\|_{\rho_0}$ verifies the following inequalities

$$c\,\gamma^{-4}\,\delta_0^{-4\sigma}\,\|e_0\|_{\rho_0} < 1\,,$$

and

$$c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} < r\,,$$

where $0 < \delta_0 \leq \min(1, \rho_0/12)$, then there exists a mapping $K_\infty \in \tilde{P}_{\rho_0 - 6\delta_0}$ and a vector $\lambda_\infty \in \mathbb{R}^{2n}$, satisfying

$$f_{\lambda_\infty} \circ K_\infty = K_\infty \circ T_\omega\,.$$

Moreover, the following inequalities hold

$$\|K_\infty - K_0\|_{\rho_0 - 6\delta_0} < c\,\gamma^2\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0}\,,$$

$$|\lambda_\infty| < c\,\gamma^2\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0}\,.$$

Remark 6. *The same result holds if we consider a d-parametric family of symplectic maps with $d \geq 2n$. Moreover, if the elements of the family are exact symplectic we can consider $d \geq n$.*

In the present case we have also local uniqueness.

Theorem 4. *Assume that $\omega \in D(\gamma, \sigma)$. Let $K_1, K_2 \in \mathcal{ND}(\rho)$ be solutions of*

$$f_{\hat{\lambda}} \circ K = K \circ T_\omega \ ,$$

satisfying the hypotheses of Theorem 3, and such that: $K_1(U_\rho) \subset \mathcal{B}_r$, and $K_2(U_\rho) \subset \mathcal{B}_r$.

There exists a constant $c > 0$ depending on n, σ, γ, ρ, ρ^{-1}, $\left|f_{\hat{\lambda}}\right|_{C^2,\mathcal{B}_r}$, $|J|_{\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$, $\left\| \frac{\partial f_\lambda}{\partial \lambda}\Big|_{\lambda=0} K_2 \right\|_\rho$, $\|K_2\|_{C^2,\rho}$, $\|N_2\|_\rho$, and $\left|(\text{avg}\,\{\Lambda_2\}_\theta)^{-1}\right|$ (with N_2, and Λ_2 are as in Definition 3, replacing K by K_2), such that if $\|K_1 - K_2\|_\rho$ satisfies

$$c \, \|K_1 - K_2\|_\rho < 1 \,,$$

where $\delta = \rho/8$, then there exists an initial phase $\tau \in \mathbb{R}^n$, such that $K_1 \circ T_\tau = K_2$ in $U_{\rho/2}$.

2.3 Vector fields

The results for a Hamiltonian vector field with Hamiltonian function $H : \mathbf{U} \to \mathbb{R}$ are based on the study of the equation

$$\partial_\omega \ K(\theta) = J \, \nabla H(K(\theta)) \,, \tag{16}$$

where J is defined in (8) and ∂_ω is the derivative in direction ω:

$$\partial_\omega \ K \stackrel{\text{def}}{=} \sum_{i=1}^{n} \omega_i \, \frac{\partial}{\partial \theta_i} K \,,$$

where $K : \mathbb{T}^n \to \mathbf{U}$ is the function to be determined and $\omega \in D_h(\gamma, \sigma) \subset \mathbb{R}^n$ (see Definition 1).

Notice that (16) implies that the range of K is invariant under the Hamiltonian vector field with Hamiltonian function H, X_H. The map K gives a parameterization of the invariant torus which makes the dynamics of the vector filed X_H restricted to the torus into a rigid rotation.

In order to simplify the notation we assume that $\Omega = dx \wedge dy$. This is not restrictive because the proof of Theorem 5 follows from Theorem 1 (see [Dou82]). Moreover, the direct construction given in Section 8 works for a general exact symplectic structure $\Omega = d\alpha$.

Definition 4. *Given a Hamiltonian function H and a frequency vector ω. The mapping $K : \mathbb{T}^n \to \mathbf{U}$ is say to be non-degenerate (for the Hamiltonian vector field X_H) if it satisfies the following conditions:*

- *K is real analytic on the set U_ρ.*

- *There exists a matrix N satisfying*

$$N(\theta)\, DK(\theta)^\top DK(\theta) = I_n.$$

- $\operatorname{avg}\{S\}_\theta$ *is invertible, with*

$$S(\theta) = N(\theta) DK(\theta)^\top \left[A(\theta)\, J - JA(\theta)\right]\, DK(\theta) N(\theta)$$

where

$$A(\theta) = \begin{pmatrix} D_x \nabla_y H(K(\theta)) & D_y \nabla_y H(K(\theta)) \\ -D_x \nabla_x H(K(\theta)) & -D_y \nabla_x H(K(\theta)) \end{pmatrix}.$$

Theorem 5. *Let ω satisfy the Diophantine condition (9). Assume that K_0 is non-degenerate (i.e. satisfies Definition 4). Assume that H is real analytic and that it can be holomorphically extended to some complex neighborhood of the image of U_ρ under K_0:*

$$\mathcal{B}_r = \{z \in \mathbb{C}^{2n} : \sup_{\theta \in U_\rho} |z - K_0(\theta)| < r\}.$$

Define the error function

$$e_0 \overset{\text{def}}{=} \partial_\omega\ K_0(\theta) - J\, \nabla H(K_0(\theta))\,.$$

There exists a constant $c > 0$, depending on σ, n, r, ρ, $|H|_{C^3, \mathcal{B}_r}$, $\|DK_0\|_\rho$, $\|N_0\|_\rho$, and $\left|(\operatorname{avg}\{S_0\}_\theta)^{-1}\right|$ (with N_0 and S_0 given by Definition 4 replacing K by K_0), such that if

$$c\, \gamma^{-4}\, \delta_0^{-4\sigma}\, \|e_0\|_\rho\ < 1\,,$$

and

$$c\, \gamma^{-2}\, \delta_0^{-2\sigma}\, \|e_0\|_{\rho_0}\ < r\,,$$

with $\delta_0 = \rho/12$, then there exists a solution for (16), K_∞, which is real analytic on $U_{\rho/2}$ and satisfies the non-degenerate conditions in Definition 4. Moreover,

$$\|K_\infty - K_0\|_{\rho/2}\ \le c\, \gamma^{-2} \delta_0^{-2\sigma} \|e_0\|_\rho\,.$$

Remark 7. *There is a local uniqueness statement for vector fields which follows by the reduction to the time one map (see [Dou82]). We do not formulate such a theorem.*

We will say that H_λ is a d-parametric family of Hamiltonian functions if there is a function $H : \mathbf{U} \times B \to \mathbb{R}$, with $B \subset \mathbb{R}^d$, such that for each $x \in \mathbf{U}$ the map $H(x, \cdot)$ is C^2 and such that for each $\lambda \in B$, the map $H_\lambda \overset{\text{def}}{=} H(\cdot, \lambda) : \mathbf{U} \to \mathbb{R}$ is a real analytic Hamiltonian function.

There is an analogue of the translated torus theorem (Theorem 3) in which we study the equation

$$\partial_\omega K(\theta) = J \nabla H_\lambda(K(\theta))\,, \tag{17}$$

where H_λ is a $2n$-parametric family of Hamiltonian functions, and $K \in \tilde{\mathcal{P}}_\rho$ (see equation (10)).

Theorem 6. *Let ω satisfy the Diophantine condition (9). Assume that $K_0 \in \tilde{\mathcal{P}}_\rho$ satisfies the following properties:*

- *There exists a matrix N_0 satisfying*

$$N_0(\theta)\, DK_0(\theta)^\top\, DK_0(\theta) = I_n.$$

- $\operatorname{avg}\{\Lambda_0\}_\theta$ *has maximum range, with*

$$\Lambda_0(\theta) = \begin{pmatrix} N_0(\theta) DK_0(\theta)^\top J \left(\frac{\partial}{\partial\lambda}\ \nabla H_\lambda(K_0(\theta))\,|_{\lambda=0} \right) \\ \\ DK_0(\theta)^\top J \left(\frac{\partial}{\partial\lambda}\ \nabla H_\lambda(K_0(\theta))\,|_{\lambda=0} \right) \end{pmatrix}.$$

Define the error function

$$e_0 \overset{\text{def}}{=} \partial_\omega\ K_0(\theta) - J\,\nabla H_0(K_0(\theta))\,.$$

There exists a constant $c > 0$, depending on σ, n, r, ρ, $|H|_{C^3,\mathcal{B}_r}$, $\|DK_0\|_\rho$, $\|N_0\|_\rho$, $\left|(\operatorname{avg}\{S_0\}_\theta)^{-1}\right|$, $\|\Lambda_0\|_\rho$, and $\left|(\operatorname{avg}\{\Lambda_0\}_\theta)^{-1}\right|$ such that if

$$c\,\gamma^{-4}\,\delta_0^{-4\sigma}\,\|e_0\|_\rho\ < 1\,,$$

and

$$c\,\gamma^{-2}\,\delta_0^{-2\sigma}\ \|e_0\|_{\rho_0}\ < r\,,$$

with $\delta_0 = \rho/12$, then there exist $\lambda_\infty \in \mathbb{R}^{2n}$ and $K_\infty \in \tilde{\mathcal{P}}_{\rho/2}$ satisfying

$$\partial_\omega K_\infty(\theta) = J\nabla H_{\lambda_\infty}(K_\infty(\theta))$$

Moreover,

$$\|K_\infty - K_0\|_{\rho/2}\ \le c\,\gamma^{-2}\delta_0^{-2\sigma}\|e_0\|_\rho\,,$$

and

$$|\lambda_\infty|\ <\ c\,\gamma^2\ \delta_0^{-2\sigma}\,\|\,e_0\|_{\rho_0}\ .$$

The perturbative case

Even if one of the strong points of the proof presented here is that it does not require that the problem is in action angle coordinates nor close to integrable, here we will present how the results here apply to the quasi-integrable case. We hope that this can help understand the meaning of the non-degeneracy conditions.

We consider $\mathbb{T}^n \times U$ endowed with the usual symplectic structure.

If we consider a map

$$f(\varphi, I) = (\varphi + A(I), I) + \Delta(\varphi, I)$$

where Δ is small, we can take as $K(\theta) = (\theta, I^*)$ where I^* is chosen in such a way that $A(I^*) = \omega$).

Then, we see that the size of the error function can be controlled by the size of Δ.

The non-degeneracy condition **N1** on the embedding is satisfied clearly since $N(\theta) = I_n$.

We have that, in this case P is the constant matrix

$$P(\theta) = \begin{pmatrix} I_n \\ 0 \end{pmatrix}$$

Since Δ is assumed to be small, we will verify the hypothesis on the main part of the map considering the Δ as an small perturbation.

Note that

$$Df(K(\theta)) = \begin{pmatrix} I_n & DA(I^*) \\ 0 & I_n \end{pmatrix} + O(\Delta)$$

Performing the calculation in the definition of S, we obtain that, in this perturbative case, $S = DA(I^*) + O(\Delta)$.

Hence, in this perturbative case, the non-degeneracy condition we assume is precisely the twist condition.

We leave to the reader the verification that in the other quasi-integrable cases, the non-degeneracy conditions we assume reduce to the standard ones.

3 Sketch of the procedure

Equations (1), (15), (16), and (17) are functional equations for the unknown function K, (and the unknown parameter λ in the case of (15) and (17)). As it is shown in [Mos66a, Zeh75], in order to prove the existence of a solution of a non-linear problem by using a modified Newton method it is enough to have an approximate solution of the corresponding linearized equation such that the new error is quadratic with respect to the original one.

The main idea is to use the geometric properties of the problem to prove that the corresponding linearized equations can be transformed into a simpler linear equation that is approximately solvable by using Fourier series.

We emphasize that, at each step of the iterative procedure, the reduction of the error is accomplished by adding a small function rather than composing with a near identity transformation (as it is customary in many proofs of the KAM theorem). This leads to a simpler and more efficient estimates, which are closer to those obtained in numerical procedures. It is also interesting to note that the procedure presented here can be implemented to compute numerical approximations of invariant tori.

Maps

The key point on finding a solution for (1) and (15) is to solve approximately the corresponding linearized equation. The main part of both linearized equations is the linear operator:

$$\mathcal{T}_{f,K,\omega}\,\Delta = Df(K(\theta))\,\Delta - \Delta\circ T_\omega\,. \tag{18}$$

Indeed, equation (1) and (15) can be formulated as finding zeros of

$$F(K) \overset{\text{def}}{=} f\circ K - K\circ T_\omega\,, \tag{19}$$

and

$$G(K,\lambda) \overset{\text{def}}{=} f_\lambda\circ K - K\circ T_\omega\,, \tag{20}$$

respectively. Therefore, at least formally we have

$$\begin{aligned} DF(K)\,\Delta &= \mathcal{T}_{f,K,\omega}\,\Delta \\ DG(K,\lambda)\,(\Delta K,\Delta\lambda) &= \mathcal{T}_{f_\lambda,K,\omega}\,(\Delta K)\;+\;\frac{\partial f_{\lambda(K)}}{\partial}\Delta\lambda\,. \end{aligned}$$

In Section 4 we prove that, under certain hypotheses, the corresponding linear operator $DF(K)$ is approximately invertible in the sense of [Zeh75]. First of all we prove that the Lagrangian character of an invariant torus, for a symplectic map, is slightly modified if the torus is only approximately invariant (see Section 4.1). The approximate Lagrangian character of an approximately invariant torus enables us to define a change of variables such that in the new variables the linear operator $DF(K)$ has a simple form (see sections 4.2 and 4.3). In Section 4.4 we show that, if some non-degeneracy conditions hold, in the new variables the linearized equation

$$DF(K)\Delta = -F(K),$$

can be solved approximately.

Similarly, in Section 7 we prove (with different non-degeneracy conditions) that

$$DG(K,\lambda)(\Delta K,\delta\lambda) = -G(K,\lambda)$$

can be solved approximately.

4 The linearized operator $DF(K)$

In order to construct a modified Newton method for the functional equation (19) we have to study the invertibility properties of the corresponding linear operator $DF(K)$. More concretely, we are interested in the solvability properties of the linearized equation

$$DF(K)\Delta(\theta) = Df(K(\theta))\,\Delta(\theta)\;-\;\Delta(\theta+\omega)\;=-\,e(\theta) \tag{21}$$

where K is an approximate solution of (1) with error function

$$e(\theta) = F(K)(\theta), \quad \forall\, \theta \in \mathbb{R}^n\ . \tag{22}$$

By using the geometric properties of the problem we will prove that (21) can be transformed into another equation which has an upper triangular form (see equation (45)) that can be solved approximately by using Fourier series. To be more precise, by using the symplectic properties of f and the algebraic properties of ω we will show that, if the error e defined in (22) is small enough, then for any $\theta \in \mathbb{T}^n$ the set

$$\left\{ \frac{\partial K(\theta)}{\partial \theta_j} \quad , J(K(\theta)) \frac{\partial K(\theta)}{\partial \theta_j} \right\}_{j=1}^{n}$$

forms a basis of $T_{K(\theta)}\mathbf{U} \sim \mathbb{R}^{2n}$ (the tangent space to $\mathbf{U}$ at the point $K(\theta)$). We use this basis to transform the linearized equation (21) into a convenient linear equation for which we can construct an approximate solution.

