Math. Z. **202** (1989) 559–608

On Elliptic Lower Dimensional Tori in Hamiltonian Systems

JÜRGEN PÖSCHEL

1 The problem

In this paper I describe a perturbation theory for *elliptic lower dimensional* invariant tori in hamiltonian systems of arbitrary dimension. These are tori in whose vicinity the linearized equations of motion possess an elliptic fixed point.

To be more precise, the unperturbed hamiltonian system is described by a real *normal form*

$$N = e + \sum_{i=1}^{n} \omega_i y_i + \tfrac{1}{2} \sum_{j=1}^{m} \Omega_j (u_j^2 + v_j^2),$$

defined on the phase space

$$\mathcal{P} = \mathbb{T}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m,$$

where $1 \le n < \infty$, $1 \le m \le \infty$, and $\mathbb{T}^n$ denotes the usual n-torus obtained from $\mathbb{R}^n$ by identifying coordinates modulo 2π. The associated symplectic form is

$$\sum_{i=1}^{n} dx_i \wedge dy_i + \sum_{j=1}^{m} du_j \wedge dv_j .$$

The equations of motion thus read

$$\begin{aligned} \dot{x} &= \omega \\ \dot{y} &= 0 \\ \dot{u} &= \varOmega v \\ \dot{v} &= -\varOmega u \end{aligned}$$

in usual vector notation, setting

$$\omega = (\omega_1, \ldots, \omega_n), \qquad \varOmega = (\varOmega_1, \ldots, \varOmega_m),$$

and identifying the latter with a diagonal matrix.

This unperturbed system features the invariant torus

$$\mathcal{T}_0 = \mathbb{T}^n \times \{0\} \times \{0\} \times \{0\}$$

carrying a quasi-periodic flow $x = \omega t + x_o$ with fixed *torus frequencies* ω. Its normal space is described by the uv-coordinates, whose origin is an elliptic fixed point with characteristic frequencies $\varOmega$ and $-\varOmega$, refered to as *normal frequencies*. The objective is to investigate the persistence of such an elliptic invariant torus under sufficiently small perturbations of the hamiltonian.

A necessary prerequisite of any such perturbation theory is a complete control over the torus frequencies ω. As our notation indicates, they are considered as *parameters* which vary freely over some n-dimensional domain and may be adjusted if necessary. This is tantamount to imposing a “nondegeneracy” or “anisochronicity” condition upon the unperturbed system—given such a nondegeneracy these frequencies are always available as parameters. They have the advantage to dispense with quadratic terms in y in the normal form thereby simplifying the formalism considerably.

The normal frequencies $\varOmega$, however, do *not* vary freely in general—they rather are functions of the torus frequencies ω. Still, if $m = 1$, this dependence may be resolved by a slight stretching of the time scale [14]. But if $m > 1$, then there are not sufficiently many further parameters available to adjust the normal frequencies. In particular, under perturbation they will not stay put in general.

This “coupling of frequencies” complicates the handling of the small divisor conditions arising in the theory. They are of the form

$$|\langle k, \omega \rangle - \langle l, \varOmega(\omega) \rangle| \geq \alpha \delta_k, \qquad |l| \leq 2, \quad |k| + |l| \neq 0,$$

where k and l are integer vectors, α is a positive parameter, and the δ_k converge to zero from above as $|k|$ tends to infinity. Due to the dependence of Ω on ω the question arises whether there are any torus frequencies ω that satisfy these infinitely many conditions for sufficiently small α.

In the metric theory of diophantine approximations this is known as the problem of approximating dependent quantities [27]. Geometrically speaking, it consists in finding badly approximable irrational points on an n-dimensional surface—the graph of the map $\omega \mapsto \Omega$—embedded in an $n+m$-dimensional ambient space. That is, this n-surface has to be intersected with the familiar Cantor set of badly approximable, *independent* points (ω, Ω) in $n+m$-space.

It turns out that such "good points" form a set of full relative n-dimensional Lebesgue measure provided the surface is sufficiently curved in one way or another. See [27], [4] and the references therein. There is no such result, however, for flat surfaces, which arise in our example in section 3. Indeed, there may be no "good points" at all if this surface happens to fall into a resonant hyperplane. Fortunately, such degeneracies are easily excluded by a few extra assumptions about the frequencies, since our small divisor conditions involve only low order combinations of the normal frequencies in view of the restriction $|l| \leq 2$ (whereas l is unrestricted in the general theory of diophantine approximations). Apart from that one may proceed as usual without refering to intricate results in number theory. Incidentally, all this is discussed independently of the perturbation theory itself. For that matter, we simply assume the nonresonance conditions to hold.

As another complication the small divisor estimates need to be preserved under small perturbations of the normal frequencies. This would be impossible if l were unrestricted. But again, since $|l| \leq 2$, this is feasable, for instance by taking into account only finitely many resonances at each step of the iterative procedure described below—an idea already used by Arnold in [1]. Then only finitely many nonresonance conditions are involved at a time, and they are preserved under sufficiently small perturbations simply by decreasing the parameter α a bit.

A perturbation theory for elliptic invariant tori was first announced by Melnikov in the sixties [11,12], but no proofs have ever been given. Meanwhile, Moser [14] proved the existence of quasi-periodic solutions for a large class of ordinary differential equations that included the case of a hamiltonian invariant torus with a *two*-dimensional (that is, one degree of freedom) elliptic fixed point. Higher codimensions, however, were not accessible by his method, since he insisted on *all* frequencies to be fixed. It took another eigtheen years before Eliasson [5] succeeded in removing this restriction by allowing the elliptic frequencies to undergo small perturbations. The details of his proof, however, are a bit unwieldy and hard to comprehend. Rüssmann [23] subsequently simplified his arguments, for instance by introducing the torus frequencies as independent parameters.

In [14] Moser also considered the case of a *hyperbolic* invariant torus, or "whiskered torus" in Arnold's terminology. Assuming all hyperbolic eigenvalues to be real and distinct, no restriction needed to be placed on their number. Later on, his result was extended by Graff [7] and Zehnder [35] to allow for hyperbolic fixed points of any kind. This is another instance of the general phenomenon of the stability of unstable structures. Undoubtedly, the present discussion may be extended to include hyperbolic directions, but the author didn't do this for the sake of brevity and clarity.

The present paper is based on Rüssmann's approach and grew out of an attempt to further improve on it. In particular, while all the results so far were restricted to systems with finitely many degrees of freedom, the theory presented here includes the case of a finite dimensional invariant torus embedded in an *infinite* dimensional hamiltonian system. For instance, the theory can be made to apply to the nonlinear wave equation

$$u_{tt} - u_{xx} + qu + u^3 = 0$$

on the strip $0 \le x \le \pi$, $-\infty < t < \infty$ with zero boundary conditions. For proper choices of the potential q there exist nontrivial solutions near the trivial zero solution that are quasi-periodic in time with any finite number of basic frequencies.

A similar result was presented by Wayne in a talk at Oberwolfach in May 1987 and appeared in written form recently [33]. Indeed, although quite different in technical aspects, his approach provided the inspiration to extend Eliasson's result to systems with infinitely many degrees of freedom.

After revising this paper the author learned about the work of Kuksin [10]. He independently generalized the work of Melnikov to infinite dimensional systems and applied it to a wave equation with a nonlinearity $\varepsilon f(u, \varepsilon)$, which is real analytic in u and vanishes for $u = 0$. Under a generic regularity condition he obtained quasi-periodic solutions u that persist for sufficiently small ε.

There are other extension of the KAM-theory to infinite dimensions, too. In the seventies Siegel's famous results on the linearization of complex analytic vectorfields in the neighbourhood of a fixed point [24,25] were extended by Ware [32], Zehnder [36] and others. Further generalizations to evolution equations in Banach spaces are due to Nikolenko and others—see [16] and the references therein—including nonlinear heat and Schrödinger equations (whereas the reference to the nonlinear wave equation in [16] is not correct).

More recently, some infinite dimensional hamiltonian systems were studied by Vittot and Bellissard [31] and Fröhlich, Spencer, Wayne [6] modelling chains of weakly coupled rotators and random anharmonic crystal lattices. Here, *infinite*-dimensional invariant tori are found that are strongly localized in space (and may therefore be considered as "quasi-finite-dimensional"). Such restrictions are necessary to conquer the difficulty of controlling small divisors that involve arbitrary integer combinations of *infinitely* many independent frequencies. For a general approach to such problems see also [19,20].

Besides that, the classical KAM theorem has advanced, too. Picking up an idea of Kozlov [9] and applied by Moser [15] to minimal foliations, Salamon and Zehnder [29] presented a new proof working with a variational formulation in configuration space. Celletti and Chierchia [3] were able to combine this approach with a finite number of computer aided estimates to obtain very realistic, close to optimal bounds for the admissable size of perturbations of some particular structure. It is conceivable that similar techniques may also improve the smallness condition of our main theorem.

At this point it is a pleasure to thank a number of people for their interest and eucouragement, for their support or simply for pointing out some mistakes—Maurice Dodson, Håkan Eliasson, Martin Kummer, Jürgen Moser, Gene Wayne and Edi Zehnder.

2 Statement of the result

Introducing complex conjugate variables

$$z = (u + iv)/\sqrt{2}, \quad \bar{z} = (u - iv)/\sqrt{2}$$

with associated symplectic form $\sum dx_i \wedge dy_i + i \sum dz_j \wedge d\bar{z}_j$, we can write

$$N = e + \langle \omega, y \rangle + \langle \Omega z, \bar{z} \rangle .$$

These coordinates are more convenient to use in the following.

We consider normal forms whose torus frequencies ω are regarded as parameters varying over a closed subset $\mathcal{O}$ of $\mathbb{R}^n$ of positive Lebesgue measure. Its normal frequencies Ω are functions of ω that are assumed to be real analytic on some complex neighbourhood

$$\mathcal{W}_h: \quad |\omega - \mathcal{O}|_\infty < h$$

of $\mathcal{O}$. Its Jacobian as a linear map from $\mathbb{R}^n$ into $\mathbb{R}^m$ is required to be uniformly bounded with respect to the operator norm $\| \cdot \|_\infty$ induced by the sup-norms on these spaces.

Usually, the energy e is also a function of ω. The dynamical properties of a hamiltonian system, however, are independent of this energy function. We therefore ignore it in the following discussion.

We consider perturbations

$$H = N + P$$

of N that are real analytic on some complex neighbourhood

$$\mathcal{D}_{r,s}: \quad |\operatorname{Im} x|_\infty < r, \quad |y|_1 < s^2, \quad |z|_2\,, |\bar z|_2 < s$$

of the torus $\mathcal{T}_0$ and are also real analytic in ω over $\mathcal{W}_h$. The norms are

$$|x|_\infty = \max_{1\le i\le n} |x_i|\,, \qquad |y|_1 = \sum_{i=1}^n |y_i|$$

and the euclidean norm $|\cdot|_2$. These different norms reflect the fact that x and y refer to *symplectic polar* coordinates, while z and $\bar z$ do not. Essentially, $\mathcal{D}_{r,s}$ is a ball of radius s in a complex space of dimension $2n+2m$ with respect to the euclidean norm.

At this point we ought to explain what we mean by a real analytic function in $z, \bar z$ when m is infinite. For our purposes there is no significant difference to the finite dimensional situation. Analytic functions are infinitely often differentiable and represented by their Taylor series expansion. We refer the reader to Appendix A of [21] for more details.

The size of P is measured in a weighted norm that is adapted to the kind of small divisor problems under consideration. Taking the Fourier series expansion

$$P = \sum_{k\in\mathbb{Z}^n} P_k e^{i\langle k,\theta\rangle},$$

let

$$\|P\|_{r,s,h} = \sum_k |\mathbf{M}P_k|_{s,h}\, e^{|k|r},$$

where $\mathbf{M}P_k$ denotes a certain majorant of P_k, and $|\cdot|_{s,h}$ denotes the sup-norm over the $y, z, \bar z$ and ω domains of $\mathcal{D}_{r,s}$ and $\mathcal{W}_h$ respectively. The precise definition is given in Appendix C.

A similar norm was introduced by Vittot [30] and Vittot, Bellissard [31], but no majorant was involved, since there were no normal coordinates. The majorant also appears in the work of Wayne [33] and Nikolenko [16] and seems to arise naturally in infinite dimensional settings.

The nonresonance conditions are

$$|\langle k, \omega\rangle - \langle l, \Omega(\omega)\rangle| \ge \alpha\delta_k, \qquad |l| \le 2, \quad |k| + |l| \ne 0, \tag{1}$$

where k and l are in $\mathbb{Z}^n$ and $\mathbb{Z}^m$ respectively, and $|\cdot|$ denotes the sum-norm of an integer vector. These conditions depend on a positive parameter α and some fixed *approximation function* Δ via the definition

$$\delta_k = \frac{1}{\Delta(|k|)}, \qquad k \in \mathbb{Z}^n.$$

Their effect in the perturbation problem is measured by a function Ψ that is defined on the positive real axis entirely in terms of Δ. See Appendix A for definitions, details and examples.

Theorem A. *Suppose N is a normal form whose normal frequencies Ω are real analytic on the complex neighbourhood $\mathcal{W}_h$ of $\mathcal{O}$ and satisfy*

$$\left\| \frac{\partial \Omega}{\partial \omega} \right\|_\infty \le M < \infty.$$

Suppose that H is a perturbation of N that is real analytic on $\mathcal{D}_{r,s} \times \mathcal{W}_h$ and satisfies

$$\frac{1}{s^2} \|H - N\|_{r,s,h} \le \frac{\varepsilon_*}{M+1} \cdot \frac{\alpha}{\Psi(\rho)} \le \frac{h}{16}$$

for some $0 < \rho < r/2$, where ε_* is an absolute positive number to be specified later.

Then there exist a Cantor set $\mathcal{O}_\alpha \subset \mathcal{O}$, a normal form N_* on $\mathcal{O}_\alpha$ such that

$$|\langle k, \omega \rangle - \langle l, \Omega_*(\omega) \rangle| \ge \alpha \delta_k / 2, \qquad |l| \le 2, \quad |k| + |l| \ne 0$$

on $\mathcal{O}_\alpha$, and a transformation

$$\mathcal{F}\colon \quad \mathcal{D}_{r-2\rho,s/2} \times \mathcal{O}_\alpha \to \mathcal{D}_{r,s} \times \mathcal{W}_h$$

of the form described in section 4, that is real analytic and symplectic for each ω and Whitney smooth in ω, such that

$$H \circ \mathcal{F} = N_* + R_*$$

with

$$\partial_y^l \partial_z^p \partial_{\bar{z}}^q R_* = 0, \qquad 2\,|l| + |p+q| \le 2.$$

Thus the perturbed system possesses elliptic invariant tori with nonresonant frequencies ω, Ω_* for each frequency vector in $\mathcal{O}_\alpha$.

There are also estimates for N_* and $\mathcal{F}$, which are given at the end of the proof of Theorem A. For the notion of Whitney differentiability and Whitney smoothness of functions on closed sets the reader is refered to [34,28,17]. For the sake of simplicity, however, only uniform continuity in ω is proven.

We obtain $\varepsilon_* = 2^{-37}$, but no effort was undertaken to improve that constant. For the classical KAM-theorem a much better bound was given by Rüssmann in [22] though no written proof has appeared. For quite realistic bounds for a special class of perturbations see the aforementioned paper by Celletti and Chierchia [3].

