INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **18** (2005) 855–895 doi:10.1088/0951-7715/18/2/020

KAM theory without action-angle variables

R de la Llave[1], A González[1], À Jorba[2] and J Villanueva[3]

[1] Department of Mathematics, University of Texas at Austin, University Station C1200, Austin, TX 78712-0257, USA
[2] Dept. de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Barcelona, Spain
[3] Dept. de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, Barcelona, Spain

Received 19 July 2004, in final form 23 November 2004
Published 21 January 2005
Online at stacks.iop.org/Non/18/855

Recommended by V Baladi

Abstract
We give a proof of a KAM theorem on existence of invariant tori with a Diophantine rotation vector for Hamiltonian systems. The method of proof is based on the use of the geometric properties of Hamiltonian systems which, in particular, do not require the Hamiltonian system either to be written in action-angle variables or to be a perturbation of an integrable one. The proposed method is also useful to compute numerically invariant tori for Hamiltonian systems. We also prove a translated torus theorem in any number of degrees of freedom.

Mathematics Subject Classification: 37J40

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 for Hamiltonian systems (symplectic maps and Hamiltonian vector fields).

What we show is that, given an approximately invariant torus which is not too degenerate, there is a true invariant torus nearby. We will refer to such proofs as 'polishing'.

The proof is constructive and the KAM method presented here leads to an algorithm that can be implemented numerically and which uses very small requirements of memory and operation.

The methodology presented here does not require the system either to be a perturbation of an integrable system or to be written in action-angle variables. This is due to the fact that the method is based on the geometric properties of Hamiltonian systems which do not require such assumptions.

Of course, action-angle variables always exist in a neighbourhood of an invariant torus. However, a change of coordinates bringing the system to action-angle variables in general cannot be explicitly computed.

The guiding principle for the proof given here is the observation that the geometry of the problem implies that KAM tori are *reducible* and approximately invariant tori are *approximately reducible*—see sections 4.2 and 8. This leads to a solution of the linearized equations without transformation theory. Moreover, the reducing transformation is given explicitly in terms of the approximately invariant torus, which forms the basis of an efficient numerical algorithm.

The use of the geometry makes the method applicable to other situations where geometric conditions imply automatic reducibility. For example, in [HdlL04b] similar ideas in spirit have been developed to prove existence of normally hyperbolic manifolds of quasi-periodic mappings. The numerical results, obtained by the implementation of the algorithm yielded by the method of the proof, are presented in [HdlL04a].

The customary KAM statements for nearly integrable systems follow taking the invariant tori for the integrable system as approximately invariant tori for the perturbed system.

Presenting the KAM theory in a polished way—as is done in [Mos66a, Mos66b, Rüs76a, Zeh75, Zeh76, SZ89, Ran87, CC97]—has certain advantages. We mention just a few, which we will develop in future papers.

- In many practical applications, one has to consider systems that are not close to integrable but nevertheless one has some approximately invariant tori with small enough error—obtained, for example, by using a non-rigorous numerical method, some asymptotic expansion etc. Then, the results presented here imply that close to these approximately invariant tori there is a true invariant torus. That is, the existence of invariant tori is validated. In numerical analysis, theorems of this type are often called '*a posteriori estimates*'.
- From a more theoretical point of view, by approximating differentiable functions by analytic ones [Mos67, Zeh75, Sal04], the results presented here more or less automatically lead to results for finitely differentiable Hamiltonian systems.
- By taking the results of formal expansions as approximately invariant tori, one obtains differentiability with respect to parameters or with respect to the frequency. In particular, the families 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 (this is sometimes called '*condensation*').
- The smooth dependence on parameters leads to several other results such as estimates on the measure occupied by the invariant tori 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 action-angle variables at all. This seems to be useful in several situations. We just mention the following:

- 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.
- On several occasions the action-angle variables are much more singular than the original system (a recent remarkable example of a very smooth integrable system with complicated action-angle variables is [RC95, RC97]).

  Two important examples of this situation, from a theoretical point of view, are:
  - *Tori near a separatrix in a perturbed pendulum.* Near a separatrix, the orbits change from rotation to libration, which are topologically different. Hence, action-angle coordinates on each side are discontinuous. There are several applications in which it is useful to have tori close to a separatrix. In one degree of freedom,

see [Neĭ84, Her83]. The paper [DdlLS03] contains a detailed study of the singularity of the action-angle variables and shows that tori near a separatrix play an important role in mechanisms of instability.

– *Tori close to an elliptic point.* The action-angle variables are singular near an elliptic critical point. Since in the vicinity of an elliptic point, the approximation of invariant tori is related to the size of the actions, the methods based on action-angle variables cannot deal with regions in which one of the actions is much smaller than some of the others. Similar phenomena occur in the proof of the Nekhoroshev theorem. The papers [FGB98, GFB98] overcome the problem of degeneracy of action-angle developing a Nekhoroshev theory not based on action-angle. Other proofs based on other methods are [Nie98, Pös99].

  We hope to come to the problem of tori in the vicinity of an elliptic point in the near future.

Let us give a general picture of the results for symplectic maps—the Hamiltonian vector fields case is similar.

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 $\mathbf{U} \subset \mathbb{R}^{2n}$, 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.

Note that (1) implies that the range of K is invariant under f. The map K gives a parametrization 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 equation (1) approximately and which is not too degenerate (see definition 2), then, there is a true solution nearby.

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 the flow by S_t, 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, $\varphi_t = \eta t$ for some vector η. Since $\varphi_1 = \omega$, we conclude that $\eta = \omega$. Hence, from the equation of invariance for the time one map and the uniqueness we obtain 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 vice versa) can be found in [Dou82]. Note that the argument presented above applies to other types of invariant objects.

Nevertheless, in section 8, we will present a direct proof for flows along the same line as the proof for maps. This has a certain interest since the direct proofs presented here can be converted into algorithms. Moreover, we also formulate the results for vector fields precisely in section 2.4.

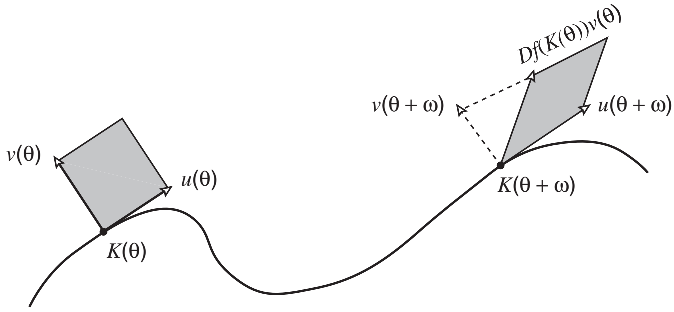

Figure 1. Geometric reason for automatic reducibility.

Reducibility of invariant tori. The method we present is based on a very simple geometric observation which has further applications. We use the geometry of the Hamiltonian system 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 in the symplectic maps case and equal to zero in the vector fields case. 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 in the symplectic maps case and equal to zero in the vector fields case. This is enough for a quasi-Newton method—see [Mos66a, Mos66b, Zeh75, Zeh76].

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

Let us consider the symplectic maps situation, the case of vector fields is similar. 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 1).

The fact that the map f preserves the 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 the invariant 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 equation (4) they become

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

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 make the averages zero.

Of course, there is little point in applying a 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 approximately satisfies (4). This is enough to get a quasi-Newton method.

Note that (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 maps 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, Mos88]. The Lagrangian automatic reducibility has the advantage that it also applies 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 the Lagrangian context, reducibility has also been 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 and proves 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-degenerate condition 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 procedure from 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 $\Omega = \mathrm{d}\alpha$ be an exact symplectic structure on $\mathbf{U}$, and let $a : \mathbf{U} \to \mathbb{R}^{2n}$ be defined by

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

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

$$\Omega_z(\xi, \eta) = \langle \xi, J(z)\eta \rangle, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ 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 $\mathbf{U}$, cf [Wei77]), that J is an almost complex structure, i.e. $J^2 = -\mathrm{Id}$, we will not make this assumption. Indeed, we have striven to deal with the problem in whatever coordinates it is presented.

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

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

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

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

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

$$\|K\|_\rho \stackrel{\text{def}}{=} \sup_{\theta \in U_\rho} |K(\theta)| < \infty,$$

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

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

Similar notation for the norm will also be 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

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

Note that it is equivalent to saying that K satisfies (10) as 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 $m \in \mathbb{Z}_+ \stackrel{\text{def}}{=} \mathbb{N} \cup \{0\}$ denote the C^m-norm of g on $\mathcal{B}$ by $|g|_{C^m,\mathcal{B}}$, i.e.

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

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 coordinates $(\partial g_i/\partial x_j)$. 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

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

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

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

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

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

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*

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

where I_n is the n-dimensional identity matrix.

N2. *The average of the matrix-valued function*

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

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 4.3. Note that **N1** depends only on K whereas **N2** depends on K and f.