4.1 Approximate Lagrangian character

In this section we formulate the well known result that a torus supporting an irrational rotation is Lagrangian in terms of a function L being zero (Lemma 3). We will also show that for parameterizations which satisfies (1) approximately, then the norm of L can be estimated by the norm of the error given in (22), e (Lemma 4).

Let L be the linear application defined by

$$\left(K^*\,\Omega\right)_\theta(\xi,\eta) = <\xi, L(\theta)\eta>, \qquad \text{for any} \quad \xi,\eta \in \mathbb{R}^{2n}, \quad \theta \in \mathbb{T}^n,$$

then

$$L(\theta) = DK(\theta)^\top J(K(\theta)) DK(\theta), \tag{23}$$

where J is defined by (8). With the previous notation the Lagrangian character of K is equivalent to the following equality

$$L(\theta) = 0\,, \qquad \forall \theta \in \mathbb{T}^n.$$

Therefore, it is natural to say that an approximately invariant torus is *approximately Lagrangian* if the norm of L is 'small'.

Exact solutions

In order to make the proof of the approximately Lagrangian character of an approximate solution of (1) clearer we present a prove of the Lagrangian character of an exact solution of (1).

Lemma 3. *Assume that K is a solution of (1) then $K^*\,\Omega$ (equivalently, L) is identically zero.*

Proof. And since K satisfies (1) and f is symplectic we have

$$f \circ K = K \circ T_\omega\,, \qquad \text{and} \qquad f^*\,\Omega = \Omega\,.$$

Then,

$$K^*\,\Omega = K^*\,(f^*\,\Omega) = (f \circ K)^*\;\Omega = (K \circ T_\omega)^*\;\Omega\,,$$

and so we have

$$K^*\,\Omega - (K \circ T_\omega)^*\;\Omega = 0.$$

Since ω is rationally independent, rotations on the torus are ergodic, this implies that $K^*\,\Omega$ is constant.

Moreover, since Ω is exact with $\Omega = d\,\alpha$, then if $K = (K_1, \dots, K_{2n})^\top$ and if in (7) $a = (a_1, \dots, a_{2n})^\top$, then

$$(K^*\alpha)_\theta = \sum_{j=1}^{n} c_j(\theta)\, d\theta_j$$

where for $s = 1, \dots, n$, c_s has the following expression

$$c_s(\theta) = \Big(DK(\theta)^\top a(K(\theta))\Big)_s = \sum_{j=1}^{2n} \frac{\partial K_j(\theta)}{\partial \theta_s}\, a_j(K(\theta)).$$

This implies $L(\theta) = Dc(\theta)^\top - Dc(\theta)$, and since the average of $Dc(\theta)$ is equal to zero one obtains that the average of L is equal to zero. ☐

Remark 8. *Lemma 3 also holds if ω is only rationally independent.*

Approximate solutions

The approximate Lagrangian character of an approximate solution of (1) follows because $K^*\Omega$ satisfies a small divisors equation, similar to (12), with the right hand side expressed in terms of derivatives of the error function e defined in (22). We make this more precise in the following statement.

Lemma 4. *Let e be the function error defined in (22), then $K^*\,\Omega$ satisfies the following linear difference equation*

$$K^*\Omega - (K \circ T_\omega)^*\,\Omega = \Omega_e, \tag{24}$$

where

$$(\Omega_e)_\theta = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{i\,j}(\theta)\, dz_i\, dz_j,$$

where $g_{i\,j}$ are the coordinates of the $n \times n$ matrix g given in (25). In particular, the average of g is equal to zero.

Moreover, if K is real analytic on the complex strip of width ρ, then there exists a constant $c > 0$, depending on n, σ, ρ, $\|DK\|_\rho$, $|f|_{C^1,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, such that, for $0 < \delta < \rho/2$

$$\|L\|_{\rho-2\delta} \leq c_1 \gamma^{-1} \delta^{-(\sigma+1)} \|e\|_\rho ,$$

where L is given in (23).

Proof. Define $\Omega_e = K^* \Omega - (K \circ T_\omega)^* \Omega$, then

$$\Omega_e = (f \circ K)^* \Omega - (K \circ T_\omega)^* \Omega$$

then

$$g(\theta) = DK(\theta)^\top Df(K(\theta))^\top J(f(K(\theta)))Df(K(\theta))DK(\theta) - L(\theta+\omega),$$

where L is defined in (23). Using the equality

$$Df(K(\theta))DK(\theta) = DK(\theta+\omega) + De(\theta) ,$$

which follows from (22), and performing some computations, one obtains

$$g(\theta) = DK(\theta+\omega)^\top \varphi(\theta)\, DK(\theta+\omega) - \psi(\theta)^\top + (Df(K(\theta))DK(\theta))^\top J(f(K(\theta)))De(\theta) \tag{25}$$

where

$$\varphi(\theta) = J(f(K(\theta))) - J(K(\theta+\omega)),$$

and

$$\psi(\theta) = DK(\theta+\omega)^\top J(f(K(\theta)))De(\theta).$$

Notice that if K is real analytic on the complex strip of width ρ, then

$$\begin{aligned} \|\varphi\|_\rho &\leq |J|_{C^1,\mathcal{B}_r} \|e\|_\rho , \\ \|\psi\|_{\rho-\delta} &\leq c\,\delta^{-1} \|e\|_\rho , \end{aligned}$$

where c is a constant depending on n, $\|DK\|_\rho$, and $|f|_{C^1,\mathcal{B}_r}$. Therefore, from (25) we have

$$\|g\|_{\rho-\delta} \leq c\,\delta^{-1} \|e\|_\rho .$$

Applying the last inequality and Proposition 1 we obtain an estimation for the norm of L. Indeed, equation (24) is equivalent to

$$L - L \circ T_\omega = g .$$

Proposition 1 implies

$$\|L\|_{\rho-2\delta} \leq c_0 \gamma^{-1} \delta^{-\sigma} \|g\|_{\rho-\delta} \leq c_0 \gamma^{-1} c\,\delta^{-(\sigma+1)} \|e\|_\rho$$

where c_0 is the constant of Proposition 1. □

4.2 Quasi-reducibility

One way of studying the solvability of the linearized equation (21) is to define a change of variables $\Delta(\theta) = M(\theta)\,\xi(\theta)$ such that in the variable ξ equation (21) is simple. Equivalently, $M(\theta)$ should be defined in such a way that

$$M(\theta+\omega)^{-1} Df(K(\theta)) M(\theta)$$

is simple.

In this section we will use the results of Section 4.1 to prove that if K is a solution for (1), then it is possible to define $M(\theta)$ in terms of $DK(\theta)$ satisfying

$$M(\theta+\omega)^{-1}\, Df(K(\theta))\, M(\theta) = C(\theta)\,, \tag{26}$$

where $C(\theta)$ is an upper triangular matrix with diagonal terms equal to 1. This is accomplished by using fact that $Df(K(\theta))$ is a linear symplectic transformation taking the tangent vector $\frac{\partial K(\theta)}{\partial \theta_j}$ into the tangent vector $\frac{\partial K(\theta+\omega)}{\partial \theta_j}$, and the Lagrangian character of an invariant torus.

When K is an approximate solution for (1), if we define $M(\theta)$ and $C(\theta)$ in the same way as we did in the case of an exact solution, even though equality (26) does not hold any more, we have

$$M(\theta+\omega)^{-1} Df(K(\theta))\, M(\theta) = C(\theta) + B(\theta)\,,$$

where (see Proposition 9)

$$\|B(\theta)\|_{\rho-2\delta} \leq \gamma^{-1}\delta^{-(\sigma+1)} c\, \|e\|_{\rho}\ .$$

Exact solutions

If K satisfies (1), taking derivatives we obtain

$$Df(K(\theta))\, DK(\theta) = DK(\theta+\omega) \tag{27}$$

hence $M(\theta)$ and $C(\theta)$ can be chosen to have the following form

$$M(\theta) = \begin{pmatrix} DK(\theta) & M_1(\theta) \end{pmatrix}\,, \quad \text{and} \quad C(\theta) = \begin{pmatrix} I_n & S(\theta) \\ 0 & H(\theta) \end{pmatrix}\,,$$

where S and H are $n \times n$ matrix-valued functions, and M_1 is an $2n \times n$ matrix-valued function, that have to be determined.

Define

$$N(\theta) \overset{\text{def}}{=} \left(DK(\theta)^{\top} DK(\theta)\right)^{-1}. \tag{28}$$

If we take M_1 to be

$$M_1(\theta) = J(K(\theta))^{-1} DK(\theta) N(\theta)\,,$$

some simple algebraic calculations show that H is the identity matrix and

$$S(\theta) = P(\theta+\omega)^\top \left[Df(K(\theta))J(K(\theta))^{-1}P(\theta) \; - J(K(\theta+\omega))^{-1}P(\theta+\omega) \right] \tag{29}$$

with

$$P(\theta) \overset{\text{def}}{=} DK(\theta)\, N(\theta)\,. \tag{30}$$

Therefore, we have proved the following result.

Proposition 5. *Assume that K satisfies (1). If*

$$M(\theta) \overset{\text{def}}{=} \begin{pmatrix} DK(\theta) & J(K(\theta))^{-1}DK(\theta)N(\theta) \end{pmatrix}, \tag{31}$$

and

$$C(\theta) \overset{\text{def}}{=} \begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix}, \tag{32}$$

with S defined in (29). Then

$$Df(K(\theta))M(\theta) = M(\theta+\omega)\, C(\theta)$$

Remark 9. *If the symplectic linear transformation satisfies that $J^{-1} = -J$, then*

$$P(\theta+\omega)^\top J(K(\theta+\omega))^{-1} P(\theta+\omega) = 0\,.$$

Therefore, S takes the following form

$$S(\theta) = -P(\theta+\omega)^\top Df(K(\theta))J(K(\theta))P(\theta)\,.$$

Remark 10. *It is possible to make further transformations that make C constant. Nevertheless, since the present form is good enough for our proposes, we do not pursue it here.*

Approximate solutions

Now assume that K is an approximate solution of (1) with error e defined in (22). Define S, M, and C by (29), (31), and (32), respectively, and let us define E as

$$E(\theta) \overset{\text{def}}{=} Df(K(\theta))M(\theta) - M(\theta+\omega)\, C(\theta). \tag{33}$$

We would like to estimate (see Proposition 9)

$$M(\theta+\omega)^{-1}\, E(\theta) = M(\theta+\omega)^{-1} Df(K(\theta))M(\theta) - C(\theta)$$

However it would be easier to use $V(\theta)^{-1}M(\theta)^\top J(K(\theta))$ (see (44)) in place of $M(\theta+\omega)^{-1}$. As we will see later (Lemma 8), this introduce very small errors since M is approximately symplectic.

Lemma 6. *Assume that the hypotheses of Lemma 4 hold. Then there exists a constant* c_2*, depending on* n*,* ρ*,* $|f|_{C^1,\mathcal{B}_r}$*,* $|J|_{C^1,\mathcal{B}_r}$*,* $\left|J^{-1}\right|_{\mathcal{B}_r}$*,* $\|DK\|_\rho$*,* $\|N\|_\rho$*, and* c_1 *such that, for any* $0<\delta<\rho/2$*, the following estimate holds*

$$\left\| M(\theta+\omega)^\top J(K(\theta+\omega))\, E(\theta) \right\|_{\rho-2\delta} \leq c_2\, \gamma^{-1}\, \delta^{-(\sigma+1)}\, \|e\|_\rho \,. \tag{34}$$

Proof. Using the following equality

$$Df(K(\theta))\, DK(\theta) - DK(\theta+\omega) = De(\theta),$$

we have

$$E(\theta) = \begin{pmatrix} De(\theta) & E_1(\theta) \end{pmatrix},$$

where

$$\begin{aligned} E_1(\theta) \;\; = \;\; & Df(K(\theta))J(K(\theta))^{-1}DK(\theta)N(\theta) - \\ & DK(\theta+\omega)S(\theta) - J(K(\theta+\omega))^{-1}DK(\theta+\omega)N(\theta+\omega). \end{aligned}$$

On the other hand, performing some simple computations, we have

$$M(\theta)^\top J(K(\theta)) = \begin{pmatrix} Q(\theta)^\top \\ -P(\theta)^\top \end{pmatrix},$$

where P is defined by equation (30), and $Q(\theta) = -J(K(\theta))DK(\theta)$. Then, we have

$$M(\theta+\omega)^\top J(K(\theta+\omega))\, E(\theta) = \begin{pmatrix} Q(\theta+\omega)^\top De(\theta) & Q(\theta+\omega)^\top E_1(\theta) \\ -P(\theta+\omega)^\top De(\theta) & -P(\theta+\omega)^\top E_1(\theta) \end{pmatrix}.$$

Moreover, E_1 satisfies

$$P(\theta+\omega)^\top E_1(\theta) = 0, \tag{35}$$

and

$$Q(\theta+\omega)^\top E_1(\theta) = \phi(\theta) - \psi(\theta) - L(\theta+\omega)S(\theta), \tag{36}$$

where $L(\theta)$, defined in (23), is bounded in Lemma 4,

$$\phi(\theta) = (Df(K(\theta))\, DK(\theta))^\top\, \varphi(\theta)\, Df(K(\theta))J(K(\theta))^{-1}DK(\theta)N(\theta),$$

with $\varphi(\theta) = J(\,K(\theta+\omega)\,) - J(f(K(\theta))$, and

$$\psi(\theta) = De(\theta)^\top J(K(\theta+\omega))Df(K(\theta))J(K(\theta))^{-1}DK(\theta)N(\theta).$$

Then inequality (34) follows from equalities (35), (36), and Lemma 4. □

4.3 Change of variables in the linearized equation

In this section we will determine conditions on $\|e\|_\rho$ such that the matrix-valued function M given by (31) is invertible, and hence we can use it to define a change of variables: $\Delta(\theta) = M(\theta)\,\xi(\theta)$.

If M is defined by (31), then a direct computation yields

$$M(\theta)^\top J(K(\theta))M(\theta) = V(\theta) \,+\, R(\theta), \tag{37}$$

where

$$V(\theta) \overset{\text{def}}{=} \begin{pmatrix} 0 & I_n \\ -I_n & -P(\theta)^\top J(K(\theta))^{-1}P(\theta) \end{pmatrix}, \tag{38}$$

and

$$R(\theta) = \begin{pmatrix} L(\theta) & 0 \\ 0 & 0 \end{pmatrix}, \tag{39}$$

with P, L and N defined in (30), (23), (28), respectively.
Notice that V is invertible, with inverse given by

$$V(\theta)^{-1} = \begin{pmatrix} -P(\theta)^\top J(K(\theta))^{-1}P(\theta) & -I_n \\ I_n & 0 \end{pmatrix}. \tag{40}$$

Exact solutions

If K satisfies (1), then Lemma 3 implies $L = 0$, this implies $R = 0$. Hence, equality (37) implies that M is invertible and its inverse is given by

$$M(\theta)^{-1} = V(\theta)^{-1}M(\theta)^\top J(K(\theta)) \,=\, \begin{pmatrix} T(\theta) \\ DK(\theta)^\top J(K(\theta)) \end{pmatrix},$$

with V^{-1} given in (40) and

$$T(\theta) \overset{\text{def}}{=} P(\theta)^\top \left[I_n - J(K(\theta))^{-1}P(\theta)DK(\theta)^\top J(K(\theta)) \right], \tag{41}$$

where $P(\theta)$ is given in (30).