The question remains under which circumstances the Cantor set $\mathcal{O}_\alpha$ is not empty. Without further assumptions, there may be no nonresonant frequencies no matter how small α is chosen. The following paragraphs describe a particularly simple criterion. More has to be done to discuss the nonlinear wave equation.

We introduce the following notion. The closed set $\mathcal{O}$ is called *essentially nonresonant*, if the following two conditions are satisfied for some positive α. *Nonresonance*: for every l with $0 < |l| \le 2$,

$$\min_{\omega\in\mathcal{O}} |\langle k, \omega\rangle - \langle l, \Omega(\omega)\rangle| \ge \alpha$$

for all $k \in \mathcal{K}_l$, the closed convex hull of the gradient set

$$\mathcal{G}_l = \{\, \partial_\omega \langle l, \Omega(\omega)\rangle : \omega \in \mathcal{O} \,\}.$$

Finiteness: the cardinality of the set

$$\mathcal{L}_k = \left\{ l : \min_{\omega\in\mathcal{O}} |\langle k, \omega\rangle - \langle l, \Omega(\omega)\rangle| \le \alpha \right\} \subset \{\, l : |l| \le 2 \,\}$$

grows not faster with $|k|$ than a constant multiple of some approximation function.

The finiteness condition is important only in the infinite dimensional case, otherwise $|\mathcal{L}_k|$ is uniformly bounded. The nonresonance condition, in contrast, is relevant also for small dimensions m. Still, by the hypotheses of Theorem A each of the sets $\mathcal{K}_l$ is contained in the closed ball of radius $2M$ around the origin. For each of the finitely many k's in that ball there are only finitely many l for which the resonant term may be smaller than α by the finiteness condition. Hence, the first condition is "finite", involving only finitely many resonances.

Theorem B. *Suppose the frequency set $\mathcal{O}$ is essentially nonresonant. If the approximation function is chosen such that*

$$\sum_k \frac{|\mathcal{L}_k|}{\Delta(|k|)} < \infty,$$

and if the smallness condition of Theorem A holds with $M+1$ *multiplied by* $e^{M\rho}$, *then*

$$m(\mathcal{O}-\mathcal{O}_\alpha) = O(\alpha d^{n-1}),$$

where d *denotes the exterior diameter of* $\mathcal{O}$ *(with respect to any norm).*

Hence, if the set $\mathcal{O}$ has positive Lebesgue measure, then the Cantor set $\mathcal{O}_\alpha \subset \mathcal{O}$ is not empty for all sufficiently small α.

The classical KAM-theorem

The classical KAM-theorem in the analytic case may be considered as a special case of the preceding results by allowing the dimension m to be zero. In that "degenerate" case the weighted norm no longer involves majorants, and we have

$$\|P\|_{r,s,h} = \sum_k |P_k|_{s,h}\, e^{|k|r},$$

which is almost Vittot's norm in [30]. In particular,

$$|P|_{r,s,h} \le \|P\|_{r,s,h} \le \coth^n\rho \; |P|_{r+2\rho,s,h}$$

for $\rho > 0$, where $|\cdot|$ denotes the usual sup-norm.

Moreover, every parameter set $\mathcal{O}$ is essentially nonresonant, and the Cantor set $\mathcal{O}_\alpha$ simply consists of all ω in $\mathcal{O}$ such that

$$|\langle k, \omega\rangle| \ge \alpha\delta_k, \qquad 0 \ne k \in \mathbb{Z}^n,$$

since there are no normal frequencies. In particular, this set is known *a priori*. Finally, the smallness conditions of the two theorems are identical when m is zero. Thus we have

Theorem C (The classical KAM-Theorem). *Suppose the Hamiltonian* H *is a perturbation of the normal form* $N = e + \langle\omega, y\rangle$ *that is real analytic on* $\mathcal{D}_{r,s} \times \mathcal{W}_h$ *and satisfies*

$$\frac{1}{s^2}\|H-N\|_{r,s,h} \le \frac{\alpha\varepsilon_*}{\Psi(\rho)} \le \frac{h}{16}$$

for some $0<\rho<r/2$, *where* ε_* *is an absolute positive number. Then there exists a transformation*

$$\mathcal{F}\colon \quad \mathcal{D}_{r-2\rho,s/2} \times \mathcal{O}_\alpha \to \mathcal{D}_{r,s} \times \mathcal{W}_h$$

of a form analogous to the one described in section 4, that is real analytic and symplectic for each ω *in* $\mathcal{O}_\alpha$ *and Whitney smooth in* ω *such that*

$$H \circ \mathcal{F} = e_* + \langle \omega, y \rangle + \dots,$$

where the dots denote terms of higher order in y*. Thus there exists a foliation of invariant tori over the Cantor set* $\mathcal{O}_\alpha$ *for the perturbed system. Moreover,*

$$m(\mathcal{O} - \mathcal{O}_\alpha) = O(\alpha d^{n-1}),$$

where d *denotes the exterior diameter of* $\mathcal{O}$ *(with respect to any norm).*

Note that in this context,

$$\mathcal{D}_{r,s}: \quad |\mathrm{Im}\, x|_\infty < r, \quad |y|_1 < s^2,$$

so the s^2 in the smallness condition is just the radius of the y-domain, not its square.

Note also that in the smallness condition of Theorem C the small divisors "enter only once" via the expression α/Ψ. This is different from other versions of the KAM-theorem such as in [22], where this term is squared. This is due to regarding the frequencies as independent parameters thereby decoupling them from the hamiltonian systems itself. More technically speaking, in the linearized problem there is only one small divisor equation to be solved. Of course, this is not a genuine improvement, and the square is restored when reducing the traditional versions of this theorem to the one above.

3 An Example

We describe an application of the preceding results to an elliptic fixed point of a hamiltonian system with a *finite* number of degrees of freedom. Placing the fixed point at the origin of the coordinate system, we assume that

$$H = \tfrac{1}{2} \sum_{k=1}^{d} \lambda_k (q_k^2 + p_k^2) + \dots, \tag{2}$$

the dots denoting terms of higher order. Thus, we assume that the linear part of the equations of motion,

$$\dot{q}_k = \lambda_k p_k + \dots, \quad \dot{p}_k = -\lambda_k q_k + \dots,$$

can be written as a system of d decoupled oscillators with characteristic frequencies

$$\lambda_1, \dots, \lambda_d, -\lambda_1, \dots, -\lambda_d.$$

This is always possible, for instance, if all characteristic frequencies are distinct, an assumption included in the nonresonance condition (3) below.

The objective is to find quasi-periodic motions of, say, the first n oscillators, $1 \le n < d$, filling a lower dimensional invariant torus, while keeping the remaining $m = n - d \ge 1$ oscillators at rest in suitable coordinates. By comparison, the classical KAM-theorem deals with quasi-periodic motions of *all* oscillators, filling an invariant torus of maximal dimension d.

Let

$$\alpha = (\lambda_1, \dots, \lambda_n), \qquad \beta = (\lambda_{n+1}, \dots, \lambda_d),$$
$$Y = (I_1, \dots, I_n), \qquad Z = (I_{n+1}, \dots, I_d),$$

where

$$I_k = \tfrac{1}{2}(q_k^2 + p_k^2), \qquad 1 \le k \le d.$$

Then

$$H = \langle \alpha, Y \rangle + \langle \beta, Z \rangle + P_3 + P_4 + \dots,$$

where P_3, P_4 denote monomials of order three and four respectively.

It is a well known fact [26] that these monomials may be put into Birkhoff normal form by a symplectic change of coordinates, provided the frequencies α, β are nonresonant up to order four. For our purposes, however, a slightly more general normal form suffices, which allows for some low order resonances. Precisely, assume that

$$\langle k, \alpha \rangle + \langle l, \beta \rangle \neq 0, \qquad 1 \le |k| + |l| \le 4, \quad |l| \le 3. \tag{3}$$

Then one can choose symplectic coordinates in such a way that

$$P_3 \equiv 0, \quad P_4 = \tfrac{1}{2}\langle AY, Y \rangle + \langle BY, Z \rangle + Q(\hat{q}, \hat{p})$$

with a symmetric $n \times n$-matrix A, an $n \times m$-matrix B and a monomial Q of fourth order in $\hat{q} = (q_{n+1}, \dots, q_d)$ and $\hat{p} = (p_{n+1}, \dots, p_d)$. The coefficients of A and B are independent of the particular choice of coordinates.

Our aim is to perturb the tori

$$Y = Y_0 > 0, \quad Z = 0$$

in arbitrarily small neighbourhoods of the origin (the inequality is understood componentwise). This is conveniently done by first shrinking all coordinates by a factor $\varepsilon > 0$ and dividing the resulting hamiltonian by ε^2, thereby introducing a fixed domain to work on and preserving the symplectic structure. Staying with the same notation as before we obtain

$$\begin{aligned} H = &\langle \alpha, Y \rangle + \langle \beta, Z \rangle \\ &+ \varepsilon^2 \big(\tfrac{1}{2} \langle AY, Y \rangle + \langle BY, Z \rangle + Q(\hat{q}, \hat{p}) \big) \\ &+ \varepsilon^3 R(q, p, \varepsilon), \end{aligned}$$

where the functions Q and R are uniformly bounded, say, on the fixed *complex* polydisc

$$I_k = \tfrac{1}{2}(q_k^2 + p_k^2) < 2, \quad 1 \le k \le d$$

for all sufficiently small $\varepsilon > 0$.

Now introduce symplectic polar coordinates x, y around each of these tori in the usual fashion such that $Y = Y_0 + y$. We obtain

$$H = \langle \omega, y \rangle + \langle \Omega, Z \rangle + S(x, y, \hat{q}, \hat{p}; Y_0, \varepsilon),$$

where

$$\omega = \alpha + \varepsilon^2 A Y_0, \quad \Omega = \beta + \varepsilon^2 B Y_0$$

and

$$S = \varepsilon^2 \big(\tfrac{1}{2} \langle Ay, y \rangle + \langle By, Z \rangle + Q(\hat{q}, \hat{p}) \big) + \varepsilon^3 R.$$

Restricting their initial positions to a "positive" domain such as

$$\mathcal{Y}_\delta: \quad Y_0 > \delta, \quad |Y_0|_1 < 1 - \delta,$$

$0 < \delta < \frac{1}{2}$, this hamiltonian is real analytic on any domain

$$\mathcal{D}_{r,s}: \quad |\operatorname{Im} x|_\infty < r, \quad |y|_1 < s^2, \quad \left|\hat{q}\right|_2, \left|\hat{p}\right|_2 < s,$$

where r and s are sufficiently small. In particular, we may fix some r and choose

$s^4 = \varepsilon$ to obtain

$$|S|_{r,s} = O\big(\varepsilon^2 s^4 + \varepsilon^3\big) = O\big(\varepsilon^3\big)\,.$$

Since everything is finite-dimensional, a similar estimate holds for the weighted norm $\|S\|_{r,s}$, so

$$\frac{1}{s^2}\|S\|_{r,s} = O\big(\varepsilon^{5/2}\big)$$

uniformly in Y_0. Indeed, we may even choose $\delta = 2\sqrt{\varepsilon}$ without affecting these estimates.

Now consider the torus and normal frequencies, ω and Ω. Imposing the nondegeneracy condition

$$\det A \neq 0, \tag{4}$$

there is a one-to-one correspondence between the positions Y_0 and the frequencies ω, and the latter may be introduced as parameters instead. In particular, we obtain Ω as an affine function of ω,

$$\Omega = \beta + BA^{-1}(\omega - \alpha),$$

that happens to be independent of ε. Moreover,

$$\left|\frac{\partial\Omega}{\partial\omega}\right|_\infty = \big|BA^{-1}\big|_\infty = \text{const} < \infty.$$

Thus, the two terms $\langle\omega, y\rangle + \langle\Omega, Z\rangle$ qualify as a real analytic normal form being perturbed by the function S.

The ω-domain, however, does depend on ε. It is the image of $\mathcal{Y}_\delta$ under the affine map $Y \mapsto \alpha + \varepsilon^2 AY$, hence comprises a complex neighbourhood $\mathcal{W}_h$ of radius

$$h = \varepsilon^2 h_0, \qquad h_0 = |A^{-1}|^{-1}$$

around a real domain $\mathcal{O}^\varepsilon$ of linear size proportional to ε^2. Hence it suffices to choose

$$\alpha = \varepsilon^{5/2} a_0$$

with a_0 sufficiently *large* in order to satisfy the smallness condition of Theorem A for all sufficiently small ε.

Finally, it's easy to make explicit the conditions under which the domain $\mathcal{O}^\varepsilon$ is essentially nonresonant. For each integer vector l with $|l| \leq 2$ the set $\mathcal{K}_l$ consists of the single point

$$\frac{\partial}{\partial\omega}\langle l, \Omega(\omega)\rangle = A^{-1}B^t l.$$

There is nothing to check unless $A^{-1}B^t l = k$ is an integer vector, in which case

$$\langle k, \omega\rangle - \langle l, \Omega(\omega)\rangle = \left\langle l, BA^{-1}\alpha - \beta\right\rangle$$

is independent of ω. It suffices to require that

$$\left\langle l, BA^{-1}\alpha - \beta\right\rangle \neq 0 \tag{5}$$

for all l with $|l| \leq 2$ such that $A^{-1}B^t l$ is an integer vector. In particular, by our choice of α,

$$m(\mathcal{O}^\varepsilon - \mathcal{O}^\varepsilon_\alpha) = O(\sqrt{\varepsilon})\, m(\mathcal{O}^\varepsilon)$$

by Theorem B for all sufficiently small ε.

Summarizing, if conditions (2–5) hold, then in every neighbourhood of the origin there exists a family of elliptic n-dimensional invariant tori parametrized smoothly over an n-dimensional Cantor set of nonresonant frequencies. Their relative n-dimensional Lebesgue measure in a ball of radius ε is $1 - \sqrt{\varepsilon}$.

4 Outline of the proof

We now give an outline of the proof of Theorem A. Its details are given in the next three sections. Section 8 contains the proof of Theorem B.

Theorem A is proven by the familiar KAM-method employing a rapidly converging iteration scheme [8,1,13]. At each step of the scheme, a hamiltonian

$$H_n = N_n + P_n$$

is considered, which is a small perturbation of some normal form N_n. A transformation $\mathcal{F}_n$ is set up so that

$$H_n \circ \mathcal{F}_n = N_{n+1} + P_{n+1}$$

with another normal form N_{n+1} and a much smaller error term P_{n+1}. For instance,

$$\|P_{n+1}\| \leq C_n \|P_n\|^\kappa$$

for some $\kappa > 1$. This transformation consists of a symplectic change of coordinates Φ_n and a subsequent change φ_n of the parameters ω and is found by linearising the above equation. Repetition of this process leads to a sequence of transformations $\mathcal{F}_0, \mathcal{F}_1, \ldots$, whose infinite product transforms the initial hamiltonian H_0 into a normal form N_* up to a certain order.

To describe the construction in more detail, let us drop the index n to simplify notation. First we write

$$\begin{aligned} H &= N + P \\ &= N + R + (P - R), \end{aligned}$$

where R is obtained from P by truncating its Fourier and Taylor series expansion in a suitable way.