2.2. Existence of an invariant torus for symplectic maps

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

Theorem 1. *Let $f : \mathbf{U} \to \mathbf{U}$ be an exact 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 neighbourhood of the image under K_0 of U_{ρ_0}:*

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

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

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

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

Define the error function e_0 by

$$e_0 \stackrel{\text{def}}{=} f \circ K_0 - K_0 \circ T_\omega. \tag{12}$$

There exists a constant $c > 0$ depending on σ, n, r, ρ_0, $|f|_{C^2,\mathcal{B}_r}$, $|a|_{C^2,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $|J^{-1}|_{C^1,\mathcal{B}_r}$, $\|DK_0\|_{\rho_0}$, $\|N_0\|_{\rho_0}$, $|(\mathrm{avg}\{S_0\}_\theta)^{-1}|$ (where N_0 and S_0 are as in definition 2, replacing K with K_0) such that, if $\|e_0\|_{\rho_0}$, defined in (12), verifies the following inequalities:

$$c\gamma^{-4}\delta_0^{-4\sigma}\|e_0\|_{\rho_0} < 1 \tag{13}$$

and

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

where $0 < \delta_0 \leqslant \min(1, \rho_0/12)$ 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,

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

Remark 3. We emphasize that theorem 1 does not assume either that the system is given in action-angle variables or that the maps are 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 11).

The quadratic estimates (see inequality (68)) contain the term γ^{-4}, this is the reason for the coefficient γ^{-4} in (13) which is the same as that obtained in some classical papers on KAM theory (e.g. [Mos66a, Mos66b, Zeh76]), but it is worse than the γ^{-2} obtained in some classical KAM theorems (e.g. [Pös01, Sal04]). 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 [Mos66a, Mos66b, Zeh76] since it is proportional to 4σ but is different from that of [Pös01, Sal04] which 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. Note that given a mapping K satisfying 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 parametrizations, 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 result states that the solutions obtained in theorem 1 are locally unique up to the irrelevant choice of origin of the phase. As we mentioned, remark 1 is a general argument that allows results for flows to be obtained 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) be two solutions of (1), 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 with 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 $U_{\rho/2}$.

2.3. A 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 let Ω be an exact symplectic structure on $\mathbf{U}$. In this section we will assume that K is a non-trivial embedding of the torus in the annulus rather than just a periodic function. We 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*

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

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

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

then 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 $\mathrm{avg}\{\Lambda\}_\theta$ *has a range of dimension* $2n$*.*

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 $f = f_0$. That is, $K(\theta + k) = K(\theta) + (k, 0)$ for $k \in \mathbb{Z}^n$. Assume that the pair f_0, K_0 satisfies definition 3, then we prove that there exists a solution of

$$f_\lambda \circ K = K \circ T_\omega, \tag{14}$$

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 and let* f_λ *be a* $2n$*-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 neighbourhood*

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

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*,* $|f_0|_{C^2, \mathcal{B}_r}$*,* $\|DK_0\|_{\rho_0}$*,* $\|(\partial f_\lambda / \partial \lambda)|_{\lambda = 0} K_0\|_{\rho_0}$*,* $\|N_0\|_{\rho_0}$ *and* $|(\mathrm{avg}\{\Lambda_0(\theta)\}_\theta)^{-1}|$ *(where* N_0 *and* Λ_0 *are as in definition 3, replacing* K *with* 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 $0 < \delta_0 \leqslant \min(1, \rho_0/12)$*, 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 $d \geqslant 2n$. Moreover, if the elements of the family are exact symplectic we can consider $d \geqslant n$.

In this case we also have 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}*,* $|f_{\hat{\lambda}}|_{C^2,\mathcal{B}_r}$*,* $|J|_{\mathcal{B}_r}$*,* $|J^{-1}|_{\mathcal{B}_r}$*,* $\|(\partial f_\lambda/\partial\lambda)|_{\lambda=0}K_2\|_\rho$*,* $\|K_2\|_{C^2,\rho}$*,* $\|N_2\|_\rho$ *and* $|(\mathrm{avg}\{\Lambda_2\}_\theta)^{-1}|$ *(with* N_2 *and* Λ_2 *as in definition 3, replacing* K *with* 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* $U_{\rho/2}$*.*

2.4. Existence of an invariant torus for 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

$$\partial_\omega K(\theta) = J\nabla H(K(\theta)), \tag{15}$$

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).

Note that (15) implies that the range of K is invariant under the Hamiltonian vector field with Hamiltonian function H, X_{H}. The map K gives a parametrization of the invariant torus which makes the dynamics of the vector field X_{H} restricted to the torus into a rigid rotation.

In order to simplify the notation we assume that $\Omega = \mathrm{d}x \wedge \mathrm{d}y$. 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 $\Omega = \mathrm{d}\alpha$.

Definition 4. *Given a Hamiltonian function* H *and a frequency vector* ω*, the mapping* $K : \mathbb{T}^n \to \mathbf{U}$ *is said 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*

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

– $\mathrm{avg}\{S\}_\theta$ *is invertible, with*

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

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 neighbourhood 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

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

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 $|(\mathrm{avg}\{S_0\}_\theta)^{-1}|$ (with N_0 and S_0 given by definition 4, replacing K with 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 (15), K_∞, which is real analytic on $U_{\rho/2}$ and satisfies the non-degenerate conditions in definition 4. Moreover,

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

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 $H_\lambda \stackrel{\text{def}}{=} H(\cdot, \lambda) : \mathbf{U} \to \mathbb{R}$ is a real analytic Hamiltonian function.

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

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

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*

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

– $\mathrm{avg}\{\Lambda_0\}_\theta$ *has maximum range, with*

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

Define the error function

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

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

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

and

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

2.5. 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 or close to integrable, we will present how the results here apply to the quasi-integrable case. We hope that this can help in understanding 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 Δ as a 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), (14)–(16) are functional equations for the unknown function K (and the unknown parameter λ in the case of (14) and (16)). As shown in [Mos66a, Zeh75], in order to prove the existence of a solution of a nonlinear 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 is customary in many proofs of the KAM theorem). This leads to 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.

3.1. Maps

The key point in finding a solution for (1) and (14) is to solve the corresponding linearized equation approximately. The main part of both linearized equations is the linear operator:

$$\mathcal{T}_{f,K,\omega}\Delta = Df(K(\theta))\Delta - \Delta \circ T_\omega. \tag{17}$$

Indeed, equations (1) and (14) can be formulated as finding zeros of

$$F(K) \stackrel{\text{def}}{=} f \circ K - K \circ T_\omega \tag{18}$$

and

$$G(K, \lambda) \stackrel{\text{def}}{=} f_\lambda \circ K - K \circ T_\omega, \tag{19}$$

respectively. Therefore, at least formally we have

$$DF(K)\Delta = \mathcal{T}_{f,K,\omega}\Delta,$$

$$DG(K, \lambda)(\Delta K, \Delta\lambda) = \mathcal{T}_{f_\lambda,K,\omega}(\Delta K) + \frac{\partial f_{\lambda(K)}}{\partial \lambda}\Delta\lambda.$$

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 $DF(K)$ is, except for quadratic terms, an upper triangular linear differential equation. In section 4.3 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 (18) 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

$$DF(K)\Delta(\theta) = Df(K(\theta))\Delta(\theta) - \Delta(\theta+\omega) = -e(\theta), \tag{20}$$

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

$$e(\theta) = F(K)(\theta), \qquad \theta \in \mathbb{R}^n. \tag{21}$$

By using the geometric properties of the problem we will prove that (20) can be transformed into another equation 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 (21) is small enough, then for any $\theta \in \mathbb{T}^n$ the set

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

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 (20) 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.

Let L be the linear application defined by

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

then

$$L(\theta) = DK(\theta)^\top J(K(\theta))DK(\theta), \tag{22}$$

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'. We will show that if a parametrization K satisfies (1) approximately, then the norm of L can be estimated by the norm of the error given in (21), e.

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

Lemma 1. *Assume that K is a solution of (1) then $K^*\Omega$ (equivalently, L defined in (22)) is identically zero.*

Proof. 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

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

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

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

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

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

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

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. □

4.1.2. Approximate solutions. We need the following result; for a proof see [Rüs75, Rüs76b, Rüs76c] (see also [dlL01]).

Proposition 2. *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*

$$v(\theta) - v(\theta + \omega) = h(\theta) \tag{23}$$

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:*

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

where c_0 is a constant depending on n and σ.

The approximate Lagrangian character of an approximate solution of (1) follows because $K^*\Omega$ satisfies a small divisors equation, similar to (23), with the right-hand side expressed in terms of derivatives of the error function e defined in (21). We make this more precise in the following statement.

Lemma 3. *Let e be the error function defined in (21), then $K^*\Omega$ satisfies the following linear difference equation:*

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

where

$$(\Omega_e)_\theta = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}(\theta)\, \mathrm{d}\theta_i\, \mathrm{d}\theta_j,$$

where g_{ij} 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 $0 < \delta < \rho/2$

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

where L is given in (22).