Since M and C satisfy equality (26), we have that $M(\theta+\omega)^{-1}Df(K(\theta))\,M(\theta) = C(\theta)$. Therefore, if the we substitute $\Delta(\theta) = M(\theta)\,\xi(\theta)$ in the linear equation (21) we have

$$C(\theta)\,\xi(\theta) \,-\, \xi(\theta+\omega) \,=\, p(\theta)\,,$$

with

$$p(\theta) = -\,V(\theta)^{-1}M(\theta)^\top J(K(\theta))\,e(\theta) = \begin{pmatrix} -T(\theta+\omega)e(\theta) \\ -DK(\theta+\omega)^\top J(K(\theta+\omega))\,e(\theta) \end{pmatrix}. \tag{42}$$

where $P(\theta) = DK(\theta)\,N(\theta)$ (see (30)), and T is defined in (41).

Remark 11. *If* $J^{-1} = -J$, *then*

$$V(\theta) = \begin{pmatrix} 0 & I_n \\ -I_n & N(\theta)L(\theta)N(\theta) \end{pmatrix},$$

where L *is defined in (23). Then*

$$M(\theta)^\top J(K(\theta))M(\theta) = J_0 + \begin{pmatrix} L(\theta) & 0 \\ 0 & N(\theta)L(\theta)N(\theta) \end{pmatrix},$$

where $J_0 = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$.
In particular, if K *is a parameterization of an invariant torus and* Ω *is the canonical symplectic form in* $\mathbb{R}^{2n}$, *then* M *is symplectic:*

$$M^\top J_0 M = J_0.$$

Approximate solutions

Assume that K is approximate solution of (1) with error given by (22), and M is defined by (31). From equality (37) one sees that if the matrix $V(\theta)+R(\theta)$ is invertible, then $M(\theta)$ is also invertible. The invertibility of $V(\theta)+R(\theta)$ depends on the size of $\|e\|_\rho$, in order to show that we first bound the norm of $V(\theta)^{-1}R(\theta)$ in terms of $\|e\,\|_\rho$. Using the expression (38) and (39) the following lemma is immediate.

Lemma 7. *Assume that the hypotheses of Lemma 4 hold. Then there exists a constant* c_3 *depending on* n, σ, ρ, $|f|_{C^1,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $|J^{-1}|_{\mathcal{B}_r}$, $\|N\|_\rho$, $\|DK\|_\rho$, *and the constant* c_1 *of Lemma 4 such that for any* $0<\delta<\rho/2$ *the following inequality holds*

$$\left\|V(\theta)^{-1}R(\theta)\,\right\|_{\rho-2\delta} < c_3\,\gamma^{-1}\,\delta^{-(\sigma+1)}\,\|e\|_\rho\,.$$

As a consequence of Lemma 7 we have.

Lemma 8. *Assume that the hypotheses of Lemma 4 hold, and let* c_3 *be the constant of Lemma 7. If* e *satisfies*

$$c_3\,\gamma^{-1}\,\delta^{-(\sigma+1)}\,\|e\|_\rho \le 1/2\,, \tag{43}$$

then M *is invertible and its inverse given by*

$$M(\theta)^{-1} = V(\theta)^{-1}M(\theta)^\top J(K(\theta)) + M_e(\theta)\,, \tag{44}$$

where

$$M_e(\theta) \overset{\text{def}}{=} -\left[I_{2n} + V(\theta)^{-1}R(\theta)\right]^{-1}\,V(\theta)^{-1}R(\theta)\,V(\theta)^{-1}\,M(\theta)^\top J(K(\theta))\,.$$

In particular, $M_e = 0$ *if* $e = 0$. *Moreover,*

$$\|\,M_e\|_{\rho-2\delta} \le c_4\,\gamma^{-1}\,\delta^{-(\sigma+1)}\,\|e\|_\rho\,,$$

where c_4 *is a constant depending on* n, c_3, $|J^{-1}|_{\mathcal{B}_r}$, $|J|_{\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N\|_\rho$.

Proof. From Lemma 7 and the assumption on $\|e\|_\rho$ one obtains

$$\| \, V(\theta)^{-1} \, R(\theta) \, \|_{\rho-2\delta} \leq c_3 \, \gamma^{-1} \, \delta^{-(\sigma+1)} \, \|e\|_\rho \, \leq \, 1/2.$$

Then $I_{2n} + V(\theta)^{-1} R(\theta)$ is invertible with

$$\left\| \left(I_{2n} + V(\theta)^{-1} R(\theta) \right)^{-1} \right\|_{\rho-2\delta} \leq \frac{1}{1 - \| \, V(\theta)^{-1} R(\theta) \, \|_{\rho-2\delta}} \leq 2 \, ,$$

and using equality (37) we have

$$\begin{aligned} M(\theta)^{-1} &= (\, V(\theta) + R(\theta) \,)^{-1} \, M(\theta)^\top J(K(\theta)) \\ &= V(\theta)^{-1} M(\theta)^\top \, J(\, K(\theta)) \, + \, M_e(\theta). \end{aligned}$$

□

Proposition 9. *Assume that* $\omega \in D(\gamma, \sigma)$, *and that* $\|e\|_\rho$ *satisfies inequality (43), where* c_3 *is the constant of Lemma 7. Then, the change of variables defined by*

$$\Delta(\theta) = M(\theta) \, \xi(\theta) \, ,$$

transforms (21) into

$$\left[\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} + B(\theta) \right] \xi(\theta) - \xi(\theta + \omega) = p(\theta) + w(\theta) \, , \tag{45}$$

where S *is defined in (29),* p *is given in (42), and*

$$B(\theta) \overset{\text{def}}{=} M(\theta + \omega)^{-1} \, E(\theta) \, , \qquad \qquad w(\theta) \overset{\text{def}}{=} -M_e(\theta + \omega) \, e(\theta) \, , \tag{46}$$

with E *is given in (33). In particular, from (34) we have*

$$\| \, B(\theta) \|_{\rho-2\delta} \, < \, c_5 \, \gamma^{-1} \, \delta^{-(\sigma+1)} \, \|e\|_\rho \, , \tag{47}$$

and

$$\|w\|_{\rho-2\delta} \, < \, c_4 \, \gamma^{-1} \, \delta^{-(\sigma+1)} \, \|e\|_\rho^2 \, , \tag{48}$$

where c_5 *is a constant which depends on* c_2 *(the constant of Lemma 6),* $\left| J^{-1} \right|_{\mathcal{B}_r}$, $\|DK\|_\rho$, *and* $\|N\|_\rho$, *and* c_4 *is the constant of Lemma 8.*

Proof. In the variable ξ, equation (21) takes the form

$$Df(K(\theta)) \, M(\theta) \, \xi(\theta) - M(\theta + \omega) \, \xi(\theta + \omega) = - \, e(\theta) \, ,$$

multiplying by $M(\theta + \omega)^{-1}$ we obtain

$$M(\theta + \omega)^{-1} Df(K(\theta)) \, M(\theta) \, \xi(\theta) - \xi(\theta + \omega) = - \, M(\theta + \omega)^{-1} \, e(\theta) \, .$$

Using equality (33) and the definition of B we have

$$M(\theta+\omega)^{-1}Df(K(\theta))\,M(\theta)=C(\theta)+B(\theta)\,.$$

Moreover, from (44) we have

$$-\,M(\theta+\omega)^{-1}\,e(\theta)=-\,V(\theta+\omega)^{-1}M(\theta+\omega)J(K(\theta+\omega))\,e(\theta)-M_e(\theta+\omega)\,e(\theta)\,,$$

defining p and w as in (42) and (46), respectively, we have the left hand part of (45). Estimate (47) follows from

$$\|B\|_{\rho-2\delta}\leq\left\|V^{-1}\right\|_{\rho}\left\|M(\theta+\omega)^{\top}J(K(\theta+\omega))E(\theta)\right\|_{\rho-2\delta}+\|M_eE\|_{\rho-2\delta}\,.$$

Finally, estimate (48) follows from Lemma 8. ☐

Remark 12. *Notice that in Proposition 9, we have* $w=0$ *and* $B=0$ *in the case that* K *parameterizes an invariant torus.*

4.4 Approximate solvability of the linearized equation

Results of sections 4.1, 4.2, and 4.3 guarantee the existence of a change of variables which takes (21) into (45). We emphasize that those results hold for any real analytic symplectic map f having a real analytic approximate solution of (1), K, with small enough error function and frequency vector $\omega\in D(\gamma,\sigma)$. In the present section we prove that if f is an exact symplectic map, and K satisfies the non-degeneracy condition **N2**, then equation (21) can be solved approximately in the customary sense of Nash-Moser theory. In order to do that we study the approximate solvability of (45).

Let us consider the following linear operator

$$\mathcal{L}_\omega\,\xi(\theta)=\begin{pmatrix} I_n & S(\theta)\\ 0 & I_n\end{pmatrix}\xi(\theta)-\xi(\theta+\omega)\,,$$

where S is defined in (29). Notice that (45) can be written as follows

$$\mathcal{L}_\omega\,\xi(\theta)+B(\theta)\,\xi(\theta)\,=\,p(\theta)+w(\theta).$$

The approximate solutions of (45) will be the solutions of

$$\mathcal{L}_\omega\,\xi(\theta)=p(\theta)-\begin{pmatrix}0\\ \operatorname{avg}\{p_y\}_\theta\end{pmatrix}\,,\tag{49}$$

where (p_x,p_y) are the symplectically conjugate variables of p. The exactness property of f enable us to prove that $\left|\operatorname{avg}\{p_y\}_\theta\right|$ is quadratic with respect to the norm of the error (see (54)). Then a solution of (49) will satisfy (45) up to errors which are 'quadratic' in the errors, which does not affect the 'quadratic' convergence of the methods [Zeh75].

The non-degeneracy condition **N2** in Definition 2 guarantees the solvability of (49), as we prove in the following result.

Proposition 10. *Assume that* $\omega \in D(\gamma, \sigma)$ *and that the following hypotheses hold*

- *(a) The average of* S, $\operatorname{avg}\{S\}_\theta$, *is non-singular.*
- *(b)* $v = (v_x, v_y) : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ *is real analytic on* U_ρ, 1*-periodic in all its arguments, with* $v_y : \mathbb{R}^n \to \mathbb{R}^n$ *a zero average function.*

There exists a unique function $\xi = (\xi_x, \xi_y) : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ *real analytic on* $U_{\rho-2\delta}$, *for any* $0 < \delta < \min(1, \rho/2)$, *satisfying*

$$\mathcal{L}_\omega \, \xi = v \,,$$

with $\operatorname{avg}\{\xi_x\}_\theta = 0$ *and*

$$\operatorname{avg}\{\xi_y\}_\theta = (\operatorname{avg}\{S\}_\theta)^{-1} \Big(\operatorname{avg}\{v_x\}_\theta \, - \, \operatorname{avg}\Big\{ \, S(\theta) \, \tilde{\xi_y} \Big\}_\theta \Big) \,, \tag{50}$$

where $\tilde{\xi_y} = \xi_y - \operatorname{avg}\{\xi_y\}_\theta$.

Moreover, there exists a constant c *depending on* σ, ρ, n, $\|N\|_\rho$, $\|K\|_\rho$, *and* $\left|(\operatorname{avg}\{S\}_\theta)^{-1}\right|$, *s.t.*

$$\|\, \xi \,\|_{\rho-2\delta} < c \, \gamma^{-2} \, \delta^{-2\sigma} \, \|v\,\|_\rho \,.$$

Proof. The linear difference equation $\mathcal{L}_\omega \, \xi = (v_x, v_y)^\top$ is written as follows

$$\begin{aligned} \xi_x(\theta) - \xi_x(\theta + \omega) &= v_x(\theta) - S(\theta) \, \xi_y(\theta) \\ \xi_y(\theta) - \xi_y(\theta + \omega) &= v_y(\theta) \end{aligned}$$

First we consider the n-dimensional equation

$$\xi_y(\theta) - \xi_y(\theta + \omega) = v_y. \tag{51}$$

Since $\omega \in D(\gamma, \sigma)$ and $\operatorname{avg}\{v_y\}_\theta = 0$, Proposition 1 implies the existence of a solution ξ_y of equation (51) which has arbitrary average and it is analytic on $U_{\rho-\delta}$, for any $0 < \delta < \rho/2$, and satisfies following estimate

$$\|\, \xi_y\|_{\rho-\delta} \leq \, c_0 \, \gamma^{-1} \, \delta^{-\sigma} \, \|\, v_y\|_\rho \, + \, \left|\operatorname{avg}\{\xi_y\}_\theta\right| \,,$$

where c_0 is the constant given in Proposition 1.

Moreover, if $\operatorname{avg}\{\xi_y\}_\theta$ is defined by (50), then the following equality holds

$$\operatorname{avg}\{\, v_x(\theta) - S(\theta) \, \xi_y(\theta)\}_\theta = 0 \,.$$

Thus, there exists a unique zero average solution ξ_x of

$$\xi_x(\theta) - \xi_x(\theta + \omega) = v_x(\theta) - S(\theta) \, \xi_y(\theta) \,, \tag{52}$$

satisfying

$$\|\xi_x\|_{\rho-2\delta} \leq \, c_0 \, \gamma^{-1} \, \delta^{-\sigma} \, \|\, v_x(\theta) - S(\theta)\xi_y(\theta)\|_{\rho-\delta} \,.$$

Finally, using the bound for $\|\xi_y\|_{\rho-\delta}$ and assuming $0 < \delta < \min(\rho, 1)$, one obtains the desired estimate. ☐

Remark 13. *We remark that Proposition 10 is the justification for our definition of the non-degeneracy condition* **N2** *(see Definition 2).*

Remark 14. *The uniqueness property of the solution ξ in Proposition 10 implies that for any other mapping $\hat{\xi} = (\hat{\xi}_x, \hat{\xi}_y)$ satisfying $\mathcal{L}_\omega \hat{\xi} = v$ will also satisfy*

$$\hat{\xi} = \xi + \begin{pmatrix} \operatorname{avg}\left\{\hat{\xi}_x\right\}_\theta \\ 0 \end{pmatrix} .$$

In order to use Proposition 10 to find an approximate solution of (45) we first notice that if we want to obtain an approximate solution of (45) with quadratic error with respect to norm of e, because of estimate (47) and the estimates given in Proposition 10, it is enough to obtain a solution of (49). We prove that if f is an exact symplectic map, then the norm of $\operatorname{avg}\{p_y\}_\theta$ is quadratic with respect to the norm $\|e\|_\rho$, and this will be enough to guarantee the existence of an approximate solution of (45) and therefore we will obtain an approximate solution of (21) (see Lemma 12).

Lemma 11. *If f is exact symplectic, then the following equality holds*

$$\operatorname{avg}\{p_y\}_\theta = -\operatorname{avg}\left\{\varphi(\theta) \;+\; De(\theta)^\top \, Da(K(\theta+\omega))\, e(\theta)\right\}_\theta ,$$

where e is defined in (22), a is defined by $\alpha = a\, dz$, and

$$\varphi(\theta) = (Df(K(\theta))DK(\theta))^\top \; \phi(\theta)$$

$$\phi(\theta) = a(f(K(\theta))) \;-\; a(K(\theta+\omega)) \;-\; Da(K(\theta+\omega))\, e(\theta). \tag{53}$$

Moreover, the following estimate holds

$$\left|\operatorname{avg}\{p_y\}_\theta\right| \le c_5\, \rho^{-1}\; \|e\|_\rho^2 , \tag{54}$$

where c_5 is a constant depending on n, ρ, $\|DK\|_\rho$, $|f|_{C^1,\mathcal{B}_r}$, and $|a|_{C^2,\mathcal{B}_r}$.