The coordinate transformation Φ is written as the time-1-map of the flow X_F^t of a hamiltonian vectorfield X_F:

$$\Phi = X_F^t\big|_{t=1}.$$

Then Φ is symplectic. Moreover, we may expand $H \circ \Phi = H \circ X_F^t\big|_{t=1}$ with respect to t at 0 using Taylor's formula. Recall that

$$\frac{d}{dt} G \circ X_F^t = \{G, F\} \circ X_F^t,$$

the Poisson bracket of G and F evaluated at X_F^t. Thus we may write

$$\begin{aligned} (N + R) \circ \Phi &= N \circ X_F^t\big|_{t=1} + R \circ X_F^t\big|_{t=1} \\ &= N + \{N, F\} + \int_0^1 (1 - t) \{\{N, F\}, F\} \circ X_F^t \, dt \\ &\qquad + R + \int_0^1 \{R, F\} \circ X_F^t \, dt \\ &= N + R + \{N, F\} \\ &\qquad + \int_0^1 \{(1 - t)\{N, F\} + R, F\} \circ X_F^t \, dt. \end{aligned}$$

The latter integral is of quadratic order in R and F and will be part of the new error term.

The point is to find F such that

$$N + R + \{N, F\} = N_+$$

is again a normal form. Equivalently, setting $N_+ = N + \hat{N}$, this amounts to solving the linear equation

$$\{F, N\} + \hat{N} = R \tag{6}$$

for F and $\hat{N}$, when R is given. Suppose such a solution exists. Then

$$(1-t)\{N, F\} + R = (1-t)\hat{N} + tR,$$

and hence

$$\begin{aligned} H \circ \Phi &= N_+ + P_+ \\ &= N_+ + Q + (P - R) \circ \Phi \end{aligned}$$

with

$$Q = \int_0^1 \{(1-t)\hat{N} + tR, F\} \circ X_F^t \, dt.$$

Q is of quadratic order in R, F and $\hat{N}$.

The solution to equation (6) is determined by the properties of the linear operator $L\colon F \mapsto \{F, N\}$. For

$$F = e^{i\langle k, \theta\rangle} y^l z^p \bar{z}^q,$$

one easily finds

$$LF = i(\langle k, \omega\rangle + \langle p - q, \Omega\rangle) \cdot F.$$

Hence each such function is an eigenfunction of L. We see that L is *diagonizable* with eigenvalues $i(\langle k, \omega\rangle + \langle p - q, \Omega\rangle)$ and a complete set of eigenfunctions $e^{i\langle k, \theta\rangle} y^l z^p \bar{z}^q$. Its domain splits into two invariant subspaces, its nullspace $\mathcal{N}$ and its range $\mathcal{R}$, on which L is invertible. If

$$R = R_{\mathcal{N}} + R_{\mathcal{R}}$$

is the corresponding decomposition of R, one may therefore choose

$$\hat{N} = R_{\mathcal{N}},$$

the projection of R onto $\mathcal{N}$. Indeed, this is the *smallest* space one could choose $\hat{N}$ from. Then it suffices to solve

$$LF = R_{\mathcal{R}}$$

uniquely in $\mathcal{R}$. Any two solutions of that equation differ only by an element in $\mathcal{N}$, so some normalization like

$$F_{\mathcal{N}} = 0$$

fixes a unique solution.

In effect, if

$$R = \sum_{k,l,p,q} R_{klpq}\, e^{i\langle k,\theta\rangle} y^l z^p \bar{z}^q$$

then

$$\hat{N} = \sum_{\langle k,\omega\rangle + \langle p-q,\Omega\rangle = 0} R_{klpq}\, e^{i\langle k,\theta\rangle} y^l z^p \bar{z}^q$$

and

$$F = \sum_{\langle k,\omega\rangle + \langle p-q,\Omega\rangle \neq 0} \frac{-i\, R_{klpq}}{\langle k,\omega\rangle + \langle p-q,\Omega\rangle}\, e^{i\langle k,\theta\rangle} y^l z^p \bar{z}^q .$$

The truncation R of P will be chosen so that these sums extend over

$$|k| \le K, \qquad 2\,|l| + |p+q| \le 2.$$

Of course, all coefficients still depend on ω as usual.

At this point it is essential that

$$\langle k,\omega\rangle + \langle p-q,\Omega\rangle = 0 \quad \Leftrightarrow \quad k = 0 \wedge p = q$$

by the nonresonance conditions. This makes the nullspace $\mathcal{N}$ of L as small as it possibly can be, so that

$$\hat{N} = \hat{e} + \langle v(\omega), y\rangle + \langle V(\omega) z, \bar{z}\rangle$$

with a *diagonal* matrix V. It suffices to change parameters by setting

$$\omega_+ = \omega + v(\omega) \tag{7}$$

to obtain a normal form $N_+ = N + \hat{N}$ which is of the same form as N. This completes one cycle of the iteration.

We conclude this section with a remark concerning the structure of the maps Φ and $\mathcal{F}$. By construction,

$$F = \sum_{|k|\le K} \sum_{2|l|+|p+q|\le 2} F_{klpq}\, e^{i\langle k,\theta\rangle} y^l z^p \bar{z}^q .$$

It follows that $\Phi = X_F^t\big|_{t=1}$ is of the form

$$\begin{aligned} x' &= U(x) \\ y' &= V(x,y,z,\bar z) \\ z' &= W(x,z,\bar z) \\ \bar z' &= \bar W(x,z,\bar z), \end{aligned}$$

where W and $\bar W$ are first order polynomials in $z, \bar z$, and V is of the same form as F. Again, the dependence of all coefficients on ω has been suppressed. This map is composed with the inverse φ of the parameter map (7) to obtain $\mathcal{F}$.

Such transformations form a *group* under composition. Hence, if $\mathcal{F}_0, \mathcal{F}_1, \ldots,$ $\mathcal{F}_n$ belong to this group, then so does their composition $\mathcal{F}_0 \circ \mathcal{F}_1 \circ \ldots \circ \mathcal{F}_n$ and its limit $\mathcal{F}$ for $n \to \infty$.

The formal aspects of the procedure outlined here are appropriately discussed within the framework of Lie algebras of vectorfields and their transformation groups. We won't go into this here, but refer the reader for example to [14].

5 The KAM-step

We begin by describing a single step of the iterative construction. In view of a stretching of the frequencies carried out later the nonresonance condition are assumed to hold with a *large* parameter A.

Suppose

$$N = e + \langle \omega, y \rangle + \langle \Omega z, \bar z \rangle$$

is a normal form whose frequencies satisfy

$$|\langle k, \omega \rangle - \langle l, \Omega(\omega) \rangle| \ge A\delta_k \tag{9}$$

on some closed subset $\mathcal{O}$ of $\mathbb{R}^n$ for $|k| \le K$ and $|l| \le 2$ not both zero. Moreover, suppose the frequencies Ω are real analytic functions of ω on the complex neighbourhood

$$\mathcal{W} = \mathcal{W}_h\colon \quad |\omega - \mathcal{O}|_\infty < h$$

of radius h around $\mathcal{O}$ such that

$$\left| \frac{\partial \Omega}{\partial \omega} \right|_\infty \le M < \infty \tag{10}$$

on this domain. The perturbation $H = N + P$ of N is assumed to be real analytic on the complex domain $\mathcal{D} \times \mathcal{W}$, where

$$\mathcal{D} = \mathcal{D}_{r,s}\colon \quad |\mathrm{Im}\, x|_\infty < r, \quad |y|_1 < s^2, \quad |z|_2\,, |\bar{z}|_2 < s,$$

such that

$$\|H - N\|_{r,s,h} \le \varepsilon/2$$

is sufficiently small with respect to the weighted norm. The precise conditions will be given later.

Unless stated otherwise, the following estimates are uniform with respect to ω in $\mathcal{W}$. Therefore the index h is usually dropped.

Approximating the perturbation

First, the Fourier and Taylor series expansion of $P = H - N$ is truncated in such a way that the remaining series R only contains those finitely many monomials

$$e^{i\langle k,\theta\rangle} y^l z^p \bar{z}^q$$

that satisfy

$$|k| \le K, \qquad 2\,|l| + |p + q| \le 2.$$

To this end, from the Fourier series expansion

$$P = \sum_k P_k e^{i\langle k,\theta\rangle}$$

all terms P_k with $|k| > K$ are discarded, and the Taylor series of the remaining coefficients with respect to $y, z, \bar{z}$ are truncated appropriately. It is easy to check that

$$\|R\|_{r,s} \le 2\|P\|_{r,s} \le \varepsilon.$$

Moreover, on any smaller domain $\mathcal{D}_{r-\rho,\alpha s}$ with $0 < \rho < r$ and $0 < \alpha < 1$ the estimate

$$\|P - R\|_{r-\rho,\alpha s} \le \left(e^{-K\rho} + \frac{\alpha^3}{1-\alpha^2} \right) \|P\|_{r,s} \tag{11}$$

holds.

Extending the small divisor estimate

The small divisor conditions (9) are assumed to hold on the real set $\mathcal{O}$ only. They are now extended to its complex neighbourhood $\mathcal{W}$ for h sufficiently small. Precisely, if

$$h \le \frac{A-1}{2M+1} \cdot \frac{1}{K\varDelta(K)}, \tag{12}$$

then

$$|\langle k, \omega \rangle - \langle l, \varOmega(\omega) \rangle| \ge \delta_k \tag{13}$$

uniformly on $\mathcal{W}$ for $|k| \le K$ and $|l| \le 2$ not both zero.

The proof is simple. Given ω in $\mathcal{W}$ there exists an ω_o in $\mathcal{O}$ such that $|\omega - \omega_o|_\infty < h$. The difference of the small divisor expressions for ω and ω_o is bounded by

$$\begin{aligned} &|\langle k, \omega - \omega_o \rangle - \langle l, \varOmega(\omega) - \varOmega(\omega_o) \rangle| \\ &\qquad \le |k|_1 \, |\omega - \omega_o|_\infty + |l|_1 \, |\varOmega(\omega) - \varOmega(\omega_o)|_\infty \\ &\qquad \le Kh + 2Mh \\ &\qquad \le (A-1)\varDelta^{-1}(K) \end{aligned}$$

in view of assumptions (10) and (12). Using the small divisor estimates for ω_o and the inequality

$$\frac{1}{\varDelta(K)} \le \delta_k, \qquad |k| \le K,$$

the claim follows.

Solving the linearized equation

The linearized equation $\{F, N\} + \hat{N} = R$ is now solved for F and $\hat{N}$ as described in the preceding section. Due to the truncation of P and estimate (13) the relevant eigenvalues $i(\langle k, \omega \rangle + \langle p - q, \varOmega \rangle)$ of the operator L vanish on $\mathcal{W}$ if and only if both $k = 0$ and $p = q$. Hence,

$$\hat{N} = \hat{e} + \langle v(\omega), y \rangle + \langle V(\omega)z, \bar{z} \rangle \tag{14}$$

with a diagonal matrix V. The estimate

$$\|\hat{N}\|_{r,s} \le \|R\|_{r,s} \le \varepsilon \tag{15}$$

is straightforward.

The Fourier coefficients of the hamiltonian F are

$$F_k = \sum_{l,p,q} \frac{-i R_{klpq}}{\langle k, \omega \rangle + \langle p - q, \Omega \rangle} y^l z^p \bar{z}^q.$$

By (13),

$$|\mathbf{M}F_k|_s \le \delta_k^{-1} \left| \sum_{l,p,q} \left| R_{klpq} y^l \right| z^p \bar{z}^q \right|_s \le \delta_k^{-1} \, |\mathbf{M}R_k|_s \, .$$

Hence,

$$\begin{aligned} \|F\|_{r-\rho,s} &= \sum_{|k|\le K} |\mathbf{M}F_k|_s \, e^{|k|(r-\rho)} \\ &\le \sum_k \delta_k^{-1} e^{-|k|\rho} \, |\mathbf{M}R_k|_s \, e^{|k|r} \\ &\le \sup_{t\ge 0} \, \Delta(t) e^{-\rho t} \|R\|_{r,s} \le \Gamma_0 \varepsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} \sum_i \left\| F_{x_i} \right\|_{r-\rho,s} &\le \sum_k |k| \, |\mathbf{M}F_k|_s \, e^{|k|(r-\rho)} \\ &\le \sup_{t\ge 0} \, t \Delta(t) e^{-\rho t} \|R\|_{r,s} \le \Gamma_1 \varepsilon. \end{aligned}$$

Here,

$$\Gamma_k(\rho) = \sup_{t\ge 0} \, (1+t)^k \Delta(t) e^{-\rho t}, \qquad k = 0, 1.$$

Since $\Gamma_0 \le \Gamma_1$ by definition and $\Gamma_0 \le \rho \Gamma_1$ by Lemma 5, these two estimates can be combined into

$$\frac{1}{\hat{\rho}} \|F\|_{r-\rho,s}, \quad \sum_i \left\| F_{x_i} \right\|_{r-\rho,s} \le \Gamma_1 \varepsilon, \tag{16}$$

where $\hat{\rho} = \min(\rho, 1)$.

Transforming the coordinates

The preceding estimates of F, F_x hold *a fortiori* with respect to the sup-norm $|\cdot|_{r-\rho,s}$. It follows from the generalized Cauchy inequality in Appendix B that

$$\frac{1}{\rho}\left|F_y\right|_{\infty},\ \frac{2}{s^2}\left|F_x\right|_1,\ \frac{1}{s}\left|F_z\right|_2,\ \frac{1}{s}\left|F_{\bar z}\right|_2 \le 2\Gamma_1\frac{\varepsilon}{s^2}$$

uniformly on

$$\mathcal{D}_a = \mathcal{D}_{r-\rho,s/2} \subset \mathcal{D}.$$

These estimates are expressed more conveniently in terms of a *weighted phase space norm* $|W\cdot|_{\mathcal{P}}$. Namely, let

$$|(x,y,z,\bar z)|_{\mathcal{P}} = \max\bigl(|x|_{\infty},\, |y|_1,\, |z|_2,\, |\bar z|_2\bigr)$$

and

$$W = \operatorname{diag}\left(\frac{1}{\rho}I_n,\ \frac{2}{s^2}I_n,\ \frac{1}{s}I_m,\ \frac{1}{s}I_m\right),$$

where I_k denotes the k-dimensional unit matrix. Then the above estimates are equivalent to

$$|WX_F|_{\mathcal{P};\mathcal{D}_a} \le 2\Gamma E, \tag{17}$$

where

$$\Gamma = \Gamma_1, \qquad E = \frac{\varepsilon}{s^2}$$

are convenient notations for the estimates to follow.

The $|W\cdot|_{\mathcal{P}}$-distance of

$$\mathcal{D}_b = \mathcal{D}_{r-2\rho,s/4} \subset \mathcal{D}_a = \mathcal{D}_{r-\rho,s/2}$$

to the boundary of $\mathcal{D}_a$ is $1/4$. Hence, if $16\Gamma E \le 1$, then the right hand side of (17) is not bigger than $1/8$ and so

$$X_F^t\colon\quad \mathcal{D}_b \to \mathcal{D}_a, \qquad 0 \le t \le 1. \tag{18}$$

In particular, the time-1-map Φ is a symplectic map from $\mathcal{D}_b$ into $\mathcal{D}_a$ satisfying

$$|W(\Phi - id)|_{\mathcal{P};\mathcal{D}_b} \le 2\Gamma E$$

by (17) and the integral equation for Φ.