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

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

and

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

where L is defined in (22). Using the equality

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

which follows from (21), 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

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

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

$$\|\varphi\|_\rho \leqslant |J|_{C^1,\mathcal{B}_r}\|e\|_\rho,$$
$$\|\psi\|_{\rho-\delta} \leqslant c\delta^{-1}\|e\|_\rho,$$

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

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

Applying the last inequality and proposition 2 we obtain an estimation for the norm of L. Indeed, equation (24) is equivalent to

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

Proposition 2 implies

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

where c_0 is the constant of proposition 2. □

4.2. *Approximate reducibility*

In this section we will use the results of section 4.1 to prove that there exists a change of variables $\Delta(\theta) = M(\theta)\,\xi(\theta)$ such that in the variable ξ equation (20) takes the form

$$(C + B)\xi - \xi \circ T_\omega = -(M \circ T_\omega)^{-1}e, \tag{26}$$

where $C(\theta)$ is an upper triangular matrix with diagonal terms equal to 1.

We will see that, in the invariant case (i.e. K is a solution for (1)) M can be chosen in such a way that B is identically equal to zero.

In the invariant case the change of variables is obtained by using the fact that $Df(K(\theta))$ is a linear symplectic transformation taking the tangent vector $(\partial K(\theta)/\partial\theta_j)$ into the tangent vector $(\partial K(\theta + \omega)/\partial\theta_j)$, and the Lagrangian character of an invariant torus. In the case that K is an approximate solution for (1) we will use the change of variables defined in the invariant case. The fact that $Df(K(\theta))$ takes $(\partial K(\theta)/\partial\theta_j)$ into $(\partial K(\theta + \omega)/\partial\theta_j) + (\partial e(\theta)/\partial\theta_j)$ and that K is approximately Lagrangian will produce the following estimate:

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

4.2.1. Exact solutions. Let us look for a matrix-valued function M such that

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

where $C(\theta)$ is an upper triangular matrix with diagonal terms equal to 1.

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:

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

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

Define

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

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

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

with

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

Therefore, we have proved the following result.

Proposition 4. *Assume that K satisfies (1). If*

$$M(\theta) \stackrel{\text{def}}{=} (DK(\theta) \quad J(K(\theta))^{-1}DK(\theta)N(\theta)) \tag{31}$$

and

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

with S defined in (29), then

$$Df(K(\theta))M(\theta) = M(\theta+\omega)C(\theta). \tag{33}$$

Remark 8. If the symplectic linear transformation satisfies $J^{-1} = -J$, 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 9. It is possible to make further transformations that make C a constant. Nevertheless, since the present form is good enough for our purposes, we do not pursue it here.

Let M be defined by (31), in what follows we show that M is invertible. A direct computation yields

$$M(\theta)^\top J(K(\theta))M(\theta) = V(\theta) + R(\theta), \tag{34}$$

where

$$V(\theta) \stackrel{\text{def}}{=} \begin{pmatrix} 0 & I_n \\ -I_n & -P(\theta)^\top J(K(\theta))^{-1}P(\theta) \end{pmatrix} \tag{35}$$

and

$$R(\theta) = \begin{pmatrix} L(\theta) & 0 \\ 0 & 0 \end{pmatrix} \tag{36}$$

with P, L and N defined in (30), (22) and (28), respectively.

Note that V is invertible, with its inverse given by

$$V(\theta)^{-1} = \begin{pmatrix} -P(\theta)^\top J(K(\theta))^{-1}P(\theta) & -I_n \\ I_n & 0 \end{pmatrix}. \tag{37}$$

If K satisfies (1), then lemma 1 implies $L = 0$, this implies $R = 0$. Hence, equality (34) 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 (37) and

$$T(\theta) \stackrel{\text{def}}{=} P(\theta)^\top[I_n - J(K(\theta))^{-1}P(\theta)DK(\theta)^\top J(K(\theta))], \tag{38}$$

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

Since M and C satisfy equality (33), we have that $M(\theta+\omega)^{-1}Df(K(\theta))M(\theta) = C(\theta)$. Therefore, if we substitute $\Delta(\theta) = M(\theta)\xi(\theta)$ in the linear equation (20) we have

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

with

$$p(\theta) = -M(\theta+\omega)^{-1}e(\theta) = \begin{pmatrix} -T(\theta+\omega)e(\theta) \\ -DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta) \end{pmatrix}, \tag{39}$$

where $P(\theta) = DK(\theta)N(\theta)$ (see (30)), and T is defined in (38).

Remark 10. 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 (22). 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 parametrization of an invariant torus and Ω is the canonical symplectic form in $\mathbb{R}^{2n}$, then M is symplectic:

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

4.2.2. Approximate solutions. Assume that K is an approximate solution of (1) with error given by (21), and let M be defined by (31). We will show that, with the change of variables $\Delta = M\xi$, equation (20) takes the form of (26).

In order to guarantee that M defines a change of variables we need M to be invertible. From equality (34) 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 expressions (36) and (37) the following lemma is immediate.

Lemma 5. *Assume that the hypotheses of lemma 3 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 3 such that for any* $0 < \delta < \rho/2$ *the following inequality holds:*

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

As a consequence of lemma 5 we have:

Lemma 6. *Assume that the hypotheses of lemma 3 hold, and let* c_3 *be the constant of lemma 5. If* e *satisfies*

$$c_3\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho \leqslant \tfrac{1}{2}, \tag{40}$$

then M *is invertible and its inverse is given by*

$$M(\theta)^{-1} = V(\theta)^{-1}M(\theta)^\top J(K(\theta)) + M_e(\theta), \tag{41}$$

where

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

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

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

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 5 and the assumption on $\|e\|_\rho$ one obtains

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

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

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

and using equality (34) 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 7. *Assume that* $\omega \in D(\gamma, \sigma)$, *and that* $\|e\|_\rho$ *satisfies inequality (40), where* c_3 *is the constant of lemma 5.*

The change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, *transforms (20) into*

$$\left[\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} + B(\theta)\right]\xi(\theta) - \xi(\theta + \omega) = p(\theta) + w(\theta), \tag{42}$$

where S *is defined in (29),* p *is given in (39), and*

$$B(\theta) \stackrel{\text{def}}{=} M(\theta + \omega)^{-1}E(\theta), \qquad w(\theta) \stackrel{\text{def}}{=} -M_e(\theta + \omega)e(\theta) \tag{43}$$

with

$$E(\theta) \stackrel{\text{def}}{=} Df(K(\theta))M(\theta) - M(\theta+\omega)C(\theta). \tag{44}$$

Moreover, the following estimates hold:

$$\|B(\theta)\|_{\rho-2\delta} < c_5\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho, \tag{45}$$

$$\|w\|_{\rho-2\delta} < c_4\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho^2, \tag{46}$$

where c_5 is a constant which depends on n, ρ, $|f|_{C^1,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $|J^{-1}|_{\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N\|_\rho$, and c_1 and c_4 are constants of lemma 6.

Proof. In the variable ξ, equation (20) 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 (44) and the definition of B we have

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

Moreover, from (41) 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 (39) and (43), respectively, we have (42).

Estimate (46) follows from lemma 6.

Let us prove estimate (45). From lemma 6 one obtains

$$\begin{aligned}\|B\|_{\rho-2\delta} &\leqslant \|V^{-1}\|_\rho\|M(\theta+\omega)^\top J(K(\theta+\omega))E(\theta)\|_{\rho-2\delta} + \|M_e E\|_{\rho-2\delta}\\ &\quad\times c\|M(\theta+\omega)^\top J(K(\theta+\omega))E(\theta)\|_{\rho-2\delta} + c\gamma^{-1}\delta^{-(\sigma+1)}|e|_\rho\end{aligned} \tag{47}$$

for some constant c.

We claim that there exists a constant c_2, depending on n, ρ, $|f|_{C^1,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $|J^{-1}|_{\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N\|_\rho$ and c_1 such that, for any $0 < \delta < \rho/2$, the following estimate holds:

$$\|M(\theta+\omega)^\top J(K(\theta+\omega))E(\theta)\|_{\rho-2\delta} \leqslant c_2\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho. \tag{48}$$

Indeed, using the following equality:

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

we have

$$E(\theta) = (De(\theta) \quad E_1(\theta)),$$

where

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

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

$$P(\theta+\omega)^\top E_1(\theta) = 0 \tag{49}$$

and

$$Q(\theta+\omega)^\top E_1(\theta) = \phi(\theta) - \psi(\theta) - L(\theta+\omega)S(\theta), \tag{50}$$

where $L(\theta)$, defined in (22), is bounded in lemma 3,

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

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 (48) follows from equalities (49), (50) and lemma 3. Therefore, estimate (45) follows from (47) and (48). □

Remark 11. Note that in proposition 7, we have $w = 0$ and $B = 0$ in the case that K parametrizes an invariant torus.

4.3. Approximate solvability of the linearized equation

Proposition 7 guarantees the existence of a change of variables which takes (20) into (42). We emphasize that proposition 7 holds for any real analytic symplectic map f for which there is a real analytic approximate solution of (1), K, with small enough error function and frequency vector $\omega \in D(\gamma, \sigma)$. In this section we prove that if f is an exact symplectic map, and K satisfies the non-degeneracy condition **N2**, then equation (20) can be solved approximately in the customary sense of Nash–Moser theory. In order to do that we study the approximate solvability of (42).