Proof. First of all notice that

$$p_x(\theta) = -T(\theta+\omega)\, e(\theta) , \qquad p_y(\theta) = -DK(\theta+\omega)^\top J(K(\theta+\omega))\, e(\theta), \tag{55}$$

where T is defined in (41). On the other hand, from (8) we have

$$\Omega_z(\zeta,\eta) = \sum_{i=1}^{2n}\sum_{j=1}^{2n} J_{i,j}(z)\, \zeta_i \eta_j , \qquad \forall \zeta, \eta \in \mathbb{R}^{2n}.$$

Then, if $\Omega = d\alpha$, with $\alpha = a\, dz$, for $i,j = 1,\ldots,2n$, we have

$$J_{i,j}(z) = \frac{\partial a_j(z)}{\partial z_i} - \frac{\partial a_i(z)}{\partial z_j} ,$$

equivalently

$$J(z) = Da(z)^\top - Da(z). \tag{56}$$

Using equality (56) we obtain

$$\begin{aligned} DK(\theta+\omega)^\top J(K(\theta+\omega))\, e(\theta) = {} & \underbrace{DK(\theta+\omega)^\top\, Da(K(\theta+\omega))^\top\, f(K(\theta))}_{(1)} - \\ & \underbrace{DK(\theta+\omega)^\top\, Da(K(\theta+\omega))^\top\, K(\theta+\omega)}_{(2)} - \\ & \underbrace{DK(\theta+\omega)^\top\, Da(K(\theta+\omega))\, e(\theta)}_{(3)} . \end{aligned}$$

Note that

$$(1) = \nabla\left[a(K(\theta+\omega))\cdot f(K(\theta))\right] - DK(\theta)^\top Df(K(\theta))^\top a(K(\theta+\omega)),$$

and

$$(2) = \nabla\left[a(K(\theta+\omega))\cdot K(\theta+\omega)\right] - DK(\theta+\omega)^\top a(K(\theta+\omega))\,.$$

Then

$$\operatorname{avg}\{\,(1)-(2)\,\}_\theta = -\operatorname{avg}\left\{\,DK(\theta)^\top Df(K(\theta))^\top a(K(\theta+\omega)) - DK(\theta)^\top a(K(\theta))\,\right\}_\theta .$$

Moreover, if f is exact symplectic the following equality holds:

$$(f\circ K)^*\,\alpha = K^*\alpha + d\,(b\circ K)\,.$$

Expressing this in coordinates and integrating we have

$$\operatorname{avg}\left\{DK(\theta)^\top a(K(\theta))\right\}_\theta = \operatorname{avg}\Big\{\,(Df(K(\theta))\,DK(\theta))^\top\, a(f(K(\theta)))\,\Big\}_\theta , \tag{57}$$

Equality (57) implies (see (53))

$$\begin{aligned} \operatorname{avg}\{(1)-(2)\}_\theta &= \operatorname{avg}\Big\{(Df(K(\theta))\,DK(\theta))^\top\,(a(f(K(\theta))) - a(K(\theta+\omega)))\Big\}_\theta \\ &= \operatorname{avg}\Big\{(Df(K(\theta))\,DK(\theta))^\top\, Da(K(\theta+\omega))\,e(\theta)\,\Big\}_\theta + \\ &\qquad + \operatorname{avg}\Big\{(Df(K(\theta))\,DK(\theta))^\top\,\phi(\theta)\Big\}_\theta . \end{aligned}$$

Therefore,

$$\begin{aligned} \operatorname{avg}\{\,(1)-(2)-(3)\,\}_\theta = {} & \operatorname{avg}\Big\{(Df(K(\theta))\,DK(\theta))^\top\,\phi(\theta)\,\Big\}_\theta + \\ & \operatorname{avg}\left\{\,De(\theta)^\top\, Da(K(\theta+\omega))\,e(\theta)\,\right\}_\theta , \end{aligned}$$

where we have used that $Df(K(\theta))\,DK(\theta) - DK(\theta+\omega) = De(\theta)$.

Cauchy's inequalities and the properties of the norms, we have the following estimate for $\left|\mathrm{avg}\,\{p_y\}_\theta\right|$

$$\left|\mathrm{avg}\,\{p_y\}_\theta\right| \le c_5\,\rho^{-1}\,\|e\|_\rho^2\,,$$

where c_5 is a constant depending on n, ρ, $\|DK\|_\rho$, $|f|_{C^1,\mathcal{B}_r}$, and $|a|_{C^2,\mathcal{B}_r}$. □

From Proposition 9, Proposition 10, and equation (54), we have that the linearized equation

$$[DF(K)\Delta]\,(\theta) = -e(\theta)$$

is approximately solvable in the following sense.

Lemma 12. *Assume that* $\omega \in D(\gamma,\sigma)$, f *is an exact symplectic map,* S *(defined in (29)) satisfies that* $\mathrm{avg}\,\{S\}_\theta$ *is non-singular, and that* $\|e\|_\rho$ *is small enough such that Proposition 9 applies.*

Let M *be defined by (31), and* ξ *be the solution of*

$$\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} \xi(\theta) - \xi(\theta+\omega) = \begin{pmatrix} p_x \\ p_y - \mathrm{avg}\,\{p_y\}_\theta \end{pmatrix},$$

The mapping $\Delta(\theta) = M(\theta)\xi(\theta)$ *satisfies*

$$[DF(K)\Delta]\,(\theta) + e(\theta) = M(\theta+\omega)\,[\,B(\theta)\,\xi(\theta) - w(\theta) - q\,]\,, \tag{58}$$

where

$$q \overset{\mathrm{def}}{=} \begin{pmatrix} 0 \\ \mathrm{avg}\,\{p_y\}_\theta \end{pmatrix}, \tag{59}$$

and B *and* w *are defined in equation (46). Moreover, the following estimates hold*

$$\|\Delta\|_{\rho-2\delta} \le c\,\gamma^{-2}\,\delta^{-2\sigma}\,\|e\|_\rho\,, \tag{60}$$

and

$$\|DF(K)\Delta + e\|_{\rho-2\delta} \le c\,\gamma^{-3}\,\delta^{-(3\sigma+1)}\,\|e\|_\rho^2\,, \tag{61}$$

where c *is a constant depending on* σ, n, ρ, $|a|_{C^2,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$, $|f|_{C^1,\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N\|_\rho$, $\left|(\mathrm{avg}\,\{S\}_\theta)^{-1}\right|$.

Proof. Equality (58) follows from Proposition 9 and the definition of Δ. From inequality (47) and Proposition 10 one obtains

$$\|M(\theta+\omega)\,B(\theta)\,\xi(\theta)\,\|_{\rho-2\delta} \le c\,\gamma^{-3}\,\delta^{-(3\sigma+1)}\,\|e\|_\rho^2\,,$$

for some constant c. On the other hand, inequality (48) implies

$$\|M(\theta+\omega)\,w(\theta)\,\|_{\rho-2\delta} \le c\,\gamma^{-1}\,\delta^{-(\sigma+1)}\,\|e\|_\rho^2\,,$$

and from inequality (54) we obtain

$$\|M(\theta+\omega)\,q\|_{\rho-2\delta} \leq c\,\gamma^{-1}\,\rho^{-1}\,\|e\|_{\rho}^{2}\,.$$

Therefore,

$$\|M(\theta+\omega)\,[\,B(\theta)\,\xi(\theta)w(\theta)+q\,]\|_{\rho} \leq c\,\gamma^{-3}\delta^{-(3\sigma+1)}\,\|e\|_{\rho}^{2}\,,$$

where we have redefined c. □

5 A modified Newton method

In this section we will assume that f, K_0, ω, and ρ_0 satisfy the hypotheses of Theorem 1. We apply the results of Section 4 to construct a modified Newton method. We will show that if $\|e_0\|_{\rho_0}$ is small enough, the procedure can be iterated indefinitely and it converges to a solution which satisfies the bound claimed in Theorem 1. In this way we construct a convergent sequence of approximate solutions for (1), and the limit function will be an exact solution of (1).

5.1 Description of the method

Starting with the approximate solution K_0 of the equation (1), we define the sequence

$$K_m = K_{m-1} + \Delta K_{m-1}\,, \qquad m \geq 1,$$

where ΔK_{m-1} is the approximate solution, produce in Lemma 12, of the following linear equation

$$DF(K_{m-1})\Delta K_{m-1} = -\,e_{m-1},$$

$$e_{m-1}(\theta) = F(K_{m-1})(\theta). \tag{62}$$

We emphasize that the symplectic map f is never modified in the Newton method described here.

The following result formulates the estimates that make precise that this quasi-Newton step improves the solution.

As it is standard in KAM theory, we will obtain that the norm of error, in a slightly smaller domain, is bounded by the quadratic norm of the original error multiplied by a power of the domain loss.

Lemma 13. *Assume that $K_{m-1} \in \mathcal{P}_{\rho_{m-1}}$ is an approximate solution of (1), satisfying conditions* **N1-N2***, and that the following holds*

$$r_{m-1} \overset{\text{def}}{=} \|K_{m-1} - K_0\|_{\rho_{m-1}} < r\,. \tag{63}$$

If e_{m-1}, defined in (62), is small enough such that Lemma 12 applies Then there exists a function $\Delta K_m \in \mathcal{P}_{\rho_{m-1}-3\delta_{m-1}}$, for any $0<\delta_{m-1}<\rho_{m-1}/3$, such that

$$\begin{aligned} \|\Delta K_{m-1}\|_{\rho_{m-1}-2\delta_{m-1}} &< c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-2\sigma}\ \|e_{m-1}\|_{\rho_{m-1}}\,, \\ \|D\,\Delta K_{m-1}\|_{\rho_{m-1}-3\delta_{m-1}} &< c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(2\sigma+1)}\ \|e_{m-1}\|_{\rho_{m-1}}\,, \end{aligned} \tag{64}$$

where c_{m-1} is a constant depending on σ, n $|f|_{C^1,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$, $\|DK_{m-1}\|_\rho$, $\|N_{m-1}\|_{\rho_{m-1}}$, and $\left|(\mathrm{avg}\,\{S_{m-1}\}_\theta)^{-1}\right|$.

Moreover, if $K_m = K_{m-1}+\Delta_{m-1}$, and

$$r_{m-1} + c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-2\sigma}\ \|e_{m-1}\|_{\rho_{m-1}} < r\,, \tag{65}$$

then we can redefine c_{m-1}, it will depend on $|f|_{C^2,\mathcal{B}_r}$ and all the previous quantities, such that the error function $e_m(\theta) = F(K_m)(\theta)$ satisfies

$$\|e_m\|_{\rho_m} \le c_{m-1}\gamma^{-4}\,\delta_{m-1}^{-4\sigma}\ \|e_{m-1}\|^2_{\rho_{m-1}}\,. \tag{66}$$

Proof. Let $\Delta_{m-1}(\theta) = M_{m-1}(\theta)\,\xi_{m-1}(\theta)$ be the approximate solution, provided by Lemma 12, of

$$DF(K_{m-1})\Delta_{m-1} = -\,e_{m-1}\,.$$

Then, estimates (64) follow from (60) and Cauchy's inequalities.

Moreover if we define $K_m = K_{m-1}+\Delta_{m-1}$, then

$$\|K_m - K_0\|_{\rho_{m-1}-2\delta_m} \le r_{m-1} + c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-2\sigma}\ \|e_{m-1}\|_{\rho_{m-1}} < r\,.$$

Hence, if (65) holds $K_m(\theta)\in\mathcal{B}_r$ for any θ with $|\mathrm{Im}\ \theta| < \rho_{m-1}-2\delta_{m-1}$. Define the remainder of Taylor expansion

$$\mathcal{R}(K,v) = F(v) - F(K) - DF(K)\,(v\,-\,K)\,,$$

then

$$e_m(\theta) = [\,e_{m-1} + DF(K)\Delta_{m-1}]\,(\theta)\,+\,\mathcal{R}(K_{m-1},K_m)(\theta)\,.$$

Therefore, estimate (66) follows from Taylor's Theorem and estimate (61). □

5.2 The non-degeneracy condition

In this section we will inductively show that if the norm $\|e_0\|_{\rho_0}$ is small enough and the first step of the method described in Section 5.1 is possible, then K_m, for $m\ge 1$, satisfies the non-degeneracy conditions **N1-N2**.

Throughout this section we will assume that $K_{m-1}\in\mathcal{P}_{\rho_{m-1}}$ satisfies conditions **N1-N2** (see Definition 2), and that inequalities (63) and (65) hold.

Since conditions **N1-N2** involves the derivative of K_m, which norm is estimated in the complex strip $U_{\rho_{m-1}-3\delta_{m-1}}$, for any $0<\delta_{m-1}<\rho_{m-1}/3$, (see Lemma 13). We define $\rho_m = \rho_{m-1}-3\delta_{m-1}$, for $0<\delta_{m-1}<\rho_{m-1}/3$ fixed.

Lemma 14. *Assume that the hypotheses of Lemma 13 hold. If* $\|e_{m-1}\|_{\rho_{m-1}}$ *is small enough (see inequality (67)), then*

1. *If* $DK_{m-1}(\theta)^\top DK_{m-1}(\theta)$ *is invertible, with inverse* N_{m-1}, *then* $DK_m(\theta)^\top DK_m(\theta)$ *is invertible, and the inverse* N_m *satisfies*
$$\|N_m\|_{\rho_m} \leq \|N_{m-1}\|_{\rho_{m-1}} + \tilde{c}_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(2\sigma+1)}\;\|e_{m-1}\|_{\rho_{m-1}}\,,$$

2. *Lest* S_{m-1} *and* S_m *be defined by replacing* K *with* K_{m-1} *and* K_m, *respectively, in (29). If* $\operatorname{avg}\{S_{m-1}\}_\theta$ *is invertible then,* $\operatorname{avg}\{S_m\}_\theta$ *is also invertible. Moreover, the following inequality holds*
$$\left|(\operatorname{avg}\{S_m\}_\theta)^{-1}\right| < \left|(\operatorname{avg}\{S_{m-1}\}_\theta)^{-1}\right| + \hat{c}_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(2\sigma+1)}\;\|e_{m-1}\|_{\rho_{m-1}}\,,$$

where $\tilde{c}_{m-1}$, *and* $\hat{c}_{m-1}$ *depend on same quantities as* c_{m-1} *in Lemma 13,* $|f|_{C^2,\mathcal{B}_r}$, *and* $\left|J^{-1}\right|_{C^1,\mathcal{B}_r}$.