In fact, the smallness condition on E was chosen such that exactly the same statements hold for the bigger domain

$$\mathcal{D}_{r-\kappa\rho,\kappa s/4} \subset \mathcal{D}_a, \quad \kappa = 3/2$$

instead of $\mathcal{D}_b$. The $|W\cdot|_{\mathcal{P}}$-distance of its boundary to $\mathcal{D}_b$ is also $1/8$. Applying the generalized Cauchy inequality of Appendix B to the last estimate it follows that

$$\left\| W(D\Phi - I)W^{-1} \right\|_{\mathcal{P};\mathcal{D}_b} \le 16\Gamma E$$

where $\|\cdot\|_{\mathcal{P}}$ denotes the operator-norm induced by $|\cdot|_{\mathcal{P}}$.

With sharper smallness assumptions (18) is also be sharpened. This is important in order to make the truncation error (11) sufficiently small later on. Specifically, if

$$2\Gamma E \le \alpha^2 \le 1/4,$$

then

$$X_F^t: \quad \mathcal{D}_\beta = \mathcal{D}_{r-2\rho,\alpha s/2} \to \mathcal{D}_\alpha = \mathcal{D}_{r-\rho,\alpha s}, \qquad 0 \le t \le 1 \tag{19}$$

by the same arguments as before using (17). The parameter α will be chosen later.

Transforming the torus frequencies

$N_+ = N + \hat{N}$ is in normal form, if

$$\omega_+ = \omega + v(\omega)$$

are introduced as new frequencies, where v is defined by (14). We have $|v|_\infty \le E$ uniformly on $|\omega - \mathcal{O}|_\infty < h$ by (15). Refering to Lemma 12 it follows that for

$$E \le h/4 \tag{20}$$

the map $id + v$ has a real analytic inverse

$$\varphi: \quad \mathcal{W}_b = \mathcal{W}_{h/4} \to \mathcal{W}_a = \mathcal{W}_{h/2},$$

for which the estimates

$$|\varphi - id|_\infty, \quad \frac{h}{4}\left| \frac{\partial\varphi}{\partial\omega_+} - I \right|_\infty \le E$$

hold uniformly on the domain $\mathcal{W}_b$.

Estimating the new normal frequencies

The normal frequencies of N_+ are $\Omega + V$ evaluated at ω, where V is defined by (14). Expressing them in terms of ω_+,

$$\Omega_+ = (\Omega + V) \circ \varphi.$$

Again, $|V|_\infty \leq E$ uniformly on $\mathcal{W}$ by (15). Hence,

$$\begin{aligned} |\Omega_+ - \Omega|_{\mathcal{W}_b} &\leq |\Omega \circ \varphi - \Omega|_{\mathcal{W}_b} + |V \circ \varphi|_{\mathcal{W}_b} \\ &\leq \left|\frac{\partial \Omega}{\partial \omega}\right|_{\mathcal{W}_a} |\varphi - id|_{\mathcal{W}_b} + |V|_{\mathcal{W}_a} \qquad (21) \\ &\leq (M+1)E, \end{aligned}$$

where all norms are sup-norms. Similarly, introducing

$$\mu = \frac{2E}{h},$$

which is smaller than 1 by (20), we have

$$\begin{aligned} \left|\frac{\partial \Omega_+}{\partial \omega_+} - \frac{\partial \Omega}{\partial \omega} \circ \varphi\right|_{\mathcal{W}_b} &\leq \left|\frac{\partial \Omega}{\partial \omega}\right|_{\mathcal{W}_a} \left|\frac{\partial \varphi}{\partial \omega_+} - I\right|_{\mathcal{W}_b} + \left|\frac{\partial V}{\partial \omega}\right|_{\mathcal{W}_a} \left|\frac{\partial \varphi}{\partial \omega_+}\right|_{\mathcal{W}_b} \\ &\leq M \frac{\mu}{1-\mu} + \frac{2E}{h} \frac{1}{1-\mu} \\ &= (M+1) \frac{\mu}{1-\mu}. \end{aligned}$$

It follows that

$$\left|\frac{\partial \Omega_+}{\partial \omega_+}\right| \leq M + (M+1)\frac{\mu}{1-\mu} = \frac{M+\mu}{1-\mu} \qquad (22)$$

uniformly on $\mathcal{W}_b$.

By (21) the small divisor estimates (9) on the set $\mathcal{O}$ for $|k| \leq K$ deteriorate at most by the amount

$$\begin{aligned} |\langle l, \Omega_+ - \Omega \rangle| &\leq |l|_1 \, |\Omega_+ - \Omega|_\infty \\ &\leq 2(M+1)E \\ &\leq 2(M+1)\Delta(K)E \cdot \delta_k. \end{aligned}$$

By the definition of μ and assumption (12),

$$2(M+1)\Delta(K)E = \mu(M+1)\Delta(K)h \le \mu A.$$

Hence, on $\mathcal{O}$ the new frequencies satisfy

$$|\langle k, \omega_+\rangle - \langle l, \Omega_+(\omega_+)\rangle| \ge (1-\mu)A\delta_k \tag{23}$$

for $|k| \le K$ and $|l| \le 2$ not both zero.

Estimating the new error term

The new error term is

$$P_+ = \int_0^1 \{R_t, F\} \circ X_F^t \, dt \; + \; (R-P) \circ X_F^1,$$

where $R_t = (1-t)\hat{N} + tR$. Assuming that

$$2C_0 \Gamma E < \alpha^2 \le 1/4, \tag{24}$$

where C_0 is the absolute constant of Lemma 11, this lemma in combination with estimate (16) yields

$$\begin{aligned}
\|P_+\|_{\mathcal{D}_\beta} &\le \int_0^1 \left\|\{R_t, F\} \circ X_F^t\right\|_{\mathcal{D}_\beta} dt + \left\|(R-P) \circ X_F^1\right\|_{\mathcal{D}_\beta} \\
&\le \int_0^1 2\|\{R_t, F\}\|_{\mathcal{D}_\alpha} dt + 2\|R-P\|_{\mathcal{D}_\alpha}
\end{aligned}$$

with the domains as in (19). Note that (24) is stronger than the similar conditions above. Clearly,

$$\|R_t\|_{\mathcal{D}} \le \varepsilon, \qquad 0 \le t \le 1$$

by the estimates for R and $\hat{N}$ and therefore

$$\|\{R_t, F\}\|_{\mathcal{D}_\alpha} \le 16\Gamma \frac{\varepsilon^2}{s^2} = 16\Gamma E \varepsilon$$

by estimate (16) and the special case of Lemma 10 with $\sigma = s/2$. By (11),

$$\|R-P\|_{\mathcal{D}_\alpha} \le \left(e^{-K\rho} + \frac{\alpha^3}{1-\alpha^2}\right)\varepsilon,$$

since $\mathcal{D}_\alpha = \mathcal{D}_{r-\rho,\alpha s}$. Altogether it follows that

$$\|P_+\|_{\mathcal{D}_\beta} \le c_a \Gamma E \varepsilon + c_b e^{-K\rho} \varepsilon + c_c \alpha^3 \varepsilon \tag{25}$$

with absolute constants $c_a = 32$, $c_b = 2$ and $c_c = 3$.

6 Iteration

The KAM-step is repeated infinitely often to make the perturbation finally vanish. How to set things up for this purpose is not entirely obvious. Therefore we begin with some heuristic considerations.

Choosing

$$e^{-K\rho} \sim \Gamma_* E, \qquad \alpha^3 \sim \Gamma_* E$$

with $\Gamma_* \sim \Gamma$ all terms in the error estimate (25) are roughly of the same size, and

$$\varepsilon_+ \sim \Gamma_* E \varepsilon.$$

Choosing $s_+ \sim \alpha s$ this gives

$$\frac{\varepsilon_+}{s_+^2} \sim \frac{\Gamma_* E^2}{\alpha^2} \sim {\Gamma_*}^{1/3} E^{4/3}.$$

That is,

$$E_+ \sim \Gamma_*^{\kappa-1} E^\kappa, \quad \kappa = \tfrac{4}{3}.$$

This estimate is iterated with small divisor functions $\Gamma_0 \le \Gamma_1 \le \Gamma_2 \le \dots$ in place of Γ_* arising from a nonincreasing sequence $\rho_0 \ge \rho_1 \ge \rho_2 \ge \dots > 0$. After n steps,

$$E_n \le \prod_{\nu=0}^{n-1} \Gamma_\nu^{(\kappa-1)\kappa^{n-\nu-1}} E_0^{\kappa^n} = \left(\prod_{\nu=0}^{n-1} \Gamma_\nu^{\kappa_\nu} E_0 \right)^{\kappa^n},$$

where

$$\kappa_\nu = \frac{\kappa - 1}{\kappa^{\nu+1}}.$$

The infinite product of the $\Gamma_\nu^{\kappa_\nu}$ converges to a constant multiple of the function Ψ defined in Appendix A, which by hypotheses is finite for an appropriate choice of the ρ_ν. Thus, if E_0 is sufficiently small, then the E_n converge to zero exponentially fast.

The actual choice of Γ_* and thus of the Γ_n has to take into account an important constraint. Condition (12) turns out to be tantamount to

$$h \le \frac{\Gamma_* E}{\Gamma} \qquad \text{or} \qquad \frac{E}{h} \ge \frac{\Gamma}{\Gamma_*}.$$

Since E/h must converge to zero to make the iteration convergent, Γ/Γ_* must converge to zero.

The iterative lemma

In the following, a, b, c, d, e denote absolute positive integers for which a choice is given in (33).

Given $\rho > 0$ there exists a sequence $\rho_0 \ge \rho_1 \ge \ldots > 0$ such that

$$\Psi(\rho) = \prod_{\nu=0}^{\infty} \Gamma(\rho_\nu)^{\kappa_\nu}, \qquad \sum_{\nu=0}^{\infty} \rho_\nu = \rho.$$

See Appendix A for the argument. We fix such a sequence, and for $n \ge 0$ set

$$\Gamma_n = 2^{n+a}\Gamma(\rho_n), \quad \Theta_n = \prod_{\nu=0}^{n-1} \Gamma_\nu^{\kappa_\nu}, \quad E_n = (\Theta_n E_0)^{\kappa^n}$$

with $\Theta_0 = 1$. The parameters K_n, α_n are defined implicitly by the equations

$$e^{-K_n\rho_n} = 2^{-b}\Gamma_n E_n, \quad \alpha_n^3 = 2^{-3c}\Gamma_n E_n,$$

and r_n, s_n, h_n by the equations

$$r_n = r_0 - 2\sum_{\nu=0}^{n-1}\rho_\nu, \quad s_n = s_0 \prod_{\nu=0}^{n-1}\frac{\alpha_\nu}{2}, \quad h_n = 2^{n+d}E_n,$$

where the sum and product are understood to be zero and one respectively for $n = 0$. These sequences define the complex domains

$$\begin{aligned} \mathcal{D}_n &= \mathcal{D}_{r_n,s_n}, & \mathcal{W}_n &= \mathcal{W}_{h_n}, \\ \mathcal{D}_n^* &= \mathcal{D}_{r_{n+1},s_n/4}, & \mathcal{W}_n^* &= \mathcal{W}_{h_n/4}, \end{aligned}$$

where the parameter domains are complex neighbourhoods of a closed set $\mathcal{O}_n$ in $\mathbb{R}^n$

to be defined during the iteration. Finally, set

$$M_n = \frac{M_0 + \lambda_n}{1 - \lambda_n}, \quad A_n = (1 - \lambda_n) A_0,$$

where $\lambda_n = (1 - 2^{-n})/2 = 2^{-2} + \ldots + 2^{-n-1}$.

Iterative Lemma. *If*

$$E_0 \le \varepsilon_{*0} \Psi^{-1}(\rho), \qquad A_0 \ge 8(M_0 + 1), \tag{26}$$

where $\varepsilon_{*0} = 2^{-e}$*, then the following holds for all* $n \ge 0$*.*

Let N_n *be a normal form whose frequencies satisfy*

$$|\langle k, \omega \rangle - \langle l, \Omega_n(\omega) \rangle| \ge A_n \delta_k$$

on a closed subset $\mathcal{O}_n$ *of* $\mathbb{R}^n$ *for* $|k| \le K_n$ *and* $|l| \le 2$ *not both zero. Moreover, suppose that its normal frequencies* Ω_n *are real analytic and satisfy*

$$\left\| \frac{\partial \Omega_n}{\partial \omega} \right\|_\infty \le M_n$$

on the complex neighbourhood $\mathcal{W}_n$ *of* $\mathcal{O}_n$*. Finally, let* H_n *be a hamiltonian that is real analytic on* $\mathcal{D}_n \times \mathcal{W}_n$ *and satisfies*

$$\frac{1}{s_n^2} \| H_n - N_n \|_{r_n, s_n, h_n} \le E_n. \tag{27}$$

Then there exists a normal form N_{n+1} *that is real analytic on* $\mathcal{W}_n^*$*, a real analytic transformation*

$$\mathcal{F}_n\colon \quad \mathcal{D}_n^* \times \mathcal{W}_n^* \to \mathcal{D}_n \times \mathcal{W}_n$$

of the form (8), and a closed subset $\mathcal{O}_{n+1} \subset \mathcal{O}_n$ *such that for* N_{n+1} *and*

$$H_{n+1} = H_n \circ \mathcal{F}_n$$

the same set up as before holds for $n + 1$*.*

The set $\mathcal{O}_{n+1}$ *is obtained from* $\mathcal{O}_n$ *by removing the relatively open set of all points* ω *which satisfy at least* one *of the conditions*

$$|\langle k, \omega \rangle - \langle l, \Omega_{n+1}(\omega) \rangle| < A_{n+1} \delta_k$$

for $K_n < |k| \le K_{n+1}$ *and* $|l| \le 2$*.*

Auxiliary equalities and inequalities

Before giving the proof of the lemma we collect some useful facts. The κ_ν are such that

$$\sum_{\nu=0}^{\infty} \kappa_\nu = 1, \qquad \sum_{\nu=0}^{\infty} \nu \kappa_\nu = \frac{1}{\kappa - 1}.$$

By the monotinicity of the function Γ we thus have

$$\Gamma_n = \prod_{\nu=n}^{\infty} \Gamma_n^{\kappa_\nu \kappa^n} \leq \left(\prod_{\nu=n}^{\infty} \Gamma_\nu^{\kappa_\nu} \right)^{\kappa^n}.$$

Together with the definition of E_n and the smallness condition of the Iterative Lemma we obtain

$$\begin{aligned} \Gamma_n E_n \leq \left(\prod_{\nu=0}^{\infty} \Gamma_\nu^{\kappa_\nu} E_0 \right)^{\kappa^n} &= \left(2^{3+a} \Psi E_0 \right)^{\kappa^n} \\ &\leq 2^{(3+a-e)\kappa^n} \leq 2^{-n-3}, \end{aligned} \tag{28}$$

provided that $e \geq a + 6$. Moreover,

$$\Gamma_n^{\kappa-1} E_n^{\kappa} = E_{n+1} \tag{29}$$

by a straightforward calculation. Finally, if $d \geq 3$, then

$$\mu_n = \frac{2E_n}{h_n} \leq 2^{-n-2}. \tag{30}$$

This implies that

$$\frac{M_n + \mu_n}{1 - \mu_n} \leq M_{n+1} \leq 2M_0 + 1, \tag{31}$$

$$(1 - \mu_n) A_n \geq A_{n+1} \geq A_0/2, \tag{32}$$

as one easily verifies.