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). Note that (42) can be written as follows:

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

The approximate solutions of (42) will be the solutions of

$$\mathcal{L}_\omega \xi(\theta) = p(\theta) - \begin{pmatrix} 0 \\ \operatorname{avg}\{p_y\}_\theta \end{pmatrix}, \tag{51}$$

where (p_x, p_y) are the symplectically conjugate variables of p. The exactness property of f will enable us to prove that $|\operatorname{avg}\{p_y\}_\theta|$ is quadratic with respect to the norm of the error. Then a solution of (51) will satisfy (42) 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 (51), as we prove in the following result.

Proposition 8. *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, one-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 $\mathrm{avg}\{\xi_x\}_\theta = 0$ *and*

$$\mathrm{avg}\{\xi_y\}_\theta = (\mathrm{avg}\{S\}_\theta)^{-1}(\mathrm{avg}\{v_x\}_\theta - \mathrm{avg}\{S(\theta)\tilde{\xi}_y\}_\theta), \tag{52}$$

where $\tilde{\xi}_y = \xi_y - \mathrm{avg}\{\xi_y\}_\theta$*.*

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

$$\xi_x(\theta) - \xi_x(\theta+\omega) = v_x(\theta) - S(\theta)\xi_y(\theta),$$
$$\xi_y(\theta) - \xi_y(\theta+\omega) = v_y(\theta).$$

First we consider the n-dimensional equation

$$\xi_y(\theta) - \xi_y(\theta+\omega) = v_y. \tag{53}$$

Since $\omega \in D(\gamma, \sigma)$ and $\mathrm{avg}\{v_y\}_\theta = 0$, proposition 2 implies the existence of a solution ξ_y of equation (53) which has arbitrary average and is analytic on $U_{\rho-\delta}$, for any $0 < \delta < \rho/2$, and satisfies the following estimate:

$$\|\xi_y\|_{\rho-\delta} \leqslant c_0\gamma^{-1}\delta^{-\sigma}\|v_y\|_\rho + |\mathrm{avg}\{\xi_y\}_\theta|,$$

where c_0 is the constant given in proposition 2.

Moreover, if $\mathrm{avg}\{\xi_y\}_\theta$ is defined by (52), then the following equality holds:

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

Thus, there exists a unique zero average solution ξ_x of

$$\xi_x(\theta) - \xi_x(\theta+\omega) = v_x(\theta) - S(\theta)\xi_y(\theta), \tag{54}$$

satisfying

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

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

Remark 12. We remark that proposition 8 is the justification for our definition of the non-degeneracy condition **N2** (see definition 2).

Remark 13. The uniqueness property of the solution ξ in proposition 8 implies that for any other mapping $\hat{\xi} = (\hat{\xi}_x, \hat{\xi}_y)$ satisfying $\mathcal{L}_\omega \hat{\xi} = v$ will also satisfy

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

Let ξ be the solution of (51), with $\mathrm{avg}\{\xi\}_\theta$ as in proposition 8, and define $\Delta(\theta) = M(\theta)\xi(\theta)$, then

$$F(K)\Delta(\theta) = -e(\theta) + M(\theta+\omega)\left[B(\theta)\xi(\theta) - w(\theta) - \begin{pmatrix} 0 \\ \mathrm{avg}\{p_y\}_\theta \end{pmatrix}\right], \tag{55}$$

B and w are defined in equation (43).

Note that if we want to obtain an approximate solution of (20) with quadratic error with respect to the norm of e, because of (45), (46) and the estimates given in proposition 8, it is enough to prove that the norm of $\mathrm{avg}\{p_y\}_\theta$ is quadratic with respect to the norm $\|e\|_\rho$.

Lemma 9. *If f is exact symplectic, then the following equality holds:*

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

where e is defined in (21), a is defined by $\alpha = a\,\mathrm{d}z$, and

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

$$\phi(\theta) = a(f(K(\theta))) - a(K(\theta+\omega)) - Da(K(\theta+\omega))e(\theta). \tag{56}$$

Moreover, the following estimate holds:

$$|\mathrm{avg}\{p_y\}_\theta| \leqslant c_5\rho^{-1}\|e\|_\rho^2, \tag{57}$$

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 note that

$$p_x(\theta) = -T(\theta+\omega)e(\theta), \qquad p_y(\theta) = -DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta), \tag{58}$$

where T is defined in (38). Moreover, 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 $\Omega = \mathrm{d}\alpha$, with $\alpha = a\,\mathrm{d}z$, 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

$$J(z) = Da(z)^\top - Da(z). \tag{59}$$

Using equality (59) we obtain

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

Note that

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

and

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

Then

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

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

$$\mathrm{avg}\{DK(\theta)^\top a(K(\theta))\}_\theta = \mathrm{avg}\{(Df(K(\theta))DK(\theta))^\top a(f(K(\theta)))\}_\theta, \tag{60}$$

Equality (60) implies (see (56))

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

Therefore,

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

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

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

$$|\operatorname{avg}\{p_y\}_\theta|\leqslant c_5\rho^{-1}\|e\|_\rho^2,$$

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 propositions 7, 8 and equation (57), we have that the linearized equation

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

is approximately solvable in the following sense.

Lemma 10. *Assume that $\omega\in D(\gamma,\sigma)$, f is an exact symplectic map, S (defined in (29)) satisfies that $\operatorname{avg}\{S\}_\theta$ is non-singular, and that $\|e\|_\rho$ is small enough such that proposition 7 applies.*

Let M be defined by (31), and ξ be the solution of (51). The mapping $\Delta(\theta)=M(\theta)\xi(\theta)$ satisfies (55). Moreover, the following estimates hold:

$$\|\Delta\|_{\rho-2\delta}\leqslant c\gamma^{-2}\delta^{-2\sigma}\|e\|_\rho \tag{61}$$

and

$$\|DF(K)\Delta+e\|_{\rho-2\delta}\leqslant c\gamma^{-3}\delta^{-(3\sigma+1)}\|e\|_\rho^2, \tag{62}$$

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

Proof. Estimate (61) follows from proposition 8. Equality (55) follows from proposition 7 and the definition of Δ. From inequality (45) and proposition 8 one obtains

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

for some constant c. Inequality (46) implies

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

and from inequality (57) we obtain

$$\left\|M(\theta+\omega)\begin{pmatrix}0\\ \operatorname{avg}\{p_y\}_\theta\end{pmatrix}\right\|_{\rho-2\delta}\leqslant c\gamma^{-1}\rho^{-1}\|e\|_\rho^2.$$

Applying these estimates to (55) we obtain (62). □

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 equation (1), we define the sequence

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

where ΔK_{m-1} is the approximate solution of the following linear equation:

$$DF(K_{m-1})\Delta K_{m-1} = -e_{m-1}, \tag{63}$$

$$e_{m-1}(\theta) = F(K_{m-1})(\theta). \tag{64}$$

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 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 11. *Assume that $K_{m-1} \in \mathcal{ND}(\rho_{m-1})$ (see definition 2) is an approximate solution of (1) and that the following holds:*

$$r_{m-1} \stackrel{\text{def}}{=} \|K_{m-1} - K_0\|_{\rho_{m-1}} < r. \tag{65}$$

If e_{m-1}, defined in (64), is small enough such that lemma 10 applies, then there exists a function $\Delta K_{m-1} \in \mathcal{P}_{\rho_{m-1}-3\delta_{m-1}}$, for any $0 < \delta_{m-1} < \rho_{m-1}/3$, such that

$$\begin{aligned} \|\Delta K_{m-1}\|_{\rho_{m-1}-2\delta_{m-1}} &< c_{m-1}\gamma^{-2}\delta_{m-1}^{-2\sigma}\|e_{m-1}\|_{\rho_{m-1}}, \\ \|D\Delta K_{m-1}\|_{\rho_{m-1}-3\delta_{m-1}} &< c_{m-1}\gamma^{-2}\delta_{m-1}^{-(2\sigma+1)}\|e_{m-1}\|_{\rho_{m-1}}, \end{aligned} \tag{66}$$

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

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

$$r_{m-1} + c_{m-1}\gamma^{-2}\delta_{m-1}^{-2\sigma}\|e_{m-1}\|_{\rho_{m-1}} < r, \tag{67}$$

then we can redefine c_{m-1}, which 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

$$\|e_m\|_{\rho_m} \leqslant c_{m-1}\gamma^{-4}\delta_{m-1}^{-4\sigma}\|e_{m-1}\|_{\rho_{m-1}}^2. \tag{68}$$

Proof. Let $\Delta K_{m-1}(\theta) = M_{m-1}(\theta)\xi_{m-1}(\theta)$ be the approximate solution, provided by lemma 10, of (63). Then, estimates (66) follow from (61) and Cauchy's inequalities.

Moreover, if $K_m = K_{m-1} + \Delta K_{m-1}$, then

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

Hence, if (67) holds $K_m(\theta) \in \mathcal{B}_r$ for any θ with $|\text{Im}\,\theta| < \rho_{m-1} - 2\delta_{m-1}$. Define the remainder of the Taylor expansion

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

then

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

Therefore, estimate (68) follows from Taylor's theorem and (62). □

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 satisfies the non-degeneracy conditions **N1**, **N2**, for $m \geqslant 1$.

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 (65) and (67) hold.

Since conditions **N1**, **N2** involve the derivative of K_m, its 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 11). We define $\rho_m = \rho_{m-1} - 3\delta_{m-1}$, for $0 < \delta_{m-1} < \rho_{m-1}/3$ fixed.