Proof. Performing some computations we have

$$\begin{aligned} S_m &= S_{m-1} + \Phi_{m-1}\,,\\ DK_m(\theta)^\top DK_m(\theta) &= N_{m-1}(\theta)^{-1} + Z_{m-1}(\theta)\,, \end{aligned}$$

where Z_{m-1} satisfies

$$\begin{aligned} \|Z_{m-1}(\theta)\|_{\rho_{m-1}} &\leq 2\,\|DK_{m-1}\|_{\rho_{m-1}}\,\|D\,\Delta K_{m-1}\|_{\rho_{m-1}}\\ &\qquad + \|D\,\Delta K_{m-1}\|_{\rho_{m-1}}\,\|\Delta K_{m-1}\|_{\rho_{m-1}}\\ &\leq c_{m-1}\delta_{m-1}^{-(2\sigma+1)}\;\|e_{m-1}\|_{\rho_{m-1}}\,, \end{aligned}$$

where we have used estimates given in (64) and we have ignored the quadratic terms at the price of redefining c_{m-1}. Similarly, we have

$$\|\Phi_{m-1}\|_{\rho_{m-1}} \leq c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(2\sigma+1)}\;\|e_{m-1}\|_{\rho_{m-1}}\,.$$

Therefore, if

$$b_{m-1} \stackrel{\text{def}}{=} \max\left(\left|(\operatorname{avg}\{S_{m-1}\}_\theta)^{-1}\right|\,,\|\,N_{m-1}\,\|\right)$$

and e_{m-1} satisfies

$$b_{m-1}\,c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(2\sigma+1)}\;\|e_{m-1}\|_{\rho_{m-1}} < 1/2\,, \tag{67}$$

then, we have that the matrices

$$\left(I_n + (\operatorname{avg}\{S_{m-1}\}_\theta)^{-1}\operatorname{avg}\{\Phi_{m-1}\}_\theta\right)\,,\qquad \text{and}\qquad (I_n + N_{m-1}Z_{m-1})\,,$$

are invertible. Hence, Neuman's series theorem implies that the matrices

$$DK_m(\theta)^\top DK_m(\theta) = N_{m-1}(\theta)^{-1}\left[I_n + N_{m-1}(\theta)Z(\theta)\right]$$

$$\operatorname{avg}\{S_m\}_\theta = \operatorname{avg}\{S_{m-1}\}_\theta\left[I_n + (\operatorname{avg}\{S_{m-1}\}_\theta)^{-1}\operatorname{avg}\{\Phi_{m-1}\}_\theta\right]$$

are invertible. ☐

Remark 15. *For typographical reasons we redefine constant c_{m-1}, given in Lemma 13, to be*

$$\max\left(\tilde{c}_{m-1}, \hat{c}_{m-1}, b_{m-1}c_{m-1}\right) .$$

5.3 Convergence

The convergence of the modified Newton method described in Section 5.1 is standard in KAM theory. We will present full details for the sake of completeness.

We have seen that if $K_{m-1}\left(U_{\rho_{m-1}}\right) \subset \mathcal{B}_r$ and $\|e_{m-1}\|_{\rho_{m-1}}$ is small enough, then $K_m \in \mathcal{P}_{\rho_m}$ and it satisfies conditions **N1-N2** . More precisely, there exists a suitable constant (see lemmas 13, and 14), such that if $\|e_{m-1}\|_{\rho_{m-1}}$ satisfies

$$\begin{aligned} r_{m-1} + c_{m-1}\,\gamma^{-2}\delta_{m-1}^{-2\sigma}\,\|e_{m-1}\|_{\rho_{m-1}} &< r\,, \\ c_{m-1}\,\gamma^{-2}\,\delta_{m-1}^{-(\sigma+1)}\,\|e_{m-1}\|_{\rho_{m-1}} &\le 1/2\,, \end{aligned} \tag{68}$$

then, one step of the modified Newton method produces a mapping $K_m \in \mathcal{P}_{\rho_m}$ which satisfies conditions **N1-N2** and the corresponding error function e_m satisfies the quadratic estimate (66), (see Lemma 13).

We will establish conditions on $\|e_0\|_{\rho_0}$ that guarantee that inequalities (68) hold for all $m \ge 1$. Then we will prove that the Newton method described in Section 5.1 is quadratically convergent and we obtain bounds for $\|K_0 - K_\infty\|_{\rho_\infty}$. The obtained estimates are similar to those in [Zeh75], hence, the argument follows similar lines.

Remark 16. *In principle, the constant c_{m-1} depends on the parameters of the problem n, σ, γ, $|f|_{C^2,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $\left|J^{-1}\right|_{C^1,\mathcal{B}_r}$. It also depends on $\rho_{m-1} \le \rho_0$, and on the following quantities, related to the approximation K_{m-1},*

$$d_m \overset{\text{def}}{=} \|DK_{m-1}\|_{\rho_{m-1}}\,, \quad \nu_m \overset{\text{def}}{=} \|N_{m-1}\|_{\rho_{m-1}}\,, \quad \tau_m \overset{\text{def}}{=} \left|(\operatorname{avg}\{S_{m-1}\}_\theta)^{-1}\right| .$$

Moreover, it is possible to see that the dependence of c_n on (d_m, ν_m, τ_m) is polynomial. That is there exists a polynomial, $\lambda(y_1, y_2, y_3)$, with positive coefficients depending on the parameters of the problem, and such that

$$c_{m-1} = \lambda\left(d_{m-1}, \nu_{m-1}, \tau_{m-1}\right)\,, \qquad \forall m \ge 1. \tag{69}$$

Since we are looking for a solution of (1) which is close to K_0*, it is natural to expect that the quantities* d_m*,* ν_m*, and* τ_m *to vary not more than a small quantity from* d_0*,* ν_0*, and* τ_0*. If this is the case, the constant* c_{m-1} *would be bounded*

$$\lambda\,(d_0+\beta, \nu_0+\beta, \tau_0+\beta\,)\,,$$

where β *will be conveniently determined.*

Lemma 15. *Let* $\{c_m\}_{m\geq 0}$ *be the sequence of positive numbers given in (69). For a fixed* $0<\delta_0<\min(\rho_0/12,1)$ *define*

$$\delta_m \overset{\text{def}}{=} \delta_0\, 2^{-m}, \quad \varepsilon_m \overset{\text{def}}{=} \|e_m\|_{\rho_m}\,, \quad r_m \overset{\text{def}}{=} \|K_m-K_0\|_{\rho_m}\,, \qquad m\geq 0.$$

There exists a constant c*, depending on* σ*,* n*,* r*,* ρ_0*,* $|f|_{C^2,\mathcal{B}_r}$*,* $|a|_{C^2,\mathcal{B}_r}$*,* $|J|_{C^1,\mathcal{B}_r}$*,* $\left|J^{-1}\right|_{C^1,\mathcal{B}_r}$*,* $\|DK_0\|_{\rho_0}$*,* $\|N_0\|_{\rho_0}$*,* $\left|(\operatorname{avg}\{S_0\}_\theta)^{-1}\right|$*, such that if* $\|e_0\|_{\rho_0}$ *satisfies the following inequalities*

$$\kappa \overset{\text{def}}{=} 2^{4\sigma}\, c\,\gamma^{-4}\,\delta_0^{-4\sigma}\,\|e_0\|_{\rho_0} \leq 1/2, \tag{70}$$

$$\left[1+\frac{2^{4\sigma}}{2^{2\sigma}-1}\right] c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0} < r\,, \tag{71}$$

then, the modified Newton step can be iterated indefinitely and it converges to a mapping $K_\infty \in \mathcal{P}_{\rho_0-6\delta_0}$*, which satisfies the non-degenerate conditions* **N1** *and* **N2** *(Definition 2), and*

$$f\circ K_\infty = K_\infty \circ T_\omega.$$

Moreover, the following holds

$$\|K_\infty - K_0\|_{\rho_0-6\delta_0} \leq \left[\frac{2^{2\sigma}}{2^{2\sigma}-1}\right] c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\|e_0\|_{\rho_0}\,. \tag{72}$$

Proof. In order to prove that the modified Newton described in Section 5.1 can be iterated indefinitely it is enough to prove that conditions (70) and (71) imply conditions (68) for all $m\geq 0$. Define

$$c \overset{\text{def}}{=} \lambda\,(d_0+\beta, \nu_0+\beta, \tau_0+\beta\,)\,, \tag{73}$$

with

$$\beta \overset{\text{def}}{=} \gamma^2\,\delta_0^{2\sigma-1}\,2^{-(4\sigma+1)}\left(1+2^{4\sigma-1}\right)\,.$$

We prove, using induction, that for all integer $m\geq 0$ the following holds

C1(m) $r_m \leq \left[1+\kappa\,\frac{2^{4\sigma}}{2^{2\sigma}-1}\right] c\,\gamma^{-2}\,\delta_0^{-2\sigma}\,\varepsilon_0 < r.$

C2(m) $\varepsilon_m \leq 2^{-4\sigma(m-1)}\,\kappa^{(2^m-1)}\,\varepsilon_0,$

C3(m) $c_m \le c$,

C4(m) $c_m\, \gamma^{-2}\, \delta_m^{-(2\sigma+1)}\, \varepsilon_m \le 1/2$.

Notice that this implies that conditions (68) hold for all $m \ge 1$.

C1(0) and **C2(0)** are immediate. Note that from the definition $c_0 \le c$, i.e. **C3(0)** holds, and since we are assuming $\sigma > 1$ and $0 < \gamma \le 1$, inequality (70) implies

$$c_0\, \gamma^{-2} \delta_0^{-(2\sigma+1)}\, \varepsilon_0 \le \left(c_0\, \gamma^{-4}\, \delta_0^{-4\sigma}\, \varepsilon_0\right) \delta_0^{2\sigma-1} < 2^{-4\sigma} < 1/2$$

this proves **C4(0)**.

Assume that **C1(j)-C4(j)** hold for $j = 1, \dots, m-1$ (then it is possible to perform $m-1$ steps of the modified Newton method described in Section 5.1). In particular, for $j = 0, \dots, m-1 \ge 0$, c_j is bounded by c, then for $j = 1, \dots, m$, one has the following estimate of the error (see (66)):

$$\begin{aligned}
\varepsilon_j \quad &\le \quad c\,\gamma^{-4} \delta_{j-1}^{-4\sigma}\, \varepsilon_{j-1}^2 \le \left(c\,\gamma^{-4}\right)^{(1+2)} \left(\delta_{j-1}\, \delta_{j-2}^{2}\right)^{-4\sigma}\, \varepsilon_{j-2}^4 \le \dots \\
\dots \quad &\le \quad \left(c\,\gamma^{-4}\right)^{1+2+\dots+2^{j-1}} \left(\delta_{j-1}\, \delta_{j-2}^2 \dots\, \delta_0^{2^{j-1}}\right)^{-4\sigma} \varepsilon_0^{2^j} \\
&= \quad \left(c\,\gamma^{-4}\, \delta_0^{-4\sigma}\right)^{1+2+\dots+2^{j-1}} \left(2^{4\sigma}\right)^{2^0 (j-1)+2(j-2)+\dots+2^{j-2}} \varepsilon_0^{2^j} \\
&\le \quad \left(c\,\gamma^{-4}\, \delta_0^{-4\sigma}\right)^{2^j-1} 2^{4\sigma\,(2^j-j)}\, \varepsilon_0^{2^j} \\
&\le \quad \Bigg(\underbrace{c\,\gamma^{-4}\, \delta_0^{-4\sigma} 2^{4\sigma} \varepsilon_0}_{\kappa}\Bigg)^{2^j-1} 2^{-4\sigma(j-1)}\, \varepsilon_0 \le \kappa^{2^j-1} \left(2^{-4\sigma(j-1)}\right) \varepsilon_0\,,
\end{aligned}$$

where we have used that $2^0(j-1)+2(j-2)+\dots+2^{j-2} = 2^{j-1} \sum_{s=1}^{j-1} s\, 2^{-s} \le 2^j - j$.

Let us prove that **C1(m)-C4(m)** hold. We first prove **C1(m)**:

$$\begin{aligned}
r_m &\le r_{m-1} + c_{m-1}\, \gamma^{-2}\, \delta_{m-1}^{-2\sigma}\, \|e_{m-1}\|_{\rho_{m-1}} \le \dots \le c\,\gamma^{-2}\, \delta_0^{-2\sigma} \varepsilon_0 + c\,\gamma^{-2} \sum_{j=1}^{m-1} \delta_j^{-2\sigma} \varepsilon_j \\
&\le c\,\gamma^{-2}\, \delta_0^{-2\sigma} \varepsilon_0 + c\,\gamma^{-2}\, \delta_0^{-2\sigma}\, \kappa\, \varepsilon_0 \sum_{j=1}^{m-1} 2^{2j\sigma}\, 2^{-4\sigma(j-1)} \\
&\le c\,\gamma^{-2}\, \delta_0^{-2\sigma} \varepsilon_0 \left(1 + \kappa\, 2^{4\sigma} \sum_{j=1}^{\infty} 2^{-2j\sigma}\right) = c\,\gamma^{-2}\, \delta_0^{-2\sigma} \varepsilon_0 \left(1 + \kappa\, \frac{2^{4\sigma}}{2^{2\sigma}-1}\right)
\end{aligned}$$

and since $\kappa \le 1/2$, (71) implies **C1(m)**.

Therefore, inequalities (66) and (70) imply

$$\varepsilon_j \; < \; \kappa^{2^j-1} \varepsilon_0 \, 2^{-4\sigma(j-1)} , \qquad \text{for}\, j = 1, \ldots, m, \tag{74}$$

this proves **C2(m)**.

We now prove **C3(m)**. The first inequality in (64) and Lemma 14 imply

$$\begin{aligned} d_m &\le d_{m-1} \, + \, c_{m-1} \, \gamma^{-2} \, \delta_{m-1}^{-(2\sigma+1)} \varepsilon_{m-1} \le d_0 \, + \, s_{m-1} \, , \\ \nu_m &\le \nu_{m-1} + c_{m-1} \, \gamma^{-2} \, \delta_{m-1}^{-(2\sigma+1)} \varepsilon_{m-1} \le \nu_0 + s_{m-1} \, , \\ \tau_m &\le \tau_{m-1} + c_{m-1} \, \gamma^{-2} \, \delta_{m-1}^{-(2\sigma+1)} \varepsilon_{m-1} \le \tau_0 + s_{m-1} \, , \end{aligned}$$

where

$$s_{m-1} \overset{\text{def}}{=} \sum_{j=0}^{m-1} c_j \, \gamma^{-2} \, \delta_j^{-(2\sigma+1)} \, \varepsilon_j .$$

Let us estimate s_{m-1}. Inequality (74) and **C3(j)** imply, for $j = 1, \ldots, m-1$,

$$c_j \, \gamma^{-2} \, \delta_j^{-(2\sigma+1)} \, \varepsilon_j \; \le \left(c \, \gamma^{-2} \, \delta_0^{-(2\sigma+1)} \, \varepsilon_0 \right) 2^{4\sigma} \kappa^{2^j-1} \, 2^{-j(2\sigma-1)} \, ,$$

then, using inequality (70) one has

$$\begin{aligned} s_{m-1} \quad &\le \quad c \, \gamma^{-2} \, \delta_0^{-(2\sigma+1)} \, \varepsilon_0 \left(1 + \kappa 2^{4\sigma} \sum_{j=1}^{\infty} 2^{-j(2\sigma-1)} \right) \\ &= \quad \kappa \gamma^2 \, \delta_0^{2\sigma-1} \, 2^{-4\sigma} \left(1 + \kappa \, \frac{2^{4\sigma}}{2^{2\sigma-1} - 1} \right) \\ &\le \quad \gamma^2 \, \delta_0^{2\sigma-1} \, 2^{-(4\sigma+1)} \left(1 + 2^{4\sigma-1} \right) = \kappa_1 \end{aligned}$$

Then

$$d_m \le d_0 + \beta \, , \qquad \nu_m \le \nu_0 + \beta \, , \qquad \tau_m \le \tau_0 + \beta \, ,$$

this implies $c_m \le c$, i.e. **C3(m)** holds. Form (74) and (70) we have

$$\begin{aligned} c_m \, \gamma^{-2} \, \delta_m^{-(2\sigma+1)} \, \varepsilon_m \quad &\le \quad \left(c \, \gamma^{-2} \, \delta_0^{-(2\sigma+1)} \, \varepsilon_0 \right) 2^{4\sigma} \kappa^{(2^m-1)} \, 2^{-m(2\sigma-1)} \\ &\le \quad \kappa \, \gamma^2 \, \delta_0^{2\sigma-1} \, \kappa^{2^m-1} \, 2^{-m(2\sigma-1)} \le 1/2 \, . \end{aligned}$$

Hence, we have proved **C j(m)** for $j = 1, 2, 3, 4$.