Proof of the lemma

The hypotheses of the lemma and of the KAM-step concerning the normal form and its perturbation are identical, if we set

$$\varepsilon_n = 2E_n s_n^2$$

and drop the index n. The construction of the KAM-step is applicable if also the conditions (12), (20) and (24) are satisfied for all $n \geq 0$.

By the above estimates for M_n and A_n and hypotheses (26),

$$\frac{A_n - 1}{2M_n + 1} \geq \frac{A_0/2 - 1}{2(2M_0 + 1) + 1} = \frac{A_0 - 2}{8M_0 + 6} \geq 1.$$

Combining this with the definitions of the function Γ and the parameters K_n and h_n we get

$$\begin{aligned} \frac{A_n - 1}{2M_n + 1} \frac{1}{K_n \Delta(K_n)} &\geq \frac{e^{-K_n \rho_n}}{K_n \Delta(K_n) e^{-K_n \rho_n}} \\ &\geq \frac{e^{-K_n \rho_n}}{\Gamma(\rho_n)} = 2^{n+a-b} E_n \geq h_n, \end{aligned}$$

if $a \geq b + d$. This is condition (12). Condition (20) holds for all $n \geq 0$ in view of (30) if $d \geq 3$. Finally, taking the third power and using the definitions of Γ_n and α_n condition (24) is equivalent to

$$C_0^3 2^{-3(n+a-1)} \Gamma_n^3 E_n^3 < 2^{-6c} \Gamma_n^2 E_n^2 \leq 2^{-6}.$$

Since $C_0 \leq 20$ and $\Gamma_n E_n \leq 1$, this holds for all $n \geq 0$ if $a \geq 2c + 6$. Thus, all the prerequisite conditions are satisfied by proper choices for a, b, c, d as in (33), and the KAM-step is applicable.

We check the claims of the lemma. Clearly, $\varphi_n \colon \mathcal{W}_n^* \to \mathcal{W}_n$, and the normal form N_{n+1} is well defined on $\mathcal{W}_n^*$. The estimate

$$|\langle k, \omega \rangle - \langle l, \Omega_{n+1}(\omega) \rangle| \geq A_{n+1} \delta_k$$

holds on $\mathcal{O}_n$ for $|k| \leq K_n$ and $|l| \leq 2$ not both zero by (23) and (32). Removing the corresponding relatively open resonance zones for $K_n < |k| \leq K_{n+1}$ a closed set $\mathcal{O}_{n+1} \subset \mathcal{O}_n$ results with the required properties. The estimate

$$\left| \frac{\partial \Omega_{n+1}}{\partial \omega} \right|_{\infty} \leq \frac{M_n + \mu_n}{1 - \mu_n} \leq M_{n+1}$$

holds on $\mathcal{W}_n^*$ by (22) and (31).

Now consider the coordinate transformation Φ_n and the new hamiltonian H_{n+1}. By (18),

$$\Phi_n\colon \quad \mathcal{D}_n^* \to \mathcal{D}_n,$$

since $r_{n+1} = r_n - 2\rho_n$. Setting

$$\mathcal{F}_n = \Phi_n \circ \varphi_n,$$

the new hamiltonian $H_{n+1} = H_n \circ \mathcal{F}_n$ and the new error term P_{n+1} are well defined on $\mathcal{D}_n^* \times \mathcal{W}_n^*$. This domain contains $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$, since

$$s_{n+1} = \alpha_n s_n/2 \le s_n/4$$

by the definition of α_n and (28), and

$$\frac{h_{n+1}}{h_n} = \frac{2E_{n+1}}{E_n} = 2\,(\Gamma_n E_n)^{\kappa-1} \le 2^{1+(3+a-e)/3} \le \frac{1}{4}$$

by (29) and (28) provided that $e \ge a + 12$.

As to the new error term, let $\|\cdot\|_{n+1}$ denote the weighted norm for this smaller domain $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$. Condition (24) is satisfied, so according to (25) we have

$$\begin{aligned} \|P_{n+1}\|_{n+1} &\le c_a \Gamma E_n \varepsilon_n + c_b e^{-K_n\rho_n} \varepsilon_n + c_c \alpha_n^3 \varepsilon_n \\ &\le \left(2^{5-a} + 2^{1-b} + 2^{2-3c}\right) \Gamma_n E_n \varepsilon_n \end{aligned}$$

by the definition of Γ_n, K_n, α_n. Dividing by $s_{n+1}^2 = \alpha_n^2 s_n^2/4$ and using again the definition of α_n and ε_n,

$$\begin{aligned} \frac{1}{s_{n+1}^2}\|P_{n+1}\|_{n+1} &\le 2^{2c+3}\left(2^{5-a} + 2^{1-b} + 2^{2-3c}\right) \Gamma_n^{\kappa-1} E_n^{\kappa} \\ &= \left(2^{8-a+2c} + 2^{4-b+2c} + 2^{5-c}\right) E_{n+1} \\ &\le E_{n+1}, \end{aligned}$$

provided that $a \ge 2c + 10$, $b \ge 2c + 6$ and $c \ge 6$. These conditions as well as the other ones are satisfied, if we choose for example

$$a = 22, \quad b = 18, \quad c = 6, \quad d = 4, \quad e = 34. \tag{33}$$

This completes the proof of the Iterative Lemma.

For the convergence proof a couple of further estimates are necessary.

Iterative Estimates. *Under the hypotheses of the Iterative Lemma,*

$$|\Omega_{n+1} - \Omega_n|_\infty \,, \quad \frac{h_n}{4}\left|\frac{\partial \Omega_{n+1}}{\partial \omega} - \frac{\partial \Omega_n}{\partial \omega} \circ \varphi_n\right|_\infty \le 2(M_0+1)E_n$$

on $\mathcal{W}_n^*$. *Moreover, on* $\mathcal{D}_n^* \times \mathcal{W}_n$ *and* $\mathcal{W}_n^*$ *respectively,*

$$|W_n\,(\Phi_n - id)|_{\mathcal{P}} \,, \quad \frac{1}{8}\left\|W_n\,(D\Phi_n - I)\,W_n^{-1}\right\|_{\mathcal{P}} \le 2^{1-a}\Gamma_n E_n$$

and

$$|\varphi_n - id|_\infty \,, \quad \frac{h_n}{4}\left|\frac{\partial \varphi_n}{\partial \omega} - I\right|_\infty \le E_n,$$

where $W_n = \operatorname{diag}\left(\rho_n^{-1} I_n, 2s_n^{-2} I_n, s_n^{-1} I_m, s_n^{-1} I_m\right)$.

These estimates are easily read off from the corresponding estimates for the KAM-step.

7 The convergence proof

Suppose the hamiltonian H and the normal form N satisfy the hypotheses of Theorem A. A preliminary transformation is applied first so that the Iterative Lemma is applicable.

Preliminary transformation

Multiply both hamiltonians by the factor

$$\lambda = \frac{8(M+1)}{\alpha}.$$

Dynamically this amounts to a stretching of the time scale by the factor λ^{-1}. Since

$$\lambda N = \langle \lambda\omega, y\rangle + \langle \lambda\Omega z, \bar{z}\rangle \,,$$

the frequencies ω, Ω and the parameter domain $\mathcal{O}$ are stretched by the same factor λ. Consequently, in the nonresonance conditions (1) the parameter α may be replaced by $\lambda\alpha = 8(M+1)$, and the radius h of analyticity in ω may be replaced by λh. On the other hand,

$$\frac{\partial(\lambda\Omega)}{\partial(\lambda\omega)} = \frac{\partial\Omega}{\partial\omega},$$

so this Jacobian is not affected.

Writing again $H, N, \ldots$ after stretching we have

$$\left|\frac{\partial\Omega}{\partial\omega}\right|_\infty \le M$$

on $\mathcal{W}_h$ and

$$\frac{1}{s^2}\|H - N\|_{r,s,h} \le E = \frac{\varepsilon_{**}}{\Psi(\rho)} \le \frac{h}{16}$$

with $\varepsilon_{**} = 8\varepsilon_*$. The Iterative Lemma applies if we choose

$$E_0 = E, \quad M_0 = M, \quad A_0 = A = 8(M+1),$$

and so on, and remove from $\mathcal{O}$ the resonance zones up to order K_0. In particular, the domain $\mathcal{D}_{r_0,s_0} \times \mathcal{W}_{h_0}$ to begin with is contained in $\mathcal{D}_{r,s} \times \mathcal{W}_h$ by letting $r_0 = r$, $s_0 = s$ and taking into account the definition of h_0 in terms of E_0.

Proof of the theorem

The decreasing sequence of closed subsets $\mathcal{O}_n$ constructed during the iteration converges to a closed set

$$\mathcal{O}_* = \bigcap_{n\ge 0} \mathcal{O}_n,$$

which we assume to be nonempty for the following discussion. On this set, the Ω_n converge uniformly to a limit Ω_* such that

$$|\langle k,\omega\rangle - \langle l, \Omega_*(\omega)\rangle| \ge A\delta_k/2, \qquad |l| \le 2, \quad |k| + |l| \ne 0$$

holds on $\mathcal{O}_*$. Hence, the normal forms N_n converge to a normal form N_* that—after undoing the stretching—satisfies the claims of the theorem.

By the Iterative Estimates the Jacobians of the Ω_n too converge uniformly on $\mathcal{O}_*$ to a limit $\partial\Omega_*$, which is the Jacobian of the map Ω_* on $\mathcal{O}_*$ in the sense of Whitney.

Now consider the transformations

$$\mathcal{F}^n = \mathcal{F}_0 \circ \ldots \circ \mathcal{F}_{n-1}\colon \quad \mathcal{D}_n \times \mathcal{W}_n \to \mathcal{D}_0 \times \mathcal{W}_0,$$

where $\mathcal{F}^0 = id$. We claim that

$$\big|\bar{W}_0(\mathcal{F}^{n+1} - \mathcal{F}^n)\big|_{\bar{\mathcal{P}}} \leq 2\max\big(2^{1-a}\Gamma_n E_n,\, E_n/h_n\big) \tag{34}$$

on $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$ for $n \geq 0$. Here, $|\cdot|_{\bar{\mathcal{P}}}$ denotes an extension of the phase space norm $|\cdot|_{\mathcal{P}}$ obtained by appending $|\omega|_\infty$, and $\bar{W}_0$ is obtained from the weight matrix W_0 by appending $h_0^{-1} I_n$.

For the proof of this estimate let $|\cdot|_n$ be short for $|\cdot|_{\bar{\mathcal{P}},\mathcal{D}_n\times\mathcal{W}_n}$. For $n \geq 0$, we have

$$\begin{aligned}
&\big|\bar{W}_0(\mathcal{F}^{n+1} - \mathcal{F}^n)\big|_{n+1}\\
&\quad= \big|\bar{W}_0(\mathcal{F}^n \circ \mathcal{F}_n - \mathcal{F}^n)\big|_{n+1}\\
&\quad\leq \big\|\bar{W}_0 \bar{D}\mathcal{F}^n \bar{W}_{n-1}^{-1}\big\|_n \big\|\bar{W}_{n-1}\bar{W}_n^{-1}\big\|_{\bar{\mathcal{P}}} \big|\bar{W}_n\,(\mathcal{F}_n - id)\big|_{n+1},
\end{aligned}$$

where the weight matrices $\bar{W}_n, \bar{W}_{n-1}$ are defined analogously to $\bar{W}_0$.

The middle factor is bounded by 1, since the weights of $\bar{W}_n^{-1}$ overpower those of $\bar{W}_{n-1}$. The last factor is bounded by the maximum of $2^{1-a}\Gamma_n E_n$ and E_n/h_n by the Iterative Estimates. For the first factor note that

$$\begin{aligned}
&\big\|\bar{W}_n \bar{D}\mathcal{F}_n \bar{W}_n^{-1}\big\|_{n+1}\\
&\quad\leq \max\big(\big\|W_n D\Phi_n W_n^{-1}\big\|_{\mathcal{P}} + h_n \|W_n \partial_\omega \Phi_n\|_{\mathcal{P},\infty},\, \|\partial_\omega \varphi_n\|_\infty\big)\\
&\quad\leq \max\big(1 + 2^{5-a}\Gamma_n E_n,\, 1 + 4E_n/h_n\big)\\
&\quad\leq 1 + 2^{2-d-n}
\end{aligned}$$

by the same estimates, the definition of h_n and (28). Hence

$$\begin{aligned}
\big\|\bar{W}_0 \bar{D}\mathcal{F}^n \bar{W}_{n-1}^{-1}\big\|_n &\leq \prod_{\nu=0}^{n-1} \big\|\bar{W}_\nu \bar{D}\mathcal{F}_\nu \bar{W}_\nu^{-1}\big\|_{\nu+1}\\
&\leq \prod_{\nu=0}^{\infty}\big(1 + 2^{2-d-\nu}\big) \leq 2
\end{aligned}$$

for $d \geq 4$. This proves (34).

Thus the $\mathcal{F}^n$ converge uniformly on

$$\bigcap_{n\geq 0} \mathcal{D}_n \times \mathcal{W}_n = \mathcal{D}_* \times \mathcal{O}_*, \qquad \mathcal{D}_* = \mathcal{D}_{r-2\rho,0}$$

to a map $\mathcal{F}_*$ that is uniformly continuous in ω and real analytic in x and satisfies the estimate

$$\left|\bar{W}_0(\mathcal{F}_* - id)\right|_{\tilde{\mathcal{P}}} \leq 1/4$$

on $\mathcal{D}_* \times \mathcal{O}_*$.

The derivatives of the $\mathcal{F}^n$ have to be handled a bit more carefully. In fact, with the given set up the partials with respect to the parameter ω can *not* be shown to converge—we can handle only the partials with respect to $y, z, \bar{z}$. For those it suffices to consider the first and second order derivatives at the origin since our mappings are of quadratic order in these variables.

Let the symbol $\tilde{\mathcal{F}}^n$ denote the $y, z, \bar{z}$-components of the map $\mathcal{F}^n$. The symbols $\tilde{\mathcal{F}}_n, \tilde{D}, \ldots$ have an analogous meaning. Since x, ω are transformed independently of $y, z, \bar{z}$, we have

$$\tilde{D}\tilde{\mathcal{F}}^{n+1} = \tilde{D}\tilde{\mathcal{F}}^n \cdot \tilde{D}\tilde{\mathcal{F}}_n.$$

The derivatives are evaluated at different points which we do not indicate for simplicity. Moreover,

$$\tilde{D}\tilde{\mathcal{F}}_n = \begin{pmatrix} * & * & * \\ 0 & * & 0 \\ 0 & 0 & * \end{pmatrix}$$

in view of the structure (8) of the map $\mathcal{F}_n$. Our weight matrices therefore *worsen* the estimate of $\tilde{D}\tilde{\mathcal{F}}_n$, that is, we have

$$|\tilde{D}\tilde{\mathcal{F}}_n - I|_{\tilde{\mathcal{P}}} \leq \left|\tilde{W}_n\,(\tilde{D}\tilde{\mathcal{F}}_n - I)\,\tilde{W}_n^{-1}\right|_{\tilde{\mathcal{P}}} \leq 2^{4-a}\varGamma_n E_n \leq 2^{1-a-n} \tag{35}$$

uniformly on $\mathcal{D}_n^* \times \mathcal{W}_n$ by (28). It follows from standard arguments that these derivatives converge uniformly on $\mathcal{D}_* \times \mathcal{O}_*$ to a limit $\tilde{D}\tilde{F}_*$ such that

$$|\tilde{D}\tilde{F}_* - I|_{\tilde{\mathcal{P}}} \leq 2^{3-a}$$

on $\mathcal{D}_* \times \mathcal{O}_*$.