Lemma 12. *Assume that the hypotheses of lemma 11 hold. If* $\|e_{m-1}\|_{\rho_{m-1}}$ *is small enough, 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*

$$\|N_m\|_{\rho_m} \leqslant \|N_{m-1}\|_{\rho_{m-1}} + \tilde{c}_{m-1}\gamma^{-2}\delta_{m-1}^{-(2\sigma+1)}\|e_{m-1}\|_{\rho_{m-1}}.$$

(2) Let S_{m-1} *and* S_m *be defined by replacing* K *with* K_{m-1} *and* K_m*, respectively, in (29). If* $\mathrm{avg}\{S_{m-1}\}_\theta$ *is invertible, then* $\mathrm{avg}\{S_m\}_\theta$ *is also invertible. Moreover, the following inequality holds:*

$$|(\mathrm{avg}\{S_m\}_\theta)^{-1}| < |(\mathrm{avg}\{S_{m-1}\}_\theta)^{-1}| + \hat{c}_{m-1}\gamma^{-2}\delta_{m-1}^{-(2\sigma+1)}\|e_{m-1}\|_{\rho_{m-1}},$$

where $\tilde{c}_{m-1}$ *and* $\hat{c}_{m-1}$ *depend on the same quantities as* c_{m-1} *in lemma 11,* $|f|_{C^2,\mathcal{B}_r}$ *and* $|J^{-1}|_{C^1,\mathcal{B}_r}$*.*

Proof. Performing some computations we have

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

where Z_{m-1} satisfies

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

where we have used estimates given in (66) and have ignored the quadratic terms at the price of redefining c_{m-1}. Similarly, we have

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

Therefore, if

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

and e_{m-1} satisfies

$$b_{m-1}c_{m-1}\gamma^{-2}\delta_{m-1}^{-(2\sigma+1)}\|e_{m-1}\|_{\rho_{m-1}} < \tfrac{1}{2}, \tag{69}$$

then we have that the matrices

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

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

$$DK_m(\theta)^\top DK_m(\theta) = N_{m-1}(\theta)^{-1}[I_n + N_{m-1}(\theta)Z(\theta)],$$
$$\mathrm{avg}\{S_m\}_\theta = \mathrm{avg}\{S_{m-1}\}_\theta[I_n + (\mathrm{avg}\{S_{m-1}\}_\theta)^{-1}\mathrm{avg}\{\Phi_{m-1}\}_\theta]$$

are invertible. □

Remark 14. For typographical reasons we redefine constant c_{m-1}, given in lemma 11, to be

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

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 $K_{m-1}(U_{\rho_{m-1}}) \subset \mathcal{B}_r$ 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 11 and 12), such that if $\|e_{m-1}\|_{\rho_{m-1}}$ satisfies

$$
\begin{aligned}
& r_{m-1} + c_{m-1}\gamma^{-2}\delta_{m-1}^{-2\sigma}\|e_{m-1}\|_{\rho_{m-1}} < r, \\
& c_{m-1}\gamma^{-2}\delta_{m-1}^{-(\sigma+1)}\|e_{m-1}\|_{\rho_{m-1}} \leqslant \tfrac{1}{2},
\end{aligned}
\tag{70}
$$

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 (68) (see lemma 11).

We will establish conditions on $\|e_0\|_{\rho_0}$ that guarantee that inequalities (70) hold for all $m \geqslant 1$. Then we will prove that the Newton method described in section 5.1 is quadratically convergent and 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 15. In principle, the constant c_{m-1} depends on the parameters of the problem n, σ, γ, $|f|_{C^2,\mathcal{B}_r}$, $|J|_{C^1,\mathcal{B}_r}$, $|J^{-1}|_{C^1,\mathcal{B}_r}$. It also depends on $\rho_{m-1} \leqslant \rho_0$, and on the following quantities, related to the approximation K_{m-1},

$$
d_m \stackrel{\text{def}}{=} \|DK_{m-1}\|_{\rho_{m-1}}, \qquad \nu_m \stackrel{\text{def}}{=} \|N_{m-1}\|_{\rho_{m-1}}, \qquad \tau_m \stackrel{\text{def}}{=} |(\text{avg}\{S_{m-1}\}_\theta)^{-1}|.
$$

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

$$
c_{m-1} = \lambda(d_{m-1}, \nu_{m-1}, \tau_{m-1}), \qquad \forall m \geqslant 1. \tag{71}
$$

Since we are looking for a solution of (1) which is close to K_0, it is natural to expect 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 13. *Let $\{c_m\}_{m\geqslant 0}$ be the sequence of positive numbers given in (71). For a fixed $0 < \delta_0 < \min(\rho_0/12, 1)$ define*

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

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

$$
\kappa \stackrel{\text{def}}{=} 2^{4\sigma} c\gamma^{-4}\delta_0^{-4\sigma}\|e_0\|_{\rho_0} \leqslant \tfrac{1}{2}, \tag{72}
$$

$$
\left[1 + \frac{2^{4\sigma}}{2^{2\sigma} - 1}\right] c\gamma^{-2}\delta_0^{-2\sigma}\|e_0\|_{\rho_0} < r, \tag{73}
$$

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:

$$\|K_\infty - K_0\|_{\rho_0-6\delta_0} \leqslant \left[\frac{2^{2\sigma}}{2^{2\sigma}-1}\right] c\gamma^{-2}\delta_0^{-2\sigma}\|e_0\|_{\rho_0}. \tag{74}$$

Proof. It is enough to prove that conditions (72) and (73) imply conditions (70) for all $m \geqslant 0$. Define

$$c \stackrel{\text{def}}{=} \lambda(d_0+\beta, \nu_0+\beta, \tau_0+\beta) \tag{75}$$

with

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

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

C1(*m*) $r_m \leqslant [1+(\kappa(2^{4\sigma})/(2^{2\sigma}-1))]c\gamma^{-2}\delta_0^{-2\sigma}\varepsilon_0 < r$,
C2(*m*) $\varepsilon_m \leqslant 2^{-4\sigma(m-1)}\kappa^{(2^m-1)}\varepsilon_0$,
C3(*m*) $c_m \leqslant c$,
C4(*m*) $c_m\gamma^{-2}\delta_m^{-(2\sigma+1)}\varepsilon_m \leqslant \frac{1}{2}$.

Note that this implies that conditions (70) hold for all $m \geqslant 1$.

C1(0) and **C2(0)** are immediate. Note that from the definition $c_0 \leqslant c$, i.e. **C3(0)** holds, and since we are assuming $\sigma > 1$ and $0 < \gamma \leqslant 1$, inequality (72) implies

$$c_0\gamma^{-2}\delta_0^{-(2\sigma+1)}\varepsilon_0 \leqslant (c_0\gamma^{-4}\delta_0^{-4\sigma}\varepsilon_0)\delta_0^{2\sigma-1} < 2^{-4\sigma} < \tfrac{1}{2}.$$

This proves **C4(0)**.

Assume that **C1(*j*)**–**C4(*j*)** holds for $j = 1, \ldots, m-1$ (then it is possible to perform $m-1$ steps of the modified Newton method described in section 5.1). In particular, for $j = 0, \ldots, m-1 \geqslant 0$, c_j is bounded by c, then for $j = 1, \ldots, m$, one has the following estimates of the error (see (68)):

$$\begin{aligned}
\varepsilon_j &\leqslant c\gamma^{-4}\delta_{j-1}^{-4\sigma}\varepsilon_{j-1}^2 \leqslant (c\gamma^{-4})^{(1+2)}(\delta_{j-1}\delta_{j-2}^2)^{-4\sigma}\varepsilon_{j-2}^4\\
&\leqslant \cdots \leqslant (c\gamma^{-4})^{1+2+\cdots+2^{j-1}}(\delta_{j-1}\delta_{j-2}^2\cdots\delta_0^{2^{j-1}})^{-4\sigma}\varepsilon_0^{2^j}\\
&= (c\gamma^{-4}\delta_0^{-4\sigma})^{1+2+\cdots+2^{j-1}}(2^{4\sigma})^{2^0(j-1)+2(j-2)+\cdots+2^{j-2}}\varepsilon_0^{2^j}\\
&\leqslant (c\gamma^{-4}\delta_0^{-4\sigma})^{2^j-1}2^{4\sigma(2^j-j)}\varepsilon_0^{2^j}\\
&\leqslant (\underbrace{c\gamma^{-4}\delta_0^{-4\sigma}2^{4\sigma}\varepsilon_0}_{\kappa})^{2^j-1}2^{-4\sigma(j-1)}\varepsilon_0 \leqslant \kappa^{2^j-1}(2^{-4\sigma(j-1)})\varepsilon_0,
\end{aligned}$$

where we have used that $2^0(j-1)+2(j-2)+\cdots+2^{j-2} = 2^{j-1}\sum_{s=1}^{j-1}s2^{-s} \leqslant 2^j - j$.

Let us prove that **C1(*m*)–C4(*m*)** holds. We first prove **C1(*m*)**:

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

and since $\kappa \leqslant 1/2$, (73) implies **C1(*m*)**.