Therefore, the modified Newton method described in Section 5 produces a sequence of mappings K_m satisfying Lemma 13. In particular, from the first inequality in (64) we have that $\{K_m\}_{m \ge 0}$ is a Cauchy sequence in the scale of Banach spaces $\mathcal{P}_{\rho_m}$. Let $K_\infty = \lim_{m \to \infty} K_m$, then $K_\infty \in \mathcal{P}_{\rho_\infty}$ (see Definition 2), with $\rho_\infty \overset{\text{def}}{=} \lim_{m\to\infty} \rho_m = \rho_0 - 6\delta_0$,. Moreover inequality (66) implies that K_∞ satisfies (1). Finally, estimate (72) follows from first inequality in (64) and

$$\|K_\infty - K_0\|_{\rho_\infty} \leq \sum_{m=0}^{\infty} \|\Delta K_m\|_{\rho_\infty} \leq \sum_{m=0}^{\infty} \|\Delta K_m\|_{\rho_m} .$$

□

6 Local uniqueness

In this section, we prove Theorem 2. The proof is rather standard. It suffices to show that the operator $DF(K)$ has a approximate left inverse (see [Zeh75]).

In our context the existence of the approximate left inverse amounts to the uniqueness statement of Proposition 10 and Remark 14. The uniqueness up to additive constants in these results amount to the uniqueness of re-parameterization of K (see Remark 5).

From now on we assume that K_1 and K_2 satisfy the hypotheses of Theorem 2. We note that if $F(K_1) = F(K_2) = 0$, by Taylor's theorem we have

$$0 = F(K_1) - F(K_2) = DF(K_2)(K_1 - K_2) + \mathcal{R}(K_1, K_2)\,, \tag{75}$$

where

$$\|\mathcal{R}\|_\rho \leq c\, \|K_1 - K_2\|_\rho^2 .$$

Since K_2 is an exact solution of (1), the results of Section 4.3 apply and Proposition 9 implies that the equation (75) can be transformed into

$$\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} \Delta(\theta) - \Delta(\theta + \omega) = -\, M^{-1}(\theta+\omega)\, \mathcal{R}(K_1, K_2)(\theta)$$

where S is defined by (29), replacing K by K_2, and

$$\Delta = M(\theta)^{-1}(K_1 - K_2)(\theta) = \begin{pmatrix} T(\theta) \\ DK_2(\theta)^\top J(K_2(\theta)) \end{pmatrix} (K_1 - K_2)(\theta),$$

with T defined in (41) (replacing K by K_2).

By the uniqueness statement in Proposition 10 and Remark 14 we obtain that

$$\left\| \Delta - (\operatorname{avg}\{\Delta_x\}_\theta\,, 0)^\top \right\|_{\rho - 2\delta} \leq c\,\gamma^{-2}\delta^{-2\sigma}\, \|\mathcal{R}\|_\rho^2 \leq c\,\gamma^{-2}\delta^{-2\sigma}\, \|K_1 - K_2\|_\rho^2 \,,$$

where

$$\operatorname{avg}\{\Delta_x\}_\theta = \operatorname{avg}\{T(\theta)\,[K_1 - K_2]\,(\theta)\}_\theta .$$

Lemma 16. *There exists a constant* $\tilde{c}$, *depending on* n, ρ^{-1}, $|J|_{\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$, $\|N\|_\rho$, $\|K_1\|_{C^2,\rho}$, *such that if*

$$\tilde{c}\,\|K_1 - K_2\|_\rho \leq 1\,,$$

then there exists an initial phase $\tau_1 \in \Theta \overset{\mathrm{def}}{=} \left\{\tau \in \mathbb{R}^n : |\tau| < \|K_1 - K_2\|_\rho\right\}$ *such that,*

$$\operatorname{avg}\left\{T(\theta)\left[\,K_1 \circ T_{\tau_1} - K_2\right](\theta)\right\}_\theta = 0\,. \tag{76}$$

Therefore, for any $0 < \delta < \rho/2$ *we have*

$$\|K_1 \circ T_{\tau_1} - K_2\,\|_{\rho - 2\delta} < \hat{c}\,\gamma^{-2}\delta^{-2\sigma}\,\|K_1 - K_2\|_\rho^2\,, \tag{77}$$

for some constant $\hat{c}$ *depending on* n, σ, γ, ρ, ρ^{-1}, $|f|_{C^2,\mathcal{B}_r}$, $|J|_{\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$ $\|DK_2\|_\rho$, $\|N\|_\rho$, $\left|(\operatorname{avg}\{S\}_\theta)^{-1}\right|$, *with* N, *and* S *as in Definition 2, replacing* K *by* K_2.

Proof. Some direct computations (see (41)) show that

$$T(\theta)DK_2(\theta) = I_n\,,$$

then, for $x \in \mathbb{R}^n$

$$T(\theta)\left[K_1(\theta + x) - K_2(\theta)\right] = T(\theta)\left[K_1(\theta + x) - K_2(\theta)\right] - T(\theta)DK_2(\theta)\,x + x\,.$$

Therefore, a solution τ_1 of (76), is a fix point of

$$\Psi(x) \overset{\mathrm{def}}{=} -\operatorname{avg}\left\{T(\theta)\left[K_1(\theta + x) - K_2(\theta)\right] - T(\theta)DK_2(\theta)\,x\right\}_\theta$$

Performing some computations we see that for any $|x|\,, |y| < \|K_1 - K_2\|_\rho$ (i.e. $x, y \in \Theta$), the following holds

$$\begin{aligned}|\Psi(y) - \Psi(x)| &\leq c_1\,|y - x|^2 + c_2\,|x|\,|y - x| + c_3\,\|D(K_1 - K_2)\|_0\,|y - x| \\ &\leq c_4\,\|K_1 - K_2\|_\rho\,|y - x|\end{aligned}$$

where the constant c_4 depends on n, ρ^{-1}, $\|K_2\|_{C^2,\rho}$, and $\|T\|_\rho$ (i.e. $|J|_{\mathcal{B}_r}$, $\left|J^{-1}\right|_{\mathcal{B}_r}$, $\|DK_2\|_\rho$, and $\|N\|_\rho$, see (41)). Therefore, if $\tilde{c} = c_4$, and

$$\tilde{c}\,\|K_1 - K_2\|_\rho < 1,$$

then the mapping $\Psi : \Theta \to \Theta$ is a contraction.

Let $|\tau_1| < \|K_1 - K_2\|_\rho$ satisfy (76), then $K_1 \circ T_{\tau_1}$ is a solution of (1) such that if

$$\Delta(\theta) = M(\theta)^{-1}\left(K_1 \circ T_{\tau_1} - K_2\right)(\theta)\,,$$

then the uniqueness statement in Proposition 10, Remark 14, and equality (76) imply

$$\|\,\Delta\|_{\rho - 2\delta} \leq c\,\gamma^{-2}\delta^{-2\sigma}\,\|\mathcal{R}\|_\rho^2 \leq c\,\gamma^{-2}\delta^{-2\sigma}\,\|K_1 - K_2\|_\rho^2\,,$$

Therefore, we have

$$\begin{aligned}\|K_1 \circ T_{\tau_1} - K_2\|_{\rho-2\delta} &\le \|M\|_\rho \, \|\Delta\|_{\rho-2\delta} \\ &\le \hat c\,\gamma^{-2}\delta^{-2\sigma}\,\|K_1 - K_2\|_\rho^2\end{aligned}$$

for any $0 < \delta, \rho/2$. ☐

Let τ_1 be as in Lemma 16, since $K_1 \circ T_{\tau_1}$ is also a solution of (1), we can apply Lemma 16 to the solutions $K_1 \circ T_{\tau_1}$ and K_2 (notice that we do not change K_2, this implies in particular, that $\tilde c$ and $\hat c$ do not change). Hence, we obtain a sequence $\{\tau_m\}_{m\ge 1}$ such that

$$|\tau_m - \tau_{m-1}| < \left\|K_1 \circ T_{\tau_{m-1}} - K_2\right\|_{\rho_{m-1}} ,$$

and

$$\begin{aligned}\|K_1 \circ T_{\tau_m} - K_2\|_{\rho_m} &\le c\,\gamma^{-2}\,\delta_m^{-2\sigma}\,\left\|K_1 \circ T_{\tau_{m-1}} - K_2\right\|_{\rho_{m-1}}^2 \\ \dots &\le \left(\hat c\,\gamma^{-2}\right)^{2^m}\left[\delta_m\delta_{m-1}^2\delta_{m-2}^{2^2}\dots\delta_1^{2^{m-1}}\right]^{-2\sigma}\,\|K_1 - K_2\|_\rho^{2^m} \\ &\le \left(\hat c\,\gamma^{-2}\,\delta_1^{-2\sigma}\,2^{2\sigma}\,\|K_1 - K_2\|_\rho\right)^{2^m}\;2^{-2\sigma(j-1)} ,\end{aligned}$$

where $\delta_1 = \rho/8$, $\delta_{m+1} = \frac{\delta_m}{2}$, and $\rho_m = \rho - \sum_{j=1}^m \delta_j$ for $m \ge 1$.

Therefore, if

$$c = \max\left(\tilde c,\ \hat c\,2^{2\sigma}\right) ,$$

and

$$\gamma^{-2}\,\delta_1^{-2\sigma} c\,\|K_1 - K_2\|_\rho < 1$$

then the sequence $\{\tau_m\}$ converges, and its limit τ_∞ satisfies

$$\|K_1 \circ T_{\tau_\infty} - K_2\|_{\rho/2} = 0\,.$$

7 Proof of Theorem 3

We formulate a general step of the procedure, the iterative procedure and the argument for the convergence is identical to that presented in Section 5.

7.1 A general step of a modified Newton method

Let $K \in \tilde{\mathcal{P}}_\rho$ be an approximate solution of (15) with error function

$$e = f_\lambda \circ K - K \circ T_\omega \tag{78}$$

and assume that K and f_λ satisfy Definition 3. In order to construct a new approximate solution by using a modified Newton method, we look for an increment of the parameter $\Delta\lambda$ and an increment function ΔK which is 1-periodic in each variable and such that the elements of order 1 of $G(K + \Delta K, \lambda + \Delta\lambda)$ are equal to zero, with G defined in (20). This leads to study the solvability properties of the linearized equation:

$$\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right) \Delta\lambda + Df(K(\theta))\Delta K(\theta) - \Delta K(\theta+\omega) \;=\; -\,e(\theta)\,. \tag{79}$$

Notice that, except for the term $\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda$, the right hand part of (79) has the same form as the linear operator $DF(K)$. Hence, the geometric procedure described in Section 4, will work in this case with small modifications.

We emphasize that the geometric properties of an approximate solution of (1) in Section 4.1 and the results on the reducibility of the linearized equation in sections 4.2 and 4.3 follow from the symplectic properties of f and the size of the error function, then the results of sections 4.1, 4.2, and 4.3 also hold for the symplectic map f_λ. The main difference is that the approximate solvability, proved in Section 4.4, was achieved using the exactness character of f and the non-degeneracy condition **N2** (see Proposition 10 and Lemma 11). In the present case this will be achieved by using the increment of the parameter $\Delta\lambda$.

The approximate solution of the transformed equation will produce an approximate solution of the linear equation (79) with quadratic error.

In order to ensure that the procedure can be iterated, we have to verify that the correction obtained by adding to K the approximate solution of the linearized equation (79) and the symplectic map $f_{\lambda+\Delta\lambda}$ satisfy the same conditions as K and f_λ (this part is identical to Section 5.2).

Proposition 9, applied to the map f_λ and K, implies that with the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, with M defined by (31) replacing f by f_λ, transforms (79) into

$$\left[\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} + B(\theta)\right]\xi(\theta) - \xi(\theta) = p(\theta) + w(\theta) + M(\theta+\omega)^{-1}\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda\,, \tag{80}$$

where S is defined in (29) replacing f by f_λ,

$$p(\theta) = -\,V(\theta+\omega)^{-1} M(\theta+\omega)^\top J(K(\theta+\omega))\,e(\theta)\,, \tag{81}$$

where V^{-1} is given in (40), and B and w satisfy estimates

$$\|\,B(\theta)\|_{\rho-2\delta} \;\le\; c\,\gamma^{-1}\,\delta^{-(\sigma+1)}\;\|e\|_\rho\;,$$
$$\|w\|_{\rho-2\delta} \;\le\; c\,\gamma^{-1}\,\delta^{-(\sigma+1)}\;\|\,e\|_\rho^2$$

for some constant $c > 0$.

Moreover, Lemma 8 implies

$$M(\theta+\omega)^{-1}\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda \;= \Lambda(\theta)\Delta\lambda\; + q(\theta)$$

with

$$\Lambda(\theta) \overset{\mathrm{def}}{=} V(\theta+\omega)^{-1}M(\theta+\omega)^\top J(K(\theta+\omega))\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right), \tag{82}$$

and

$$\|q\|_{\rho-2\delta} \;\leq c\,\gamma^{-1}\delta^{-(\sigma+1)}\left\|\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right\|_\rho |\Delta\lambda|\;\|e\|_\rho$$

Let us prove that the reduced equation obtained by removing the terms B, w, and q from equation (80) can be solved for $(\xi, \Delta\lambda)$.