The second order derivatives are handled similarly. Use the identity

$$\tilde{D}^2\tilde{\mathcal{F}}^{n+1} = \tilde{D}^2\tilde{\mathcal{F}}^n \cdot \left(\tilde{D}\tilde{\mathcal{F}}_n\right)^2 + \tilde{D}\tilde{\mathcal{F}}^n \cdot \tilde{D}^2\tilde{\mathcal{F}}_n$$

and the estimate

$$\left|\tilde{D}^2\tilde{\mathcal{F}}_n\right|_{\tilde{\mathcal{P}}} \leq 2^{7-a}\varGamma_n E_n \leq 2^{4-a-n} \tag{36}$$

on $\mathcal{D}_n^* \times \mathcal{W}_n$. The second order derivatives converge uniformly to a limit $\tilde{D}^2\tilde{\mathcal{F}}_*$

satisfying

$$\left|\tilde{D}^2 \tilde{\mathcal{F}}_*\right|_{\tilde{\mathcal{P}}} \leq 2^{7-a}$$

on $\mathcal{D}_* \times \mathcal{O}_*$.

Since our mappings are quadratic in $y, z, \bar{z}$, these estimates imply their uniform convergence on *any* domain $\mathcal{D}_{r-2\rho,\sigma} \times \mathcal{O}_*$ to a map $\mathcal{F}_*$ of the form (8) that is real analytic and symplectic for each ω. In particular,

$$\mathcal{F}_*: \quad \mathcal{D}_{r-2\rho,s/2} \times \mathcal{O}_* \to \mathcal{D}_{r,s} \times \mathcal{W}_h$$

by straighforward estimates.

It remains to prove the last statement of Theorem A. Clearly,

$$\frac{1}{s_n^2} \left\| H \circ \mathcal{F}^n - N_n \right\|_{\mathcal{D}_n \times \mathcal{W}_n} \leq E_n.$$

The $\mathcal{F}^n$ converge uniformly to $\mathcal{F}_*$ with their partial derivatives $\partial_y^l \partial_z^p \partial_{\bar{z}}^q \mathcal{F}^n$ for $2\,|l| + |p+q| \leq 2$. In the same sense the N_n converge to N_* and $H \circ \mathcal{F}^n - N_n$ converges to zero. Hence,

$$H \circ \mathcal{F}_* - N_* = R_*$$

such that

$$\partial_y^l \partial_z^p \partial_{\bar{z}}^q R_* = 0, \qquad 2\,|l| + |p+q| \leq 2.$$

This completes the proof of Theorem A.

Estimates

So far all estimates were given in terms of E instead of

$$\varepsilon = \frac{1}{s^2} \|H - N\|_{r,s,h} \leq E \tag{37}$$

to make the construction more transparent. This, however, gives only crude bounds for $\Omega_* - \Omega$ and $\mathcal{F}_* - id$. But they are easily improved by substituting

$$\hat{E}_n = \frac{\varepsilon}{E} \cdot E_n$$

for E_n everywhere in the Iterative Lemma and the Iterative Estimates. That is, all estimates are simply scaled by the factor $\varepsilon/E \leq 1$. This does not affect their validity as one easily checks. In particular, the estimate of the new error term is scaled by this factor. Hence, with this modification,

$$\begin{aligned} |\Omega_* - \Omega|_\infty &\le \sum_{n\ge 0} |\Omega_{n+1} - \Omega_n|_\infty \\ &\le \sum_{n\ge 0} 2(M_0+1)\hat{E}_n \le 4(M+1)\varepsilon \end{aligned}$$

and similarly

$$\left\| \frac{\partial \Omega_*}{\partial \omega} - \frac{\partial \Omega}{\partial \omega} \circ \varphi_* \right\|_\infty \le \sum_{n\ge 0} 8(M_0+1)\frac{\hat{E}_n}{h_n} \le (M+1)\frac{\varepsilon}{E}$$

on $\mathcal{O}_*$. The same modification applies to (34), (35) and (36) and yields

$$\left| \bar{W}_0(\mathcal{F}_* - id) \right|_{\bar{\mathcal{P}}} \le \frac{\varepsilon}{2E}$$

on $\mathcal{D}_{r-2\rho,s/2} \times \mathcal{O}_*$ by elementary estimates.

Undoing the stretching of the time scale and returning to the situation considered in Theorem A we thus obtain

Estimates to Theorem A. *For the normal form N_* and the transformation $\mathcal{F}$ the following estimates hold. If*

$$\frac{1}{s^2}\|H - N\|_{r,s,h} = \varepsilon \le E = \frac{\varepsilon_*}{M+1} \cdot \frac{\alpha}{\Psi(\rho)},$$

then

$$|\Omega_* - \Omega|_\infty , \quad 4E\left\| \frac{\partial \Omega_*}{\partial \omega} - \frac{\partial \Omega}{\partial \omega} \circ \varphi_* \right\|_\infty \le 4(M+1)\varepsilon$$

on Ω_* and

$$\left| \bar{W}_0(\mathcal{F}_* - id) \right|_{\bar{\mathcal{P}}} \le \frac{\varepsilon}{2E}$$

on $\mathcal{D}_{r-2\rho,s/2} \times \Omega_*$.

8 The measure estimate

In this section the measure of the set $\mathcal{O} - \mathcal{O}_*$ of "bad" frequencies is estimated. Recall that by construction,

$$\mathcal{O}_* = \bigcap_{n \geq 0} \mathcal{O}_n,$$

where $\mathcal{O} \supset \mathcal{O}_0 \supset \mathcal{O}_1 \supset \dots$ is a decreasing sequence of closed sets defined inductively during the iteration process by

$$\mathcal{O}_n = \mathcal{O}_{n-1} - \bigcup_{k,l} \mathcal{R}_{kl}^n, \qquad \mathcal{O}_{-1} = \mathcal{O},$$

with

$$\mathcal{R}_{kl}^n = \left\{ \omega \in \mathcal{O}_{n-1} \colon |\langle k, \omega \rangle - \langle l, \Omega_n(\omega) \rangle| < A_n \delta_k \right\},$$

$$K_{n-1} < |k| \leq K_n, \ |l| \leq 2.$$

For $n = 0$, however, the lower bound for $|k|$ is dropped, assuming $|k| + |l| \neq 0$ instead.

The first step is to estimate the Lebesgue measure of the individual resonance zones $\mathcal{R}_{kl}^n$. Let $\mathcal{K}_l^n$ be the closed convex hull of

$$\mathcal{G}_l^n = \{ \partial_\omega \langle l, \Omega_n(\omega) \rangle : \omega \in \mathcal{O}_{n-1} \}.$$

There is a fairly straightforward estimate of this measure, when k has a positive euclidean distance $\operatorname{dist}(k, \mathcal{K}_l^n)$ to this set.

Lemma 1. *If* $\operatorname{dist}(k, \mathcal{K}_l^n) = \sigma > 0$, *then*

$$m(\mathcal{R}_{kl}^n) \leq 2 A_n \delta_k \cdot D^{\bar{n}-1} / \sigma,$$

where D *denotes the exterior diameter of* $\mathcal{O}$ *with respect to the maximum norm and* $\bar{n}$ *the dimension of the ambient space.*

Proof. Let

$$\varphi(\omega) = \langle l, \Omega_n(\omega) \rangle, \qquad \Phi(\omega) = \langle k, \omega \rangle - \varphi(\omega).$$

Since k has euclidean distance $\sigma > 0$ to the closed convex set $\mathcal{K}_l^n$, there exists a unit vector v such that

$$\langle k, v\rangle - \langle \partial_\omega \varphi, v\rangle \geq \sigma$$

uniformly on $\mathcal{O}_{n-1}$. Hence,

$$\frac{d}{dt}\,\Phi(\omega + tv) = \langle k, v\rangle - \langle \partial_\omega \varphi, v\rangle \geq \sigma$$

for all t and all ω for which this function is defined. It follows that the intersection of the real line $\omega + tv$, $-\infty < t < \infty$, with the set $\mathcal{R}^n_{kl}$ is contained in an interval of length $2A_n\delta_k/\sigma$, whose position, of course, depends on ω.

On the other hand, the projection of $\mathcal{R}^n_{kl}$ along v onto any hyperplane with normal v is contained in a cube with $(n-1)$-dimensional volume $D^{\bar{n}-1}$, where D is the outer diameter of $\mathcal{O} \supset \mathcal{R}^n_{kl}$ with respect to the maximum norm. The claim thus follows from Fubini's theorem. ∎

Next we show that all k outside of $\mathcal{K}^n_l$ are in fact uniformly bounded way from this set, provided the smallness condition of Theorem B holds.

Lemma 2. *There is a positive constant s such that*

$$\mathcal{R}^n_{kl} \neq \emptyset \quad \Rightarrow \quad \operatorname{dist}\bigl(k, \mathcal{K}^n_l\bigr) \geq s$$

for all k, l, n, for which the resonance zones were defined.

Proof. First let $n = 0$. Since $\mathcal{O}_{-1} = \mathcal{O}$ is assumed to be essentially nonresonant, the set $\mathcal{R}^0_{kl}$ is empty, if $k \in \mathcal{K}_l = \mathcal{K}^0_l$ by nonresonance. Hence, $k \notin \mathcal{K}^0_l$ for a nonempty $\mathcal{R}^0_{kl}$. The sets $\mathcal{K}^0_l$ are all closed and contained in the closed ball of radius $2M$ around the origin, and for each k in that ball there are only finitely many l's for which $\mathcal{R}^0_{kl}$ is not empty by finiteness. Hence, the distance of k to $\mathcal{K}^0_l$ is uniformly bounded away from zero for all k and l, for which $\mathcal{R}^0_{kl}$ was defined.

Now let $n > 0$. Then $|k| > K_{n-1}$. We have

$$e^{-K_{n-1}\rho_{n-1}} = 2^{-b}\Gamma_{n-1}E_{n-1} \leq 2^{-b}2^{3+a-e}e^{-4M\rho}$$

for all $n \geq 1$ by the definition of K_n, estimate (28) and the smallness condition of Theorem B. Assuming for simplicity that $1 \geq \rho_0 \geq \rho_1 \geq \ldots$ (which only requires a bound on the growth of Δ for small t), it follows that

$$K_{n-1} \geq 2(2M+1) + 1 \geq 2M_n + 1$$

for $n \geq 1$. Since $\mathcal{K}_l^n$ is contained in the closed ball of radius $2M_n$ around the origin, the distance of k to $\mathcal{K}_l^n$ is always greater than one for $n > 0$. This proves the lemma. ∎

Thus,

$$m(\mathcal{R}_{kl}^n) \leq C_s A D^{\bar{n}-1} \delta_k$$

for *all* resonance zones with $C_s = 2/s$. It remains to check that for each k there are not too many nonempty zones.

Lemma 3. *For each n and k,*

$$\mathcal{R}_{kl}^n \neq \emptyset \quad \Rightarrow \quad l \in \mathcal{L}_k,$$

where $\mathcal{L}_k = \{\, l : \min_{\omega \in \mathcal{O}} |\langle k, \omega \rangle - \langle l, \Omega(\omega) \rangle| \leq 8(M+1) \,\}$.

Proof. If $\mathcal{R}_{kl}^n$ is not empty, then

$$|\langle k, \omega \rangle - \langle l, \Omega_n(\omega) \rangle| \leq A_n \delta_k \leq A_n = (1 - \lambda_n) A_0$$

for some ω in $\mathcal{O}_{n-1} \subset \mathcal{O}$. By adding up the Iterative Estimates,

$$|\Omega_n - \Omega|_\infty \leq 4(M_0 + 1) E_0 \cdot (1 - 2^{-n}) \leq 8(M_0 + 1)\lambda_n$$

on $\mathcal{O}_{n-1}$. Since $A_0 = 8(M_0 + 1)$ and $M_0 = M$ we obtain

$$|\langle k, \omega \rangle - \langle l, \Omega(\omega) \rangle| \leq 8(M+1)$$

at this point ω, whence $l \in \mathcal{L}_k$. ∎

Now the proof of Theorem B is easily completed. The measure of the set $\mathcal{O} - \mathcal{O}_*$ is bounded by

$$\sum_{n \geq 0} \sum_{k,l} m(\mathcal{R}_{kl}^n) \leq \sum_{n \geq 0} \sum_{K_{n-1} < |k| \leq K_n} C_s A D^{\bar{n}-1} \delta_k \, |\mathcal{L}_k|$$
$$\leq C_s A D^{\bar{n}-1} \sum_k \delta_k \, |\mathcal{L}_k| \,.$$

Taking into account the stretching of the frequencies by the factor λ – see section 7 – the sum in the last line is finite by the hypotheses of Theorem B. Thus,

$$m(\mathcal{O} - \mathcal{O}_*) = O(A D^{\bar{n}-1}).$$

Returning to unstretched quantities, this estimate is multiplied by $\lambda^{-\bar{n}}$. Thus,

$$m(\mathcal{O} - \mathcal{O}_\alpha) = O(\alpha d^{\bar{n}-1}),$$

since $\lambda^{-1}A = \alpha$ and $\lambda^{-1}D = d$. Theorem B is proven.

A Approximation functions

In [22] Rüssmann introduced the notion of an approximation function in order to characterize a large class of small divisors to which the KAM procedure is applicable. A similar characterization was already used by Brjuno [2] in his extension of Siegel's famous result on the linearization of complex mappings in the plane. Incidentally, those results do *not* rely on an iteration technique but on an ingenious application of the majorant method, a brief exposition of which may also be found in [18].

A nondecreasing function

$$\Delta\colon [0, \infty) \to [1, \infty)$$

is called an *approximation function,* if

$$\frac{\log \Delta(t)}{t} \searrow 0, \qquad 0 \le t \to \infty \tag{38}$$

and

$$\int^{\infty} \frac{\log \Delta(t)}{t^2}\, dt < \infty. \tag{39}$$

In addition, the normalization $\Delta(0) = 1$ is imposed for definiteness.

Suppose the small divisors are characterized in terms of an approximation function Δ as it is done in Section 2. Then their effect in a perturbation problem is described by two functions Γ and Ψ which are defined on the positive real axis in terms of Δ. For $k \ge 0$ and $\kappa > 1$,

$$\Gamma(\rho) = \sup_{t \ge 0}\, (1+t)^k \Delta(t) e^{-\rho t}$$

and

$$\Psi(\rho) = \inf \prod_{\nu=0}^{\infty} \Gamma(\rho_\nu)^{\kappa_\nu}, \qquad \kappa_\nu = \frac{\kappa - 1}{\kappa^{\nu+1}},$$

where the infimum is taken over all sequences $\rho_0 \ge \rho_1 \ge \cdots > 0$ such that $\rho_0 + \rho_1 + \cdots \le \rho$. The parameters k and κ are different for different kinds of small divisor problems. In our case,

$$k = 1, \quad \kappa = \tfrac{4}{3}.$$

For simplicity they are not made explicit in the notation.