Therefore, inequalities (68) and (72) imply

$$\varepsilon_j < \kappa^{2^j-1}\varepsilon_0 2^{-4\sigma(j-1)}, \qquad \text{for } j = 1, \dots, m, \tag{76}$$

this proves **C2(*m*)**.

We now prove **C3(*m*)**. The first inequality in (66) and lemma 12 imply

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

where

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

Let us estimate s_{m-1}. Inequality (76) and **C3(*j*)** imply, for $j = 1, \dots, m-1$,

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

then, using inequality (72) one has

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

Then

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

this implies $c_m \leqslant c$, i.e. **C3(*m*)** holds. From (76) and (72) we have

$$\begin{aligned} c_m\gamma^{-2}\delta_m^{-(2\sigma+1)}\varepsilon_m &\leqslant (c\gamma^{-2}\delta_0^{-(2\sigma+1)}\varepsilon_0)2^{4\sigma}\kappa^{(2^m-1)}2^{-m(2\sigma-1)} \\ &\leqslant \kappa\gamma^2\delta_0^{2\sigma-1}\kappa^{2^m-1}2^{-m(2\sigma-1)} \leqslant \tfrac{1}{2}. \end{aligned}$$

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 11. In particular, from the first inequality in (66) we have that $\{K_m\}_{m\geqslant 0}$ 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 $\rho_\infty \stackrel{\text{def}}{=} \lim_{m\to\infty}\rho_m = \rho_0 - 6\delta_0$. Moreover, inequality (68) implies that K_∞ satisfies (1). Finally, estimate (74) follows from the first inequality in (66) and

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

□

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 an approximate left inverse (see [Zeh75]).

In our context the existence of the approximate left inverse amounts to the uniqueness statement of proposition 8 and remark 13. The uniqueness up to additive constants in these results amounts to the uniqueness of re-parametrization 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

$$0 = F(K_1) - F(K_2) = DF(K_2)(K_1 - K_2) + \mathcal{R}(K_1, K_2), \tag{77}$$

where

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

Since K_2 is an exact solution of (1), the results of section 4.2 apply and proposition 7 implies that equation (77) 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 with 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 (38) (replacing K with K_2).

By the uniqueness statement in proposition 8 and remark 13 we obtain

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

where

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

Lemma 14. *There exists a constant* $\tilde{c}$, *depending on* n, ρ^{-1}, $|J|_{\mathcal{B}_r}$, $|J^{-1}|_{\mathcal{B}_r}$, $\|N\|_\rho$, $\|K_1\|_{C^2,\rho}$, *such that if*

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

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

$$\operatorname{avg}\{T(\theta)[K_1 \circ T_{\tau_1} - K_2](\theta)\}_\theta = 0. \tag{78}$$

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

$$\|K_1 \circ T_{\tau_1} - K_2\|_{\rho-2\delta} < \hat{c}\gamma^{-2}\delta^{-2\sigma}\|K_1 - K_2\|_\rho^2 \tag{79}$$

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

Proof. Some direct computations (see (38)) show that

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

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

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

Therefore, a solution τ_1 of (78) is a fixed point of

$$\Psi(x) \stackrel{\text{def}}{=} -\operatorname{avg}\{T(\theta)[K_1(\theta + x) - K_2(\theta)] - T(\theta)DK_2(\theta)x\}_\theta.$$

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

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

where the constant c_4 depends on n, ρ^{-1}, $\|K_2\|_{C^2,\rho}$ and $\|T\|_\rho$ (i.e. $|J|_{\mathcal{B}_r}$, $|J^{-1}|_{\mathcal{B}_r}$, $\|DK_2\|_\rho$ and $\|N\|_\rho$, see (38)). 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 (78), then $K_1 \circ T_{\tau_1}$ is a solution of (1) such that if

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

then the uniqueness statement in proposition 8, remark 13 and equality (78) imply

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

Therefore, we have

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

for any $0 < \delta, \rho/2$. □

Let τ_1 be as in lemma 14, since $K_1 \circ T_{\tau_1}$ is also a solution of (1), we can apply lemma 14 to the solutions $K_1 \circ T_{\tau_1}$ and K_2 (note 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 $\{\tau_m\}_{m\geqslant 1}$ such that

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

and

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

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

Therefore, if

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

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 are identical to those 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 (14) with error function

$$e = f_\lambda \circ K - K \circ T_\omega \tag{80}$$

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 one-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 (19). This leads to a study of the solvability properties of the linearized equation:

$$\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda + Df(K(\theta))\Delta K(\theta) - \Delta K(\theta+\omega) = -e(\theta). \tag{81}$$

Note that, except for the term $(\partial f_\lambda(K(\theta))/\partial\lambda)\Delta\lambda$, the left-hand part of (81) 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) and the results on the reducibility of the linearized equation in section 4.2 follow from the symplectic properties of f and the size of the error function, then the results of sections 4.1 and 4.2 also hold for the symplectic map f_λ. The main difference is that in section 4.3 the approximate solvability was achieved using the exactness character of f and the non-degeneracy condition **N2** (see proposition 8 and lemma 9). In this 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 (81) 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 (81) 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 7, applied to the map f_λ and K, implies that if M is defined by (31), replacing f with f_λ, then the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$ transforms (81) into

$$\left[\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} + B(\theta)\right]\xi(\theta) - \xi(\theta+\omega) = p(\theta) + w(\theta) + M(\theta+\omega)^{-1}\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\Delta\lambda, \tag{82}$$

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

$$p(\theta) = -V(\theta+\omega)^{-1}M(\theta+\omega)^\top J(K(\theta+\omega))e(\theta), \tag{83}$$

where V^{-1} is given in (37), and B and w satisfy estimates

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

for some constant $c > 0$.

Moreover, lemma 6 implies

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

with

$$\Lambda(\theta) \stackrel{\text{def}}{=} V(\theta+\omega)^{-1}M(\theta+\omega)^\top J(K(\theta+\omega))\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right) \tag{84}$$

and

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

Let us prove that the reduced equation obtained by removing the terms B, w and q from equation (82) can be solved for $(\xi, \Delta\lambda)$.

Proposition 15. *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*

$$\begin{pmatrix} I_n & S(\theta) \\ 0 & I_n \end{pmatrix} \xi(\theta) - \xi(\theta + \omega) = R(\theta), \tag{85}$$

where

$$R(\theta) \stackrel{\text{def}}{=} -V(\theta+\omega)^{-1} M(\theta+\omega)^\top J(K(\theta+\omega)) \left[e(\theta) + \left(\frac{\partial f_\lambda(K(\theta))}{\partial \lambda} \right) \Delta\lambda \right]. \tag{86}$$

Moreover, there exists a constant c, depending on n, σ, ρ, r, $|f|_{C^2,\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N(\theta)\|_\rho$, $\|(\partial f_\lambda(K(\theta))/\partial\lambda)\|_\rho$ and $|(\text{avg}\{\Lambda\}_\theta)^{-1}|$ (see (84)) such that the following inequalities hold:

$$\|\xi\|_{\rho-2\delta} \leqslant c\gamma^{-2}\delta^{-2\sigma}\|e\|_\rho, \tag{87}$$

$$|\Delta\lambda| \leqslant c|(\text{avg}\{\Lambda(\theta)\}_\theta)^{-1}|\|e\|_\rho. \tag{88}$$

Proof. Consider the symplectically conjugated coordinates $R = (R_x, R_y)$ and $\xi = (\xi_x, \xi_y)$, then equation (85) can be written as

$$\xi_x(\theta) - \xi_x(\theta+\omega) = R_x(\theta) - S(\theta)\xi_y(\theta),$$
$$\xi_y(\theta) - \xi_y(\theta+\omega) = R_y(\theta).$$

From (86) one can see that

$$R(\theta) = \begin{pmatrix} -T(\theta+\omega)e(\theta) \\ -DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta) \end{pmatrix} + \Lambda(\theta)\Delta\lambda \tag{89}$$

with Λ given in (84).