Proposition 17. *Assume that* $\omega \in D(\gamma,\sigma)$ *and that* K *and* f_λ *satisfy Definition 3. If* $\|e\|_\rho$ *is small enough then there exists a mapping* ξ*, analytic on* $U_{\rho-2\delta}$*, and a vector* $\Delta\lambda \in \mathbb{R}^{2n}$*, satisfying*

$$\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix}\xi(\theta) - \xi(\theta) = \;R(\theta)\,, \tag{83}$$

where

$$R(\theta) \overset{\mathrm{def}}{=} -\;V(\theta+\omega)^{-1}M(\theta+\omega)^\top J(K(\theta+\omega))\left[e(\theta) + \left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda\right]. \tag{84}$$

Moreover, there exists a constant c*, depending on* n*,* σ*,* ρ*,* r*,* $|f|_{C^2,\mathcal{B}_r}$*,* $\|DK\|_\rho$*,* $\|N(\theta)\|_\rho$*,* $\left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho$*, and* $\left|(\mathrm{avg}\,\{\Lambda\}_\theta)^{-1}\right|$ *(see (82)) such that the following inequalities hold*

$$\|\,\xi\|_{\rho-2\delta} \leq\; c\,\gamma^{-2}\,\delta^{-2\sigma}\;\|\,e\|_\rho\;, \tag{85}$$

$$|\Delta\lambda| \leq\; c\;\left|(\mathrm{avg}\,\{\Lambda(\theta)\}_\theta)^{-1}\right|\,\|\,e\,\|_\rho\;, \tag{86}$$

Proof. Consider the symplectically conjugated coordinates $R = (R_x, R_y)$ and $\xi = (\xi_x, \xi_y)$, then equation (83) can be written as

$$\begin{aligned} \xi_x(\theta) - \xi_x(\theta+\omega) &= R_x(\theta) - S(\theta)\,\xi_y(\theta), \\ \xi_y(\theta) - \xi_y(\theta+\omega) &= R_y(\theta). \end{aligned}$$

From (84) one can see that

$$R(\theta) = \begin{pmatrix} -T(\theta+\omega)e(\theta) \\ -DK(\theta+\omega)^\top J(K(\theta+\omega))\,e(\theta) \end{pmatrix} + \Lambda(\theta)\,\Delta\lambda\,, \tag{87}$$

with Λ given in (82).

Since f_λ and K satisfy Definition 3 we have that $\operatorname{avg}\{\Lambda\}_\theta$ has maximum range (i.e. $2n$). Therefore we can fix the last n-coordinates of $\Delta\lambda \in \mathbb{R}^{2n}$ such that

$$\operatorname{avg}\{R_y\}_\theta = 0\,.$$

Proposition 1 implies that there exists unique zero average analytic mapping on $U_{\rho-\delta}$ satisfying

$$\xi_y(\theta) - \xi_y(\theta+\omega) = R_y(\theta)\,,$$

and satisfying the estimate

$$\|\xi_y\|_{\rho-\delta} \le c\,\gamma^{-1}\,\delta^{-\sigma}\,\|R_y\|_\rho\,. \tag{88}$$

Now we let $\Delta\lambda$ be completely determinate by the equation

$$\operatorname{avg}\{R_x(\theta) - S(\theta)\,\xi_y(\theta)\}_\theta = 0\,,$$

By Proposition 1 there exists a unique mapping ξ_x satisfying

$$\xi_x(\theta) - \xi_x(\theta+\omega) = p_x(\theta) - S(\theta)\,\xi_y(\theta)\,,$$

with

$$\|\xi_x\|_{\rho-2\delta} \le c\,\gamma^{-1}\,\delta^{-\sigma}\,\|R_x(\theta) - S(\theta)\,\xi_y(\theta)\|_{\rho-\delta}\,. \tag{89}$$

Notice that we defined $\Delta\lambda \in \mathbb{R}^{2n}$ to be the unique solution of

$$\operatorname{avg}\{\Lambda(\theta)\}_\theta\,\Delta\lambda = \begin{pmatrix} \operatorname{avg}\{-T(\theta+\omega)e(\theta) - S(\theta)\,\xi_y(\theta)\}_\theta \\ \operatorname{avg}\left\{-DK(\theta+\omega)^\top J(K(\theta+\omega))\,e(\theta)\right\}_\theta \end{pmatrix} \tag{90}$$

Let $(\varphi_x,\varphi_y)^\top = \operatorname{avg}\{\Lambda(\theta)\}_\theta\,\Delta\lambda$, then, form (87) and (90) we have the estimates:

$$\|R_y\|_\rho \le c\left(\|e\|_\rho + \left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho |\varphi_y|\right) \le \tilde{c}\,\|e\|_\rho \tag{91}$$

and

$$\begin{aligned} \|R_x\|_\rho &\le c\left(\|e\|_\rho + \left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho |\varphi_x|\right) \\ &\le \tilde{c}\left(\|e\|_\rho + \left|\operatorname{avg}\{S(\theta)\,\xi_y(\theta)\}_\theta\right|\right) \\ &\le \tilde{c}\left(\|e\|_\rho + c'\gamma^{-1}\rho^{-1}\,\|R_y\|_\rho\right). \end{aligned} \tag{92}$$

Inequalities (88), (89), (91), (92) and equality (90) imply estimates (85) and (86). □

The construction of an approximate solution of the linearized equation (79) is now clear, we establish this in the following result, the prove follows from the previous construction and it is very similar to the prove of Lemma 12.

Proposition 18. *Assume that the hypotheses of Proposition 17 hold. Let* ξ *and* $\Delta\lambda$ *be as in Proposition 17, define* $\Delta K(\theta) \overset{\text{def}}{=} M(\theta)\,\xi(\theta)$*. The pair* $(\Delta K, \Delta\lambda)$ *is an approximate solution of the linearized equation (79) such that*

$$\|D\,G(K,\lambda)\,(\Delta K,\Delta\lambda)\;+\;G(u,\lambda)\,\|_{\rho-2\delta} \le c\;\delta^{-3\sigma+1}\;\|e\|_\rho^2\;,$$

for some constant c*, depending on* n*,* σ*,* r*,* $|f|_{C^1,\mathcal{B}_r}$*,* $\|DK\|_\rho$*,* $\|N(\theta)\|_\rho$*,* $\left|(\operatorname{avg}\{\Lambda\}_\theta)^{-1}\right|$ *(see (82)), and* $\left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho$*.*

Remark 17. *In the case that* f_λ *is exact symplectic we proved in Section 4.4 that (see Lemma 11)*

$$\left|\operatorname{avg}\left\{DK(\theta+\omega)^\top J\,e(\theta)\right\}_\theta\right| \le\; c\,\rho^{-1}\;\|e\|_\rho^2\;,$$

then, if $\Delta\lambda = (\Delta\lambda_x, \Delta\lambda_y)^\top$*, we have*

$$|\Delta\lambda_y| \le c'\;\left|(\operatorname{avg}\{\Lambda\}_\theta)^{-1}\right|\;\left|\operatorname{avg}\left\{DK(\theta+\omega)^\top J\,e(\theta)\right\}_\theta\right|\;\le c''\,\|e\|_\rho^2\;.$$

Moreover, if K *satisfies the non-degeneracy condition* **N2** *stated in Definition 2, i.e. the* $n\times n$ *matrix* $\operatorname{avg}\{S(\theta)\}_\theta$*, with* S *defined by (29), replacing* f *by* f_λ*, is non-singular, then we can choose* $\operatorname{avg}\{\xi_y\}_\theta$ *in Proposition 17 such that*

$$\operatorname{avg}\left\{-T(\theta+\omega)\,e(\theta) - S(\theta)\,\tilde{\xi}\right\}_\theta\;,$$

then $\Delta\lambda_y = 0$ *(see (90)).*

The iterative procedure

Let ΔK and $\Delta\lambda$ as in Proposition 18, we consider $K+\Delta K$ as an approximate of solution of the equation

$$f_{\lambda+\Delta\lambda}\circ Z = Z\circ T_\omega$$

Since K is analytic on U_ρ and the solution of (83) is analytic on $U_{\rho-2\delta}$ for any $0<\delta<\rho/2$, $K+\Delta K$ is analytic on $U_{\rho-2\delta}$. Moreover, inequalities (85) and (85) imply

$$\|\Delta K\|_{\rho-2\delta} \le c\,\delta^{-2\sigma}\;\|e\|_\rho\;,\qquad\qquad |\Delta\lambda| \le c\;\|e\|_\rho \tag{93}$$

and by using Cauchy's inequality one has

$$\|D\Delta K\|_{\rho-3\delta} \le c\,\delta^{-(2\sigma+1)}\;\|e\|_\rho\;. \tag{94}$$

One step of the modified Newton method is obtained by proving that $K+\Delta K$ and $f_{\lambda+\Delta\lambda}$ satisfy the same conditions as K and f_λ.

Lemma 19. *Assume that* $\lambda \in \mathbb{R}^n$, $K \in \tilde{\mathcal{P}}$, *and that* f_λ *and* K *satisfy Definition 3. If* $\|e\|_\rho$ *is small enough, then* $K + \Delta K$ *and* $f_{\lambda+\Delta}$ *satisfy Definition 2. Moreover, if* Λ *is defined by (82), then the following inequality hold*

$$|\operatorname{avg}\{\Lambda_1(\theta)\}_\theta| \le |\operatorname{avg}\{\Lambda(\theta)\}_\theta| + c\,\gamma^{-2}\delta^{-(2\sigma+1)}\,\|e\|_\rho\;,$$

where Λ_1 *is defined by (82) by replacing* f_λ *and* K *by* $f_{\lambda+\Delta\lambda}$ *and* $K + \Delta K$, *respectively, and* c *is a constant depending on* n, σ, r, $|f|_{C^2,\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N(\theta)\|_\rho$, $\left\|\left(\frac{\partial^2 f_\lambda(K(\theta))}{\partial^2\lambda}\right)\right\|_\rho$, *and* $\left|(\operatorname{avg}\{\Lambda\}_\theta)^{-1}\right|$.

Proof. The proof is very similar to the proof of Lemma 14 in Section 5.2, it follows for the assumption that the family f_λ is twice differentiable with respect to λ, the estimations given in (93) and (94) and the Neuman's series theorem ☐

Since the estimate and the induction hypothesis have the same structure as those in Section 5, the convergence argument and the final estimates do not need any modifications.

8 The linear equation for vector fields

Let ω be a fixed frequency vector satisfying the Diophantine condition (9), and let K be an approximate solution of (16). We study the invertibility properties of the linear operator

$$J\,D\,(\nabla H(K))\,\Delta - \partial_\omega\,\Delta. \tag{95}$$

Lagrangian character

It is known that if K is a solution of (16) with ω is rationally independent, then the torus $K(\mathbb{T}^n)$ is a Lagrangian manifold. In the case that K satisfies (16) only approximately, we will prove that $K(\mathbb{T}^n)$ is approximate Lagrangian.

Lemma 20. *Let* $K : \mathbb{T}^n \to \mathbb{T}^n$ *be a real analytic mapping on the complex strip of width* $\rho > 0$. *Define the error function*

$$e(\theta) = J\nabla H(K(\theta)) - \partial_\omega\;K(\theta). \tag{96}$$

The following hold.

a. The components of the vector $DK(\theta)^\top J\,e(\theta)$ *are zero average functions.*

b. If $e = 0$, *i.e.* K *satisfies*

$$\partial_\omega K = J\nabla H \circ K\,, \tag{97}$$

then there exists a function $b : \mathbb{T}^n \to \mathbb{R}$ *and a constant vector* $a_0 \in \mathbb{R}^n$ *such that*

$$K^*\alpha = d\,b + a_0\,d\theta.$$

c. *There exist two functions* $\tilde{b} : \mathbb{T}^n :\to \mathbb{R}$ *and* $g : \mathbb{T}^n \to \mathbb{R}^n$, *and a constant vector* $a_0 \in \mathbb{R}^n$ *such that*

$$K^*\alpha = d\,\tilde{b} + a_0\,d\theta + \sum_{j=1}^{n} g_j(\theta) d\theta_j .$$

Moreover, the function g *satisfies*

$$\partial_\omega\ g(\theta) = DK(\theta)^\top J\,e(\theta)\,. \tag{98}$$

d. *Let* $L(\theta)$ *be defined by (23). There exists a constant* c, *depending on* n, σ *and* $|DK|_\rho$ *such that*

$$|L|_{\rho-2\delta} \leq c\,\gamma^{-1}\delta^{-(\sigma+1)}\,\|e\,\|_\rho . \tag{99}$$

In particular, $L = 0$ *if* $e = 0$.

Proof. The pull-back of $\alpha = -y\,dx$ under $K = (w, v)^\top$ is given by

$$K^*\alpha = -v\,dw = -\sum_{j=1}^{n} \frac{\partial w(\theta)}{\partial \theta_j}^\top v(\theta)\,d\theta_j .$$

For $j = 1, \ldots n$, define $a_j(\theta) = \frac{\partial w(\theta)}{\partial \theta_j}^\top v(\theta)$, then

$$\begin{aligned}\partial_\omega\ a_j(\theta) &= \left(\partial_\omega\ \frac{\partial w(\theta)}{\partial \theta_j}\right)^\top v(\theta) + \frac{\partial w(\theta)}{\partial \theta_j}^\top \partial_\omega\ v(\theta)\\ &= \left[e_x(\theta)\cdot\frac{\partial v(\theta)}{\partial \theta_j} - e_y(\theta)\cdot\frac{\partial w(\theta)}{\partial \theta_j}\right] +\end{aligned}$$

$$\frac{\partial}{\partial \theta_j}\left[\underbrace{D_yH(K(\theta))v(\theta) + (H\circ K)(\theta)}_{\hbar_1} - \underbrace{e_x(\theta)\cdot v(\theta)}_{\hbar_2}\right] .$$

Notice that since $\partial_\omega\ a_j(\theta)$, $\frac{\partial}{\partial \theta_j}\hbar_1$, and $\frac{\partial}{\partial \theta_j}\hbar_2$ have average equal to zero, then

$$\operatorname{avg}\left\{e_x(\theta)\cdot\frac{\partial v(\theta)}{\partial \theta_j} - e_y(\theta)\cdot\frac{\partial w(\theta)}{\partial \theta_j}\right\}_\theta = 0,$$

this proves part (a) of Lemma 20. Moreover (c.f. [Rüs75]) since ω is Diophantine there exist functions $g : \mathbb{T}^n \to \mathbb{R}^n$ and $b_1, b_2 : \mathbb{T}^n \to \mathbb{R}$ such that, g satisfies (98), and

$$\partial_\omega\ b_1 = \hbar_1\ - \operatorname{avg}\{\hbar_1\}_\theta\,, \qquad \partial_\omega\ b_2 = \hbar_2\ - \operatorname{avg}\{\hbar_2\}_\theta\ .$$

If K is a solution of (97), then g and b_2 are constants, setting $g = 0$ and $b_2 = 0$, we have

$$K^*\alpha = -\sum_{j=1}^{n} a_j(\theta)\, d\theta_j = -\sum_j \frac{\partial}{\partial\theta_j} b_{1,j}(\theta)\, d\theta_j - \text{avg}\,\{a_j\}_\theta\, d\theta_j\,,$$

this proves part (b) of Lemma. Taking $\tilde{b} = -(b_1 - b_2)$ part (c) of Lemma 20 follows.

Let us prove part (d). From part (c) we have

$$K^*\,\Omega\; = d\,K^*\alpha = d\,(-g\,d\theta)\,,$$

in coordinate representation one has

$$L(\theta) = Dg(\theta) - Dg(\theta)^\top$$

therefore, part (d) of Lemma 20 follows from equality (98). $\square$

Change of variables in the linearized equation

As in the case of analytic exact symplectic maps, the geometric structure of K makes possible to define a change of variables which transforms the linear operator (95) into another that it is simpler one (compare with the case of maps in Section 4.3). Let A be defined by

$$A(\theta) \overset{\text{def}}{=} \begin{pmatrix} D_x\nabla_y H(K(\theta)) & D_y\nabla_y H(K(\theta)) \\ -D_x\nabla_x H(K(\theta)) & -D_y\nabla_x H(K(\theta)) \end{pmatrix} = J\,D\,(\nabla H)\,(K(\theta))\,. \tag{100}$$

Then the linearized equation

$$J\,D\,(\nabla H(K))\,\Delta - \partial_\omega\,\Delta = -\,e$$

takes the form

$$A(\theta)\,\Delta(\theta)\; - \;\partial_\omega\;\Delta(\theta) = -e(\theta). \tag{101}$$

If we perform the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, in the variables ξ the linearized equation (101) has the following expression

$$[\,A(\theta)M(\theta) - \partial_\omega\;M(\theta)\,]\;\xi(\theta) - M(\theta)\;\partial_\omega\;\xi(\theta) = -\,e(\theta)$$

where $A(\theta)$ is defined in equality (100). Similarly to the case of exact symplectic maps, we give a change of variables M such that the matrix

$$M(\theta)^{-1}\,[\,A(\theta)M(\theta) - \partial_\omega\;M(\theta)\,] \tag{102}$$

is an upper triangular matrix with diagonal terms equal to zero.