The supremum in the definition of $\varGamma$ is attained and finite in view of condition (38). The infinite product in the definition of $\varPsi$ is lower semi-continuous when considered as a function on the set of sequences over which the infimum is taken endowed with the topology of pointwise convergence. Consequently, the infimum is also attained. For every $\rho > 0$, there exists a sequence $\rho_0^* \geq \rho_1^* \geq \cdots > 0$ whose sum is not bigger than ρ such that

$$\varPsi(\rho) = \prod_{\nu=0}^{\infty} \varGamma(\rho_\nu^*)^{\kappa_\nu}.$$

Indeed, $\rho_0^* + \rho_1^* + \cdots = \rho$, for otherwise $\varPsi$ could be further minimized.

Still, $\varPsi$ may be infinite for some $\rho > 0$. The following lemma, which is due to Rüssmann [22], rules that out.

Lemma 4. *The function $\varPsi$ is finite for all $\rho > 0$. Specifically, if*

$$\frac{1}{\log\kappa} \int_T^{\infty} \frac{\log \varDelta(t)}{t^2}\, dt \leq \rho,$$

then

$$\varPsi(\rho) \leq e^{(\kappa-1)\rho T}$$

for $k = 0$.

Note that $(1+t)^k \varDelta$ is again an approximation function, so it suffices to consider the case $k = 0$.

Proof. Let $\delta = \log \varDelta$ and

$$t_\nu = \kappa^{\nu+1} T, \qquad \rho_\nu = \delta(t_\nu)/t_\nu$$

for $\nu \geq 0$. By condition (38) and the hypotheses, $\rho_0 \geq \rho_1 \geq \cdots > 0$ and

$$\sum_{\nu=0}^{\infty} \rho_\nu \leq \int_{-1}^{\infty} \frac{\delta(t_\nu)}{t_\nu}\, d\nu \leq \frac{1}{\log\kappa} \int_T^{\infty} \frac{\delta(t)}{t^2}\, dt \leq \rho.$$

Hence, we may estimate $\Psi(\rho)$ with respect to this particular sequence. Since $\delta(t) - \rho_\nu t \le 0$ for $t \ge t_\nu$ again by condition (38), the supremum of $\delta(t) - \rho_\nu t$ is attained on the interval $[0, t_\nu]$ and thus smaller than $\delta(t_\nu)$. It follows that

$$\Gamma(\rho_\nu) = \sup_{t\ge 0} e^{\delta(t)-\rho_\nu t} \le e^{\delta(t_\nu)} = e^{\rho_\nu t_\nu}$$

by the definition of ρ_ν and hence

$$\Psi(\rho) \le \prod_{\nu=0}^{\infty} e^{\kappa_\nu \rho_\nu t_\nu} \le e^{(\kappa-1)\rho T},$$

since $\kappa_\nu t_\nu = (\kappa - 1)T$. ∎

It is convenient to impose a mild growth condition on approximation functions. We call Δ *sufficiently increasing*, if Δ is absolutely continuous with

$$\frac{d}{dt} \log \Delta(t) \ge \frac{1}{1+t}$$

for almost every $t \ge 0$. Without saying so explicitly, all our approximation functions are assumed to be sufficiently increasing.

Lemma 5. *If Δ is sufficiently increasing, then*

$$\Gamma_0(\rho) \le \rho^k \Gamma_k(\rho), \qquad k \ge 0,$$

where $\Gamma_k(\rho) = \sup_{t\ge 0} (1+t)^k \Delta(t) e^{-\rho t}$.

Proof. Suppose Δ is sufficiently increasing. If $\rho(1+t) \le 1$, then

$$\frac{d}{dt}(\log \Delta(t) - \rho t) \ge \frac{d}{dt} \log \Delta(t) - \frac{1}{1+t} \ge 0.$$

It follows that $\Delta(t)e^{-\rho t}$ attains its supremum at some point t_* where the inequality $\rho(1+t_*) \ge 1$ holds. Consequently,

$$\Gamma_0(\rho) = \Delta(t_*)e^{-\rho t_*} \le \rho^k (1+t_*)^k \Delta(t_*) e^{-\rho t_*} \le \rho^k \Gamma_k(\rho),$$

as we wanted to show. ∎

Two typical approximation functions are

$$D(t) = (1 + t/n)^n\,, \qquad E(t) = \exp\left(t^\alpha/\alpha\right),$$

for $n \geq 1$ and $0 < \alpha < 1$ respectively. They are also sufficiently increasing. The product of any two of them is also a sufficiently increasing approximation function.

Lemma 6. *For $\Delta = D$ and $k = 0$,*

$$\Gamma(\rho) \leq \max\left(\frac{1}{\rho^n}, 1\right), \quad \Psi(\rho) \leq A^n \Gamma(\rho),$$

where A only depends on κ. For $\Delta = E$ and $k = 0$,

$$\Gamma(\rho) = \exp\left(\frac{1}{\beta\rho^\beta}\right), \quad \Psi(\rho) \leq \Gamma^a(\rho)$$

where $\dfrac{1}{\alpha} - \dfrac{1}{\beta} = 1$ and a only depends on α and κ.

Proof. The estimates for the Γ-functions follow from straightforward calculations determining the maximum of $\Delta(t)e^{-\rho t}$. Choosing the sequence $\rho_\nu = \kappa_\nu \rho$ for the first example and recalling that

$$\sum_{\nu=0}^{\infty} \kappa_\nu = 1, \quad \sum_{\nu=0}^{\infty} \nu\kappa_\nu = \frac{1}{\kappa - 1},$$

we obtain, for $\rho \leq 1$,

$$\Psi(\rho) \leq \frac{1}{\rho^n} \prod_{\nu=0}^{\infty} \kappa_\nu^{-n\kappa_\nu} = \left(\frac{\kappa^{\kappa/\kappa-1}}{\kappa - 1}\right)^n \frac{1}{\rho^n} \leq \left(\frac{4}{\kappa - 1}\right)^n \frac{1}{\rho^n}.$$

The estimate for $\rho \geq 1$ follows from this. For the second example, choose $\rho_\nu = \tilde{\kappa}_\nu \rho$, where $\tilde{\kappa}_\nu$ is defined analogously to κ_ν using $\tilde{\kappa} = \kappa^{1-\alpha}$. This sequence is optimal among all geometric sequences, and gives

$$\Psi(\rho) \leq \prod_{\nu=0}^{\infty} \exp\left(\frac{\kappa_\nu}{\tilde{\kappa}_\nu^\beta \beta\rho^\beta}\right) = \exp\left(\frac{1}{\beta\rho^\beta} \sum_{\nu=0}^{\infty} \kappa_\nu \tilde{\kappa}_\nu^{-\beta}\right).$$

The infinite sum equals

$$a = \frac{\kappa - 1}{\left(\kappa^{1-\alpha} - 1\right)^{1/(1-\alpha)}},$$

as one easily checks. ∎

B The Cauchy inequality

Let A and B be two complex Banach spaces with norms $|\cdot|_A$ and $|\cdot|_B$, and let F be an analytic map from an open subset of A into B. The first derivative $d_v F$ of F at v is a linear map from A into B, whose induced operator norm is

$$\|d_v F\|_{B,A} = \max_{u \neq 0} \frac{|d_v F(u)|_B}{|u|_A}.$$

The Cauchy inequality can be stated as follows.

Lemma 7 (Generalized Cauchy inequality). *Let F be an analytic map from the open ball of radius r around v in A into B such that $|F|_B \leq M$ on this ball. Then the inequality*

$$\|d_v F\|_{B,A} \leq \frac{M}{r}$$

holds.

Proof. Let $u \neq 0$ in A. Then $f(z) = F(v + zu)$ is an analytic map from the complex disc $|z| < r/|u|_A$ in $\mathbb{C}$ into B that is uniformly bounded by M. Hence

$$|d_0 f|_B = |d_v F(u)|_B \leq \frac{M}{r} \cdot |u|_A$$

by the usual Cauchy inequality. The above statement follows, since $u \neq 0$ was arbitrary. ∎

C Weighted norms

We consider functions that are analytic on the complex domain

$$\mathcal{D}_{r,s}: \quad |\mathrm{Im}\, x|_\infty < r, \quad |y|_1 < s^2, \quad |z|_2, |\bar{z}|_2 < s$$

in $\mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^m \times \mathbb{C}^m$ and have period 2π in each of its x-variables. The dimension n must be finite, but m may be infinite. A function F of this kind has a Fourier series expansion,

$$F = \sum_{k \in \mathbb{Z}^n} F_k e^{i\langle k, \theta \rangle},$$

whose coefficients are analytic in $y, z, \bar{z}$ around the origin. Each coefficient admits a Taylor series expansion with respect to $z, \bar{z}$ at the origin,

$$F_k = \sum_{p,q \in \mathbb{N}^m} F_{k,p,q} z^p \bar{z}^q$$

in usual multi-index notation. We take the smallest majorant of F_k with respect to $z, \bar{z}$, namely

$$\mathbf{M}F_k = \sum_{p,q} \left| F_{k,p,q} \right| z^p \bar{z}^q ,$$

and define the weighted norm of F by

$$\|F\|_{r,s} = \sum_k |\mathbf{M}F_k|_s \, e^{|k|r} ,$$

where $|\cdot|_s$ denotes the sup-norm over

$$\mathcal{B}_s\colon \quad |y|_1 < s^2, \quad |z|_2 \,, |\bar{z}|_2 < s.$$

If F depends on parameters such as ω then the sup-norm of $\mathbf{M}F_k$ is extended over their range. Since this does not change any of the statements to follow, such a parameter dependence will be ignored in the following.

If $\|F\|_{r,s}$ is finite, then F is bounded on $\mathcal{D}_{r,s}$, and

$$|F|_{r,s} \le \sum_k |F_k|_s \, e^{|k|r} \le \sum_k |\mathbf{M}F_k|_s \, e^{|k|r} = \|F\|_{r,s} .$$

The converse is not true.

A basic property of the weighted norm is its multiplicity. That is, the space of analytic functions on $\mathcal{D}_{r,s}$ with finite weighted norm is a Banach algebra.

Lemma 8.

$$\|FG\|_{r,s} \le \|F\|_{r,s} \|G\|_{r,s} .$$

The proof relies on some simple facts about majorants. Let U, V be analytic on $\mathcal{B}_s$. We write

$$U \prec V,$$

if $\left|U_{pq}\right| \le V_{pq}$ holds pointwise in y for all coefficients in the Taylor series expansions of U and V with respect to $z, \bar{z}$. It is immediate that

$$\begin{aligned} \mathbf{M}(U+V) &\prec \mathbf{M}U + \mathbf{M}V, \\ \mathbf{M}(UV) &\prec \mathbf{M}U\,\mathbf{M}V. \end{aligned} \tag{40}$$

This implies

$$\begin{aligned} |\mathbf{M}(U+V)|_s &\le |\mathbf{M}U|_s + |\mathbf{M}V|_s\,, \\ |\mathbf{M}(UV)|_s &\le |\mathbf{M}U|_s\,|\mathbf{M}V|_s \end{aligned} \tag{41}$$

by the symmetry of the domain $\mathcal{B}_s$.

Proof of Lemma 8. The kth Fourier coefficient of FG is given by $(FG)_k = \sum_l F_{k-l}G_l$. Hence, by (41),

$$\begin{aligned} \|FG\|_{r,s} &= \sum_k |\mathbf{M}(FG)_k|_s\, e^{|k|r} \\ &\le \sum_{k,l} |\mathbf{M}(F_{k-l}G_l)|_s\, e^{|k|r} \\ &\le \sum_{k,l} |\mathbf{M}F_{k-l}|_s\, |\mathbf{M}G_l|_s\, e^{|k-l|r} e^{|l|r} \\ &= \|F\|_{r,s}\|G\|_{r,s}. \quad \blacksquare \end{aligned}$$

Pointwise estimates for the gradients of F are easy to obtain by the generalized Cauchy inequality, since the weighted norm majorizes the sup-norm. A somewhat stronger result holds in particular for F_x and F_y.

Lemma 9 (Cauchy inequalities).

$$\sum_{1\le i\le n} \left\|F_{x_i}\right\|_{r-\rho,s} \le \frac{1}{e\rho}\|F\|_{r,s}$$

and

$$\sup_{1\le i\le n} \left\|F_{y_i}\right\|_{r,s-\sigma} \le \frac{1}{(2s-\sigma)\sigma}\|F\|_{r,s}.$$

There is no analogous estimate for the partials with respect to z and $\bar{z}$ that is independent of the dimension m. For example, for the quadratic function $z_1^2 + \ldots + z_m^2$

one has

$$\|F\|_{r,s}^2 = s^4,$$

but

$$\sum_j \left\|F_{z_j}\right\|_{r,s}^2 = 4ms^2$$

for all $r > 0$ and $s > 0$.

Proof of Lemma 9. We have

$$\begin{aligned}
\sum_i \left\|F_{x_i}\right\|_{r-\rho,s} &\le \sum_i \sum_k |k_i|\, |\mathbf{M}F_k|_s\, e^{|k|(r-\rho)} \\
&= \sum_k |k|\, |\mathbf{M}F_k|_s\, e^{|k|(r-\rho)} \\
&\le \sup_{t\ge 0} te^{-t\rho} \|F\|_{r,s},
\end{aligned}$$

and the supremum equals $1/e\rho$. This proves the first inequality. As to the second inequality,

$$\begin{aligned}
\mathbf{M}F_{k,y_i} &= \sum_{p,q} \left|F_{kpq,y_i}\right| z^p \bar{z}^q \\
&\prec \sum_{p,q} \frac{1}{2\pi} \int_\gamma \frac{\left|F_{k,p,q}(\eta)\right|}{|\eta - y|^2}\, |d\eta|\; z^p \bar{z}^q \\
&= \frac{1}{2\pi} \int_\gamma \frac{\mathbf{M}F_k(\eta, z, \bar{z})}{|\eta - y|^2}\, |d\eta|
\end{aligned}$$

by the familiar Cauchy representation formula with a suitable path γ in the y_i-plane. It follows that

$$\left|\mathbf{M}F_{k,y_i}\right|_{s-\sigma} \le \frac{1}{s^2 - (s-\sigma)^2}\, |\mathbf{M}F_k|_s$$

and consequently

$$\begin{aligned}
\left\|F_{y_i}\right\|_{r,s-\sigma} &= \sum_k \left|\mathbf{M}F_{k,y_i}\right|_{s-\sigma} e^{|k|r} \\
&\le \frac{1}{(2s-\sigma)\sigma} \sum_k |\mathbf{M}F_k|_s\, e^{|k|r} \\
&= \frac{1}{(2s-\sigma)\sigma} \|F\|_{r,s}.
\end{aligned}$$

This holds for all $1 \leq i \leq n$, proving the second inequality. ∎

The estimate for the Poisson bracket is stated in a form that is convenient for our applications.

Lemma 10 (Poisson bracket). *Suppose that*

$$\frac{1}{\hat{\rho}_0}\|F\|_{r_0-\rho_0,s_0}, \quad \sum_i \left\|F_{x_i}\right\|_{r_0-\rho_0,s_0} \leq M,$$

where $\hat{\rho}_0 = \min(\rho_0, 1)$. *Then*

$$\|\{F, G\}\|_{r_0-\rho_0,s_0-\sigma} \leq \frac{4M}{\sigma^2}\|G\|_{r_0,s_0}$$

for $0 < \sigma < s_0$. *More generally,*

$$\|\{F, G\}\|_{r-\rho,s-\sigma} \leq \left(\frac{1}{s\sigma} + \frac{\rho_0}{s_0(s_0 - s + \sigma)\rho} + \frac{2}{(s_0 - s + \sigma)\sigma}\right) M\|G\|_{r,s}$$

for $0 < \rho < r$ *and* $0 < \sigma < s$ *with* $r - \rho \leq r_0 - \rho_0$ *and* $s - \sigma \leq s_0$.