Since f_λ and K satisfy definition 3 we have that $\text{avg}\{\Lambda\}_\theta$ has range of dimension $2n$. Therefore, we can fix the last n-coordinates of $\Delta\lambda \in \mathbb{R}^{2n}$ such that

$$\text{avg}\{R_y\}_\theta = 0.$$

Proposition 2 implies that there exists a unique zero-average analytic mapping on $U_{\rho-\delta}$ satisfying

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

and

$$\|\xi_y\|_{\rho-\delta} \leqslant c\gamma^{-1}\delta^{-\sigma}\|R_y\|_\rho. \tag{90}$$

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

$$\text{avg}\{R_x(\theta) - S(\theta)\xi_y(\theta)\}_\theta = 0.$$

By proposition 2 there exists a unique mapping ξ_x satisfying

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

with

$$\|\xi_x\|_{\rho-2\delta} \leqslant c\gamma^{-1}\delta^{-\sigma}\|R_x(\theta) - S(\theta)\xi_y(\theta)\|_{\rho-\delta}. \tag{91}$$

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

$$\text{avg}\{\Lambda(\theta)\}_\theta \Delta\lambda = \begin{pmatrix} \text{avg}\{-T(\theta+\omega)e(\theta) - S(\theta)\xi_y(\theta)\}_\theta \\ \text{avg}\{-DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta)\}_\theta \end{pmatrix}. \tag{92}$$

Let $(\varphi_x, \varphi_y)^\top = \text{avg}\{\Lambda(\theta)\}_\theta \Delta\lambda$, then from (89) and (92) we have the following estimates:

$$\|R_y\|_\rho \leqslant c\left(\|e\|_\rho + \left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho |\varphi_y|\right) \leqslant \tilde{c}\|e\|_\rho \tag{93}$$

and

$$\begin{aligned}\|R_x\|_\rho &\leqslant c\left(\|e\|_\rho + \left\|\left(\frac{\partial f_\lambda(K(\theta))}{\partial\lambda}\right)\right\|_\rho |\varphi_x|\right)\\ &\leqslant \tilde{c}(\|e\|_\rho + |\text{avg}\{S(\theta)\xi_y(\theta)\}_\theta|)\\ &\leqslant \tilde{c}(\|e\|_\rho + c'\gamma^{-1}\rho^{-1}\|R_y\|_\rho).\end{aligned} \tag{94}$$

Inequalities (90), (91), (93), (94) and equality (92) imply estimates (87) and (88). □

The construction of an approximate solution of the linearized equation (81) is now clear, and we establish this in the following result, the proof follows from the previous construction and is very similar to the proof of lemma 10.

Proposition 16. *Assume that the hypotheses of proposition 15 hold. Let ξ and $\Delta\lambda$ be as in proposition 15, define $\Delta K(\theta) \stackrel{\text{def}}{=} M(\theta)\xi(\theta)$. The pair $(\Delta K, \Delta\lambda)$ is an approximate solution of the linearized equation (81) satisfying*

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

for some constant c, depending on n, σ, r, $|f|_{C^1,\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N(\theta)\|_\rho$, $|(\text{avg}\{\Lambda\}_\theta)^{-1}|$ (see (84)) and $\|(\partial f_\lambda(K(\theta))/\partial\lambda)\|_\rho$.

Remark 16. In the case that f_λ is exact symplectic we have proved in section 4.3 that (see lemma 9)

$$|\text{avg}\{DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta)\}_\theta| \leqslant c\rho^{-1}\|e\|_\rho^2,$$

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

$$|\Delta\lambda_y| \leqslant c'|(\text{avg}\{\Lambda\}_\theta)^{-1}||\text{avg}\{DK(\theta+\omega)^\top J(K(\theta+\omega))e(\theta)\}_\theta| \leqslant c''\|e\|_\rho^2.$$

Moreover, if K satisfies the non-degeneracy condition **N2** stated in definition (29), i.e. the $n \times n$ matrix $\text{avg}\{S(\theta)\}_\theta$, with S defined by (29), replacing f with f_λ, is non-singular, then we can choose $\text{avg}\{\xi_y\}_\theta$ in proposition 15 such that

$$\text{avg}\{-T(\theta+\omega)e(\theta) - S(\theta)\tilde{\xi}\}_\theta,$$

then $\Delta\lambda_y = 0$ (see (92)).

7.1.1. The iterative procedure. Let ΔK and $\Delta\lambda$ be as in proposition 16, we consider $K + \Delta K$ as an approximate 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 (85) 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 (87) and (88) imply

$$\|\Delta K\|_{\rho-2\delta} \leqslant c\delta^{-2\sigma}\|e\|_\rho, \qquad |\Delta\lambda| \leqslant c\|e\|_\rho \tag{95}$$

and by using Cauchy's inequality one has

$$\|D\Delta K\|_{\rho-3\delta} \leqslant c\delta^{-(2\sigma+1)}\|e\|_\rho. \tag{96}$$

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 17. *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 (84). Moreover, if* Λ *is defined by (84), then the following inequality holds:*

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

where Λ_1 *is defined by (84) by replacing* f_λ *and* K *with* $f_{\lambda+\Delta\lambda}$ *and* $K + \Delta K$, *respectively, and* c *is a constant depending on* n, σ, r, $|f|_{C^2,\mathcal{B}_r}$, $\|DK\|_\rho$, $\|N(\theta)\|_\rho$, $\|(\partial^2 f_\lambda(K(\theta))/\partial^2\lambda)\|_\rho$ *and* $|(\mathrm{avg}\{\Lambda\}_\theta)^{-1}|$.

Proof. The proof is very similar to the proof of lemma 12 in section 5.2, it follows the assumption that the family f_λ is twice differentiable with respect to λ, the estimations given in (95) and (96) and Neumann's series theorem. □

Since the estimate and the induction hypotheses 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 (15). We study the invertibility properties of the linear operator

$$JD(\nabla H(K))\Delta - \partial_\omega \Delta. \tag{97}$$

8.1. Lagrangian character

We need the version of proposition 2 for vector fields.

Proposition 18. *Assume that* ω *satisfies the Diophantine condition*

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

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 the following estimate:*

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

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

It is known that if K is a solution of (15) with ω rationally independent, then the torus $K(\mathbb{T}^n)$ is a Lagrangian manifold. In the case that K satisfies (15) only approximately, we will prove that $K(\mathbb{T}^n)$ is approximate Lagrangian.

Lemma 19. *Let* $K : \mathbb{T}^n \to \mathbf{U}$ *be a real analytic mapping on the complex strip of width* $\rho > 0$. *Define the error function*

$$e(\theta) = J\nabla H(K(\theta)) - \partial_\omega K(\theta). \tag{98}$$

The following holds:

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

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

$$\partial_\omega K = J\nabla H \circ K, \tag{99}$$

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

$$K^*\alpha = \mathrm{d}b + a_0\,\mathrm{d}\theta,$$

where $\alpha = -y\,\mathrm{d}x$.

(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*

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

Moreover, the function g *satisfies*

$$\partial_\omega g(\theta) = DK(\theta)^\top Je(\theta). \tag{100}$$

(d) *Let* $L(\theta)$ *be defined by (22). There exists a constant* c, *depending on* n, σ *and* $|DK|_\rho$, *such that*

$$|L|_{\rho-2\delta} \leqslant c\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho. \tag{101}$$

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

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

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

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

$$\begin{aligned}\partial_\omega a_j(\theta) &= \left(\partial_\omega \frac{\partial w(\theta)}{\partial\theta_j}\right)^\top v(\theta) + \frac{\partial w(\theta)}{\partial\theta_j}^\top \partial_\omega v(\theta) \\ &= \left[e_x(\theta)\cdot\frac{\partial v(\theta)}{\partial\theta_j} - e_y(\theta)\cdot\frac{\partial w(\theta)}{\partial\theta_j}\right] \\ &\quad + \frac{\partial}{\partial\theta_j}\left[\underbrace{D_yH(K(\theta))v(\theta) + (H\circ K)(\theta)}_{\hbar_1} - \underbrace{e_x(\theta)\cdot v(\theta)}_{\hbar_2}\right].\end{aligned}$$

Note that since $\partial_\omega a_j(\theta)$, $(\partial/\partial\theta_j)\hbar_1$ and $(\partial/\partial\theta_j)\hbar_2$ have average equal to zero, then

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

which proves part (a) of lemma 19. Moreover (cf [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 (100) and

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

If K is a solution of (99), then g and b_2 are constants. Setting $g = 0$ and $b_2 = 0$, we have

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

this proves part (b) of lemma 19. Taking $\tilde{b} = -(b_1 - b_2)$ part (c) of lemma 19 follows.

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

$$K^*\Omega = \mathrm{d}K^*\alpha = \mathrm{d}(-g\,\mathrm{d}\theta)$$

in coordinate representation one has

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

therefore, part (d) of lemma 19 follows from equality (100). □

8.2. *Change of variables in the linearized equation*

As in the case of analytic exact symplectic maps, the geometric structure of K makes it possible to define a change of variables which transforms the linear operator (97) into a simpler one (compare with the case of maps in section 4.2). Let A be defined by

$$A(\theta) \stackrel{\text{def}}{=} \begin{pmatrix} D_x\nabla_y H(K(\theta)) & D_y\nabla_y H(K(\theta)) \\ -D_x\nabla_x H(K(\theta)) & -D_y\nabla_x H(K(\theta)) \end{pmatrix} = JD(\nabla H)(K(\theta)). \quad (102)$$

Then the linearized equation

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

takes the form

$$A(\theta)\Delta(\theta) - \partial_\omega\Delta(\theta) = -e(\theta). \quad (103)$$

If we perform the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, in the variables ξ the linearized equation (103) 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 (102). Similar to the case of exact symplectic maps, we give a change of variables M such that the matrix

$$M(\theta)^{-1}[A(\theta)M(\theta) - \partial_\omega M(\theta)] \quad (104)$$

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

Assume that there exists a matrix N such that

$$N(\theta) = (DK(\theta)^\top DK(\theta))^{-1} \quad (105)$$

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

$$M(\theta) = (DK(\theta) \quad JDK(\theta)N(\theta)). \quad (106)$$

Lemma 20. *If K is an approximate solution of (99), with error e, then if M is defined by (106) we have*

$$A(\theta)M(\theta) - \partial_\omega M(\theta) = M(\theta)\begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix} + E(\theta) \quad (107)$$

with

$$S(\theta) = N(\theta)DK(\theta)^\top[A(\theta)J - JA(\theta)]DK(\theta)N(\theta) \quad (108)$$

and

$$\|JM(\theta)^\top JE\|_{\rho-2\delta} \leqslant c\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_\rho, \quad (109)$$

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

Proof. This is an analogue to the results for maps presented in section 4.2. Since the case when K is an approximate solution of (15) involves a lot of (but simple) computations which could hide the main idea we only consider the case when K is a solution of (15).