Assume that there exists a matrix N such that

$$N(\theta) = \left(DK(\theta)^\top DK(\theta)\right)^{-1}, \tag{103}$$

and consider the $2n \times 2n$ matrix valued function

$$M(\theta) = \begin{pmatrix} DK(\theta) & JDK(\theta)N(\theta) \end{pmatrix}. \tag{104}$$

Lemma 21. *If K is an approximate solution of (97), with error e, then if M is defined by (104) we have*

$$A(\theta)\, M(\theta) - \partial_\omega\ M(\theta) = M(\theta) \begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix} + E(\theta) \tag{105}$$

with

$$S(\theta) = N(\theta) DK(\theta)^\top \left[A(\theta)\, J - JA(\theta)\,\right] DK(\theta)N(\theta)\,, \tag{106}$$

and

$$\left\| JM(\theta)^\top JE \right\|_{\rho-2\delta} \leq c\,\gamma^{-1}\delta^{-(\sigma+1)} \left\| e \right\|_\rho\ , \tag{107}$$

where $0 < \delta < \rho/2$ and c is a constant depending on n, ρ, $\|A\|_\rho$, $\|K\|_\rho$, and $\|N\|_\rho$.

Proof. This is the analogue to the results for maps presented in Section 4.2. Since the case when K is an approximate solution of (16) involves a lot of (but simple) computations which could hide the main idea we only consider the case when K is a solution of (16).

Let us compute $W(\theta) = A(\theta)M(\theta) - \partial_\omega\ M(\theta)$ in two steps: (a) compute the first n-columns, (b) compute the last n-columns.

First note that since K satisfies (16)

$$A(\theta)\, DK(\theta) = \partial_\omega\ DK(\theta) \tag{108}$$

this gives that the first n columns of $W(\theta)$ are equal to zero.

The last n columns of $W(\theta)$ are given by

$$W_1(\theta) = A(\theta)\, JDK(\theta)N(\theta) - \partial_\omega\ (JDK(\theta)N(\theta))\ .$$

The Hamiltonian character: $A(\theta)\, J = -J\, A(\theta)^\top$, and equality (108) imply

$$W_1(\theta) = -J\, \left[A(\theta)^\top + A(\theta)\right]\, DK(\theta)N(\theta)\ -\ J\, DK(\theta)\ \partial_\omega\ N(\theta).$$

Moreover, by using equality (108), it is not difficult to see that

$$\partial_\omega\ N(\theta) = -N(\theta)DK(\theta)^\top \left[\, A(\theta)^\top + A(\theta)\,\right] DK(\theta)N(\theta)\,.$$

Since K satisfies (16) the vectors $\{\frac{\partial K(\theta)}{\partial\theta_j}, J\,\frac{\partial K(\theta)}{\partial\theta_j}\}_{j=1}^n$ form a basis of $\mathbf{T}_\theta\mathbf{U} \sim \mathbb{R}^{2n}$, then we can find $n \times n$ matrix-valued functions $S(\theta)$ and $T(\theta)$, such that

$$W_1(\theta) = DK(\theta)S(\theta) + J\, DK(\theta)\, N(\theta)\, T(\theta) \tag{109}$$

multiplying the last inequality by $-DK(\theta)^\top J$, and performing some simple substitutions one obtains

$$T(\theta) = -DK(\theta)^\top J W_1(\theta) = 0\,.$$

Moreover, multiplying by $N(\theta)DK(\theta)^\top$ equality (109) we have

$$\begin{aligned} S(\theta) &= N(\theta)DK(\theta)^\top \left[A(\theta)\, JDK(\theta)N(\theta) - J\, \partial_\omega\, (DK(\theta)N(\theta)) \right] \\ &= N(\theta)DK(\theta)^\top\, \left[A(\theta)\, J - J\, A(\theta) \right]\, DK(\theta)N(\theta), \end{aligned}$$

where we have used the two following equalities

$$DK(\theta)^\top JDK(\theta) = 0\,, \qquad \partial_\omega\, DK(\theta) = A(\theta)DK(\theta)\,.$$

Therefore, if S has the expression given in (106), and K is a solution of (16), we have

$$A(\theta)\, M(\theta) - \partial_\omega\, M(\theta) = M(\theta) \begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix}.$$

The case when K is an approximate solution of (16) with error e follows defining

$$E(\theta) = A(\theta)\, M(\theta) - \partial_\omega\, M(\theta) - M(\theta) \begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix}.$$

Performing some computations and using the equality

$$A(\theta)DK(\theta) - \partial_\omega DK(\theta) = De(\theta)$$

we obtain equality (105) and estimate (107). □

Now we prove that the matrix M defined in (104) is invertible. From the definition of M it follows that

$$M(\theta)^\top J M(\theta) = -J + R(\theta),$$

with

$$R(\theta) = \begin{pmatrix} L(\theta) & 0 \\ 0 & N^\top\, L(\theta) N(\theta) \end{pmatrix},$$

where $L = DK(\theta)^\top JDK(\theta)$.

The norm of matrix R is controlled by the approximate Lagrangian character of the torus K. In fact, from inequality (99), the norm of R verifies the following inequality

$$\|R\|_{\rho-2\delta} \leq c\, \gamma^{-1} \delta^{-(\sigma+1)}\, \|e\,\|_\rho\,.$$

Lemma 22. *If the error function e is small enough, the matrix M, given by (104), is invertible and the inverse is given by*

$$M(\theta)^{-1} = J\, M(\theta)^\top J + M_e(\theta),$$

with

$$\|M_e\|_{\rho-2\delta} \le c\,\gamma^{-1}\delta^{-(\sigma+1)}\,\|e\|_\rho,$$

for some constant c, and any $0<\delta<\rho$. In particular, if $e=0$, $M_e=0$.

When K solves approximately the equation (16), matrix M defined by (104) reduce (102) into a simple form, and the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$ will transform the linearized equation (95) into a more simple one (see (110)). The result can be summarized as follows.

Proposition 23. *The change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, with M defined in (104), transforms the linearized (101) into*

$$\left[\begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix} + B(\theta)\right] \xi(\theta) - \partial_\omega\, \xi(\theta) = p(\theta) + w(\theta) \tag{110}$$

where S defined in (106),

$$p(\theta) = -\,J\, M(\theta)^\top J\, e(\theta) = \begin{pmatrix} -N(\theta)^\top DK(\theta)^\top e(\theta) \\ DK(\theta)^\top J\, e(\theta) \end{pmatrix}. \tag{111}$$

and B and w satisfy

$$\begin{aligned} \|\,B\,\|_{\rho-2\delta} &\le c\,\gamma^{-1}\delta^{-(\sigma+1)}\|\,e\|_\rho, \\ \|\,w\,\|_{\rho-\delta} &\le c\,\gamma^{-1}\delta^{-(\sigma+1)}\,\|\,e\|_\rho^2\,, \end{aligned} \tag{112}$$

for any $0<\delta<\rho/2$, where c is a constant depending on n, $\|\,DK\|_\rho$, and $\|N\|_\rho$.

Proof. Since M is invertible, we have

$$\begin{aligned} A(\theta)M(\theta)\xi(\theta) - \partial_\omega\, (M(\theta)\xi(\theta)) &= M(\theta)\, \left[M(\theta)^{-1}\, W(\theta)\xi(\theta) - \partial_\omega\, \xi(\theta)\right] \\ &= M(\theta)\, \left[JM(\theta)^\top JW(\theta)\xi(\theta) + M_e W(\theta)\xi(\theta) - \partial_\omega\, \xi(\theta)\right]\,, \end{aligned}$$

with $W(\theta) = A(\theta)M(\theta) - \partial_\omega\, M(\theta)$. From Lemma 21 we have

$$M(\theta)^{-1}\, W(\theta) = \begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix} + M(\theta)^{-1}E(\theta)$$

Therefore, if we define

$$B(\theta) = M(\theta)^{-1}E(\theta) = JM(\theta)^\top J\, E(\theta) + M_e(\theta)\, E(\theta)$$

the estimate (107) and the estimate for the norm of M_e imply the first inequality of (112). Therefore, in the variables ξ, the linear equation (101) writes

$$[\,C(\theta) + B(\theta)\,]\,\xi(\theta) - \partial_\omega\, \xi(\theta) = -M(\theta)^{-1}\, e(\theta) = -JM(\theta)^\top J\, e(\theta) - M_e\, e(\theta).$$

□

Solvability of the reduced equation

Let us consider the reduced equation

$$\begin{aligned} S(\theta)\,\xi_y - \partial_\omega\,\xi_x &= -\,N(\theta)^\top\,DK(\theta)^\top\,e(\theta), \\ \partial_\omega\,\xi_y &= DK(\theta)^\top\,Je(\theta), \end{aligned} \tag{113}$$

The conditions guaranteeing the existence of a solution (ξ_x, ξ_y) of equation (113) are provided by the following result.

Lemma 24. *There exists a solution* (ξ_x, ξ_y) *of equation (113) satisfying the following estimate*

$$\begin{aligned} \|\,\xi_y\|_{\rho-\delta} &\leq c_1\,\gamma^{-1}\delta^{-\sigma}\,\|e\|_\rho\,, \\ \|\,\xi_x\|_{\rho-2\delta} &\leq c_2\,\gamma^{-2}\delta^{-2\sigma}\,\|e\|_\rho\,, \end{aligned}$$

for any any $0 < \delta < \rho/2$*, where* c_1 *and* c_2 *are constants depending on* n*,* σ*,* $\|N\|_\rho$*,* $\|K\|_\rho$*, and* $\left|(\operatorname{avg}\{S\}_\theta)^{-1}\right|$.

Proof. From part (c) of Lemma 20 $\operatorname{avg}\left\{DK(\theta)^\top\,Je(\theta)\right\}_\theta = 0$, then the small divisors equation

$$\partial_\omega\,\xi_y = DK(\theta)^\top\,Je(\theta),$$

has a solution $\tilde{\xi}_y$ with average equal to zero, c.f. [Rüs75]. Let $\bar{\xi}_y \in \mathbb{R}^n$ such that the function $\xi_y = \tilde{\xi}_y + \bar{\xi}_y$ satisfies

$$\operatorname{avg}\left\{S(\theta)\,\xi_y + N(\theta)^\top\,DK(\theta)^\top\,e(\theta)\right\}_\theta = 0.$$

Then, the small divisors equation

$$\partial_\omega\,\xi_x = S(\theta)\,\xi_y + N(\theta)^\top\,DK(\theta)^\top\,e(\theta)$$

has a unique solution ξ_x with average equal to zero. The estimations for ξ_x and ξ_y follow from [Rüs75]. ☐

Acknowledgements

The work of R. L. has been supported by NSF grants, a Deans Fellowship at U.T. Austin Spring 03. A. G was supported by a U.S.–Spain Science & Technology Program 2000, Project 96. À. J. and J. V. have also been supported by the Catalan grant 2001SGR-70, the Spanish grant BFM2000-0623, the INTAS grant 00-221 and DURSI.

References

[Bos86] J.-B. Bost. Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel,...). *Astérisque*, No. 133–134:113–157, 1986. Seminar Bourbaki, Vol. 1984/85.

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

[DdlLS03] A. Delshams, R. de la Llave, and T.M. Seara. A geometric mechanism for diffusion in hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model. *Preprint*, 2003.

[DG96] Amadeu Delshams and Pere Gutiérrez. Estimates on invariant tori near an elliptic equilibrium point of a Hamiltonian system. *J. Differential Equations*, 131(2):277–303, 1996.

[dlL01] Rafael de la Llave. A tutorial on KAM theory. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, pages 175–292. Amer. Math. Soc., Providence, RI, 2001.

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

[Dou82] Raphaël Douady. Une démonstration directe de l'équivalence des théorèmes de tores invariants pour difféomorphismes et champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 295(2):201–204, 1982.

[Her83] M.-R. Herman. *Sur les courbes invariantes par les difféomorphismes de l'anneau. Vol. 1*, volume 103 of *Astérisque*. Société Mathématique de France, Paris, 1983.

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

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

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

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

[Neĭ84] A. I. Neĭshtadt. The separation of motions in systems with rapidly rotating phase. *Prikl. Mat. Mekh.*, 48(2):197–204, 1984.

[Pös01] Jürgen Pöschel. A lecture on the classical KAM theorem. In *Smooth ergodic theory and its applications (Seattle, WA, 1999)*, pages 707–732. Amer. Math. Soc., Providence, RI, 2001.

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

[RC95] M. J. Raković and Shih-I Chu. New integrable systems: hydrogen atom in external fields. *Phys. D*, 81(3):271–279, 1995.

[RC97] M. J. Raković and Shih-I Chu. Phase-space structure of a new integrable system related to hydrogen atoms in external fields. *J. Phys. A*, 30(2):733–753, 1997.

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

[Rüs76a] H. Rüssmann. On a new proof of Moser's twist mapping theorem. In *Proceedings of the Fifth Conference on Mathematical Methods in Celestial Mechanics (Oberwolfach, 1975), Part I*, volume 14, pages 19–31, 1976.

[Rüs76b] H. Rüssmann. On optimal estimates for the solutions of linear difference equations on the circle. *Celestial Mech.*, 14(1):33–37, 1976.

[Rüs76c] Helmut Rüssmann. Note on sums containing small divisors. *Comm. Pure Appl. Math.*, 29(6):755–758, 1976.

[Rüs98] H. Rüssmann. Invariant tori in the perturbation theory of weakly nondegenerate integrable Hamiltonian systems. *Mainz preprint*, 14, 1998.

[Sal86] D. Salamon. The Kolmogorov-Arnold-Moser theorem. *Zürich preprint*, 1986.

[Sev96] M. B. Sevryuk. Invariant tori of Hamiltonian systems that are nondegenerate in the sense of Rüssmann. *Dokl. Akad. Nauk*, 346(5):590–593, 1996.

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

[Van02] J. Vano. *A Whitney-Zehnder implicit function theorem.* PhD thesis, University of Texas at Austin, 2002.

[Wei77] Alan Weinstein. *Lectures on symplectic manifolds.* American Mathematical Society, Providence, R.I., 1977. Expository lectures from the CBMS Regional Conference held at the University of North Carolina, March 8–12, 1976, Regional Conference Series in Mathematics, No. 29.

[WL98] Maciej P. Wojtkowski and Carlangelo Liverani. Conformally symplectic dynamics and symmetry of the Lyapunov spectrum. *Comm. Math. Phys.*, 194(1):47–60, 1998.

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

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