Proof. Recall that

$$\{F, G\} = \left\langle F_x, G_y\right\rangle - \left\langle F_y, G_x\right\rangle + i\,\langle F_z, G_{\bar{z}}\rangle - i\,\langle F_{\bar{z}}, G_z\rangle\,.$$

Let $\|\cdot\|_- = \|\cdot\|_{r-\rho,s-\sigma}$. By Lemmas 8 and 9,

$$\begin{aligned}\left\|\left\langle F_x, G_y\right\rangle\right\|_- &\leq \sum_i \left\|F_{x_i}G_{y_i}\right\|_- \leq \sum_i \left\|F_{x_i}\right\|_-\left\|G_{y_i}\right\|_- \\ &\leq \sum_i \left\|F_{x_i}\right\|_- \sup_i \left\|G_{y_i}\right\|_- \leq \frac{1}{s\sigma}M\|G\|_{r,s},\end{aligned}$$

since $2s - \sigma \geq s$. Similarly,

$$\begin{aligned}\left\|\left\langle F_y, G_x\right\rangle\right\|_- &\leq \sup_i \left\|F_{y_i}\right\|_- \sum_i \left\|G_{x_i}\right\|_- \\ &\leq \frac{1}{s_0^2 - (s-\sigma)^2}\|F\|_{r_0-\rho_0,s_0}\frac{1}{e\rho}\|G\|_{r,s} \\ &\leq \frac{\rho_0}{s_0(s_0 - s + \sigma)\rho}M\|G\|_{r,s},\end{aligned}$$

since $s_0 + s - \sigma \geq s_0$.

The other two terms require a bit more work. The kth Fourier coefficient of $\langle F_z, G_{\bar z}\rangle$ is

$$H_k = \sum_l \sum_j F_{k-l,z_j} G_{l,\bar z_j}.$$

By (40),

$$\begin{aligned} \mathbf{M}H_k &\prec \sum_l \sum_j \mathbf{M}F_{k-l,z_j} \mathbf{M}G_{l,\bar z_j} \\ &= \sum_l \sum_j (\mathbf{M}F_{k-l})_{z_j} (\mathbf{M}G_l)_{\bar z_j} . \end{aligned}$$

Hence, at every point in $|z|_2, |\bar z|_2 < s - \sigma$ with *nonnegative coordinates* and everywhere in $|y|_1 < s^2$,

$$\begin{aligned} 0 \le \mathbf{M}H_k(y, z, \bar z) &\le \sum_l \sum_j \left|(\mathbf{M}F_{k-l})_{z_j}\right| \left|(\mathbf{M}G_l)_{\bar z_j}\right| \\ &\le \sum_l \left|(\mathbf{M}F_{k-l})_z\right|_2 |(\mathbf{M}G_l)_{\bar z}|_2 \\ &\le \frac{1}{(s_0 - s + \sigma)\sigma} \sum_l |\mathbf{M}F_{k-l}|_{s_0} |\mathbf{M}G_l|_s \end{aligned}$$

by the Schwarz inequality and the generalized Cauchy inequality. The supremum of $\mathbf{M}H_k$ over the "nonnegative subset" of $|z|_2, |\bar z|_2 < s - \sigma$ is the same as its supremum over the entire domain, since this function has nonnegative coefficients. Thus,

$$|\mathbf{M}H_k|_{s-\sigma} \le \frac{1}{(s_0 - s + \sigma)\sigma} \sum_l |\mathbf{M}F_{k-l}|_{s_0} |\mathbf{M}G_l|_s .$$

This yields

$$\|\langle F_z, G_{\bar z}\rangle\|_- \le \frac{1}{(s_0 - s + \sigma)\sigma} \|F\|_{r_0-\rho_0,s_0} \|G\|_{r,s} \le \frac{1}{\sigma^2} M \|G\|_{r,s}$$

by the same arguments as in the proof of Lemma 8.

The same estimate holds for $\langle F_{\bar z}, G_z\rangle$ by symmetry. Adding up the estimates for the four terms of $\{F, G\}$ we obtain the general result. The specialized result follows immediately. ∎

Unlike the familiar sup-norm, the weighted norm of a function is very sensitive to coordinate transformations. Fortunately, we only need to consider canonical transformations that are close to the identity. The following lemma is therefore stated with our specific application in mind.

Lemma 11 (Transformation). *Suppose that*

$$\frac{1}{\hat{\rho}_0}\|F\|_{r_0-\rho_0,s_0}, \quad \sum_i \|F_{x_i}\|_{r_0-\rho_0,s_0} \le M < s^2/C_0,$$

where $\hat{\rho}_0 = \min(\rho_0, 1)$ *and* $C_0 = 20$. *Then, for* $0 < \rho_0 \le \rho < r \le r_0 - \rho_0$ *and* $0 < s \le s_0/2$, *one has*

$$\|G \circ \Phi\|_{r-\rho,s/2} \le \frac{1}{1 - C_0 M/s^2}\|G\|_{r,s},$$

where Φ *denotes the time-1-map of the hamiltonian vectorfield* X_F.

The hypotheses of the lemma imply that

$$X_F^t\colon \quad \mathcal{D}_{r-\rho,s/2} \to \mathcal{D}_{r,s}, \qquad 0 \le t \le 1.$$

This fact, however, is not used explicitly in the following proof.

Proof of Lemma 11. Consider the Lie series expansion

$$G \circ \Phi = \sum_{h \ge 0} \frac{1}{h!}\mathrm{ad}_F^h G,$$

where

$$\mathrm{ad}_F^0 G = G, \qquad \mathrm{ad}_F^h G = \left\{\mathrm{ad}_F^{h-1} G, F\right\}, \quad h > 0.$$

For arbitrary ρ, σ and positive integers h with $0 < h\rho < r$ and $0 < h\sigma < s$ we have

$$\begin{aligned} \left\|\mathrm{ad}_F^h G\right\|_h &= \left\|\{\mathrm{ad}_F^{h-1} G, F\}\right\|_h \\ &\le \left(\frac{1}{s\sigma} + \frac{\rho_0}{s_0(s_0 - s)\rho} + \frac{2}{(s_0 - s)\sigma}\right) M \left\|\mathrm{ad}_F^{h-1} G\right\|_{h-1} \\ &\le \left(\frac{3}{s\sigma} + \frac{\rho_0}{s^2\rho}\right) M \left\|\mathrm{ad}_F^{h-1} G\right\|_{h-1} \end{aligned}$$

by the preceding lemma and the assumption $s_0 \ge 2s$. The notation $\|\cdot\|_h$ is short for $\|\cdot\|_{r-h\rho,s-h\sigma}$. Iterating this estimate,

$$\left\|\mathrm{ad}_F^h G\right\|_{r-h\rho,s-h\sigma} \le \left(\frac{3}{s\sigma} + \frac{\rho_0}{s^2\rho}\right)^h M^h \|G\|_{r,s}.$$

Replacing ρ, σ by $\frac{\rho}{h}, \frac{s}{2h}$ respectively and using the assumption $\rho_0 \le \rho < r$ this yields

$$\left\| \mathrm{ad}_F^h G \right\|_{r-\rho,s/2} \le \left(\frac{7Mh}{s^2} \right)^h \| G \|_{r,s} .$$

By Stirling's formula, $\frac{h^h}{h!} \le e^h$ for $h \ge 1$. So we obtain

$$\begin{aligned} \| G \circ \Phi \|_{r-\rho,s/2} &\le \sum_{h \ge 0} \frac{1}{h!} \left\| \mathrm{ad}_F^h G \right\|_{r-\rho,s/2} \\ &\le \sum_{h \ge 0} \left(\frac{C_0 M}{s^2} \right)^h \| G \|_{r,s} \\ &= \frac{1}{1 - C_0 M / s^2} \| G \|_{r,s} \end{aligned}$$

for $C_0 M < s^2$ with $C_0 = 20$, as we wanted to show. ∎

D An inverse function theorem

The following lemma describes the inverse function theorem that is applied during the KAM-step. Recall that $\mathcal{W}_h$ is an open complex neighbourhood of radius h of some subset $\mathcal{O}$ of $\mathbb{R}^n$ with respect to the sup-norm.

Lemma 12. *Suppose f is a real analytic map from $\mathcal{W}_h$ into $\mathbb{C}^n$. If*

$$|f - id|_\infty \le \delta \le h/4$$

on $\mathcal{W}_h$, then f has a real analytic inverse φ on $\mathcal{W}_{h/4}$. Moreover,

$$|\varphi - id|_\infty , \quad \tfrac{h}{4} |\partial \varphi - I|_\infty \le \delta$$

on this domain.

Proof. Let $k = h/4$. Let u, v be two points in $\mathcal{W}_{2k}$ such that $f(u) = f(v)$. Then

$$u - v = (u - f(u)) - (v - f(v)) ,$$

hence $|u - v|_\infty \le 2\delta \le 2k$. It follows that the straight line $(1-s)u + sv$, $0 \le s \le 1$, is strictly contained in $\mathcal{W}_{3k}$. Along this line,

$$\theta = \max_{\text{line}} |\partial f - I|_\infty < \delta/k \leq 1$$

by Cauchy's inequality and so

$$|u - v|_\infty \leq |\partial f - I|_\infty \, |u - v|_\infty \leq \theta \, |u - v|_\infty$$

by the mean value theorem. It follows that $u = v$. Thus, f is one-to-one on $\mathcal{W}_{2k}$.

By elementary arguments from degree theory the image of $\mathcal{W}_{2k}$ under f covers $\mathcal{W}_k$ since $|f - id|_\infty \leq \delta$. So f has a real analytic inverse φ on $\mathcal{W}_k$, which clearly satisfies $|\varphi - id|_\infty \leq \delta$. Finally,

$$\begin{aligned} |\partial\varphi - I|_{\mathcal{W}_k} &= \left|(\partial f)^{-1} \circ \varphi - I\right|_{\mathcal{W}_k} \\ &\leq \left|(\partial f)^{-1} - I\right|_{\mathcal{W}_{2k}} \\ &\leq \left(1 - |\partial f - I|_{\mathcal{W}_{2k}}\right)^{-1} - 1 \\ &\leq \frac{1}{1 - \delta/2k} - 1 \\ &\leq \frac{\delta}{k} \end{aligned}$$

by applying Cauchy to the domain $\mathcal{W}_{2k}$. ∎

References

[1] V.I. Arnold, *Proof of a theorem by A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian*, Uspehi Math. Nauk **18** (1963), 13–40; Russian Math. Surveys **18** (1963), 9–36.

[2] A.D. Brjuno, *Analytic form of differential equations*, Trans. Moscow Math. Soc. **25** (1971), 131-288.

[3] A. Celletti, L. Chierchia, *Construction of analytic KAM surfaces and effective stability bounds*, Commun. Math. Phys. **118** (1988), 119–161.

[4] M.M. Dodson, B.P. Rynne, J.A.G. Vickers, *Metric diophantine approximation and Hausdorff dimensions on manifolds*, Math. Proc. Cam. Phil. Soc. (to appear).

[5] L.H. Eliasson, *Perturbations of stable invariant tori*, Ann. Scuola Norm. Sup. Pisa (to appear).

[6] J. Fröhlich, T. Spencer, C.E. Wayne, *Localization in disordered, nonlinear dynamical systems*, Jour. Stat. Phys. **42** (1986), 247–274.

[7] S.M. Graff, *On the continuation of hyperbolic invariant tori for hamiltonian systems,* Jour. Diff. Equ. **15** (1974), 1–69.

[8] A.N. Kolmogorov, *On the conservation of conditionally periodic motions for a small change in Hamilton's function,* Dokl. Akad. Nauk SSSR **98** (1954), 525–530.

[9] S.M. Kozlov, *Reducibility of quasi-periodic differential operators and averaging,* Trans. Moscow Math. Soc. **46** (1984), 101–126.

[10] S.B. Kuksin, *Hamiltonian perturbations of infinite-dimensional linear systems with an imaginary spectrum*, Funktsional. Anal. i Prilozhen. **21** (1987), 22–37; Functional Anal. Appl. **21** (1988), 192-205.

[11] V.K. Melnikov, *On some cases of conservation of conditionally periodic motions under a small change of the Hamiltonian function*, Soviet Math. Doklady **6** (1965), 1592–1596.

[12] V.K. Melnikov, *A family of conditionally periodic solutions of a Hamiltonian system,* Soviet Math. Doklady **9** (1968), 882–886.

[13] J. Moser, *On invariant curves of area preserving mappings of an annulus,* Nachr. Akad. Wiss. Gött., Math. Phys. Kl. (1962), 1–20.

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

[15] J. Moser, *A stability theorem for minimal foliations on a torus,* Preprint, ETH Zürich, 1988.

[16] N.V. Nikolenko, *The method of Poincaré normal forms in problems of integrability of equations of evolution type,* Russ. Math. Survey **41** (1986), 63–114.

[17] J. Pöschel, *Integrability of Hamiltonian systems on Cantor sets,* Comm. Pure Appl. Math. **35** (1982), 653–695.

[18] J. Pöschel, *On invariant manifolds of complex analytic mappings near fixed points*, Expo. Math. **4** (1986), 97–109.

[19] J. Pöschel, *A general infinite dimensional KAM-theorem,* Proceedings of the IX International Congress on Mathematical Physics, Swansea, 1988.

[20] J. Pöschel, *Small divisors with spatial structure I. The theory*, Preprint, Bonn University, 1989.

[21] J. Pöschel, E. Trubowitz, *Inverse Spectral Theory*, Academic Press, Boston, 1987.

[22] H. Rüssmann, *On the one-dimensional Schrödinger equation with a quasi-periodic potential,* Annals of the New York Acad. Sci. **357** (1980), 90–107.

[23] H. Rüssmann, Talk held at the ETH Zürich, February 1986.

[24] C.L. Siegel, *Iteration of analytic functions*, Ann. Math. **43** (1942), 607–612.

[25] C.L. Siegel, *Über die Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung,* Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl. (1952), 21–30.

[26] C.L. Siegel, J. Moser, *Lectures on Celestial Mechanics*, Springer, Berlin-Heidelberg-New York, 1971.

[27] V.G. Sprindžuk, *Theory of Diophantine Approximations*, Wiley, New York, 1979.

[28] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton, 1970.

[29] D. Salamon, E. Zehnder, *KAM-theory in configuration space*, Comment. Math. Helv. (to appear).

[30] M. Vittot, *Théorie classique des perturbations et grand nombre de degres de liberté*, Thèse de doctorat de l'universitè de Provence, 1985.

[31] M. Vittot, J. Bellissard, *Invariant tori for an infinite lattice of coupled classical rotators*, Preprint, CPT-Marseille, 1985.

[32] B. Ware, *Infinite-dimensional versions of two theorems by Carl Siegel*, Bull. Am. Math. Soc. **82** (1976), 613–615.

[33] C.E. Wayne, *Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory*, Preprint No 88027, Penn. State University, 1988.

[34] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. A.M.S. **36** (1934), 63–89.

[35] E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems I*, Comm. Pure Appl. Math. **28** (1975), 91-140; *II*, Comm. Pure Appl. Math. **29** (1976), 49–111.

[36] E. Zehnder, *Siegel's linearization theorem in infinite dimensions*, Manus. math. **23** (1978), 363–371.