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 (15)

$$A(\theta)DK(\theta) = \partial_\omega DK(\theta), \tag{110}$$

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 $A(\theta)J = -JA(\theta)^\top$ and equality (110) implies

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

Moreover, by using equality (110), it is not difficult to see that

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

Since K satisfies (15) the vectors $\{\partial K(\theta)/\partial\theta_j, J(\partial K(\theta)/\partial\theta_j)\}_{j=1}^n$ 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

$$W_1(\theta) = DK(\theta)S(\theta) + JDK(\theta)N(\theta)T(\theta). \tag{111}$$

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

$$T(\theta) = -DK(\theta)^\top JW_1(\theta) = 0.$$

Moreover, multiplying equality (111) by $N(\theta)DK(\theta)^\top$ we have

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

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 (108), and K is a solution of (15), 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 (15) 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 (107) and estimate (109). □

Now we prove that the matrix M defined in (106) is invertible. From the definition of M it follows that

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

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 (101), the norm of R verifies the following inequality:

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

Lemma 21. *If the error function e is small enough, the matrix M, given by (106), is invertible and the inverse is given by*

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

with

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

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

When K solves equation (15) approximately, matrix M defined by (106) reduces (104) into a simple form, and the change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$ will transform the linearized equation (97) into a more simple one (see (112)). The result can be summarized as follows.

Proposition 22. *The change of variables $\Delta(\theta) = M(\theta)\xi(\theta)$, with M defined in (106), transforms the linearized equation (103) into*

$$\left[\begin{pmatrix} 0 & S(\theta) \\ 0 & 0 \end{pmatrix} + B(\theta)\right]\xi(\theta) - \partial_\omega\xi(\theta) = p(\theta) + w(\theta), \tag{112}$$

where S is defined in (108),

$$p(\theta) = -JM(\theta)^{\top}Je(\theta) = \begin{pmatrix} -N(\theta)^{\top}DK(\theta)^{\top}e(\theta) \\ DK(\theta)^{\top}Je(\theta) \end{pmatrix} \tag{113}$$

and B and w satisfy

$$\begin{aligned} \|B\|_{\rho-2\delta} &\leqslant c\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_{\rho}, \\ \|w\|_{\rho-\delta} &\leqslant c\gamma^{-1}\delta^{-(\sigma+1)}\|e\|_{\rho}^{2} \end{aligned} \tag{114}$$

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

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

with $W(\theta) = A(\theta)M(\theta) - \partial_\omega M(\theta)$. From lemma 20 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

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

the estimate (109) and the estimate for the norm of M_e imply the first inequality of (114). Therefore, in the variables ξ, the linear equation (103) is written

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

8.3. Solvability of the reduced equation

Let us consider the reduced equation

$$\begin{aligned} S(\theta)\xi_y - \partial_\omega \xi_x &= -N(\theta)^\top DK(\theta)^\top e(\theta), \\ \partial_\omega \xi_y &= DK(\theta)^\top J e(\theta), \end{aligned} \tag{115}$$

The conditions guaranteeing the existence of a solution (ξ_x, ξ_y) of equation (115) are provided by the following result:

Lemma 23. *There exists a solution (ξ_x, ξ_y) of equation (115) satisfying the following estimate:*

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

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

Proof. From part (c) of lemma 19 $\text{avg}\{DK(\theta)^\top Je(\theta)\}_\theta = 0$, then the small divisors equation

$$\partial_\omega \xi_y = DK(\theta)^\top J e(\theta)$$

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

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

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]. □

Acknowledgments

The authors are grateful to Professor Niederman for his careful reading of this paper and for his valuable suggestions. The work of RL has been supported by NSF grants, a Deans Fellowship at U.T. Austin Spring 03. Visits of RL to Barcelona have been supported by ICREA. AG was supported by a US–Spain Science & Technology Programme 2000, Project 96. ÀJ and JV have also been supported by the Catalan grant 2001SGR-70, the Spanish grant MCyT/FEDER, no BFM2003-07521-C02-01, the INTAS grant 00-221 and DURSI.

References

[Bos86] Bost J-B 1986 Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, . . .) *Astérisque* **133–134** 113–57 *Seminar Bourbaki* vol 1984/85

[CC97] Celletti A and Chierchia L 1997 On the stability of realistic three-body problems *Commun. Math. Phys.* **186** 413–49

[DdlLS03] Delshams A, de la Llave R and Seara T M 2003 A geometric mechanism for diffusion in hamiltonian systems overcoming the large gap problem: heuristics and rigorous verification on a model *Preprint* mp_arc 03-182

[dlL01] de la Llave R 2001 A tutorial on KAM theory *Smooth Ergodic Theory and its Applications (Seattle, WA, 1999)* (Providence, RI: American Mathematical Society) pp 175–292

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

[Dou82] Douady R 1982 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** 201–4

[FGB98] Fassò F, Guzzo M and Benettin G 1998 Nekhoroshev-stability of elliptic equilibria of Hamiltonian systems *Commun. Math. Phys.* **197** 347–60

[GFB98] Guzzo M, Fassò F and Benettin G 1998 On the stability of elliptic equilibria *Math. Phys. Electron. J.* **4** Paper 1, p 16 (electronic)

[HdlL04a] Haro A and de la Llave R 2004 A parameterization method for the computation of whiskers in quasi periodic maps: numerical implementation *Preprint* mp_arc 04-350

[HdlL04b] Haro A and de la Llave R 2004 A parameterization method for the computation of whiskers in quasi periodic maps: rigorous results *Preprint* mp_arc 04-348

[Her83] Herman M-R 1983 *Sur les courbes invariantes par les difféomorphismes de l'anneau vol 1* (Paris: Société Mathématique de France) *Astérisque* **103**

[Koz83] Kozlov S M 1983 Reducibility of quasiperiodic differential operators and averaging *Tr. Moskov. Mat. Obshch.* **46** 99–123

Kozlov S M 1984 *Trans. Moscow Math. Soc.* **2** 101–26 (Engl. Transl.)

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

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

[Mos67] Moser J 1967 Convergent series expansions for quasi-periodic motions *Math. Ann.* **169** 136–76

[Mos88] Moser J 1988 A stability theorem for minimal foliations on a torus *Ergodic Theory Dyn. Syst.* **8*** (Charles Conley Memorial Issue) 251–81

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

[Nie98] Niederman L 1998 Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system *Nonlinearity* **11** 1465–79

[Pös99] Pöschel J 1999 On Nekhoroshev's estimate at an elliptic equilibrium *Internat. Math. Res. Notices* 203–15

[Pös01] Pöschel J 2001 A lecture on the classical KAM theorem *Smooth Ergodic Theory and its Applications (Seattle, WA, 1999)* (Providence, RI: American Mathematical Society) pp 707–32

[Ran87] Rana D 1987 Proof of accurate upper and lower bounds to stability domains in small denominator problems *PhD Thesis* Princeton University

[RC95] Raković M J and Shih-I Chu 1995 New integrable systems: hydrogen atom in external fields *Physica* D **81** 271–9

[RC97] Raković M J and Shih-I Chu 1997 Phase-space structure of a new integrable system related to hydrogen atoms in external fields *J. Phys. A: Math. Gen.* **30** 733–53

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

[Rüs76a] Rüssmann H 1976 On a new proof of Moser's twist mapping theorem *Proc. 5th Conf. on Mathematical Methods in Celestial Mechanics (Oberwolfach, 1975)* vol 14 (part I) pp 19–31

[Rüs76b] Rüssmann H 1976 On optimal estimates for the solutions of linear difference equations on the circle *Celestial Mech.* **14** 33–7

[Rüs76c] Rüssmann H 1976 Note on sums containing small divisors *Commun. Pure Appl. Math.* **29** 755–8

[Sal04] Salamon D 2004 The Kolmogorov–Arnold–Moser theorem *Math. Phys. Electron. J.* (3) 1–37

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

[Van02] Vano J 2002 A Whitney–Zehnder implicit function theorem *PhD Thesis* University of Texas at Austin

[Wei77] Weinstein A 1977 *Lectures on Symplectic Manifolds* (*Regional Conference Series in Mathematics* No 29) (Providence, RI: American Mathematical Society) (Expository lectures from *the CBMS Regional Conf. held at the University of North Carolina, 8–12 March, 1976*)

[WL98] Wojtkowski M P and Liverani C 1998 Conformally symplectic dynamics and symmetry of the Lyapunov spectrum *Commun. Math. Phys.* **194** 47–60

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

[Zeh76] Zehnder E 1976 Generalized implicit function theorems with applications to some small divisor problems. II *Commun. Pure Appl. Math.* **29** 49–111