

# A Lecture on the Classical KAM Theorem

JÜRGEN PÖSCHEL

## 1 The Classical KAM-Theorem

**a.** The purpose of this lecture is to describe the KAM theorem in its most basic form and to give a complete and detailed proof. This proof essentially follows the traditional lines laid out by the inventors of this theory, Kolmogorov, Arnold and Moser (whence the acronym ‘KAM’), and the emphasis is more on the underlying ideas than on the sharpness of the arguments. After all, KAM theory is not only a collection of specific theorems, but rather a methodology, a collection of ideas of how to approach certain problems in perturbation theory connected with “small divisors”.

**b.** The classical KAM theorem is concerned with the stability of motions in hamiltonian systems, that are small perturbations of integrable hamiltonian systems. These integrable systems are characterized by the existence of action angle coordinates such that the hamiltonian depends on the action variable alone – see [2,14] for details. Thus we are going to consider hamiltonians of the form

$$H(p, q) = h(p) + f_\epsilon(p, q), \quad f_\epsilon(p, q) = \epsilon f_*(p, q, \epsilon)$$

for small  $\epsilon$ , where  $p = (p_1, \dots, p_n)$  are the action variables varying over some domain  $D \subset \mathbb{R}^n$ , while  $q = (q_1, \dots, q_n)$  are the conjugate angular variables, whose domain is the usual  $n$ -torus  $\mathbb{T}^n$  obtained from  $\mathbb{R}^n$  by identifying points whose components differ by integer multiples of  $2\pi$ . Thus,  $f_\epsilon$  has period  $2\pi$  in each component of  $q$ . Moreover, all our hamiltonians are assumed to be *real analytic* in all arguments.

The equations of motion are, as usual,

$$\dot{p} = -H_q(p, q), \quad \dot{q} = H_p(p, q)$$

in standard vector notation, where the dot indicates differentiation with respect to the time  $t$ , and the subscripts indicate partial derivatives. The underlying phase space is  $D \times \mathbb{T}^n \subset \mathbb{R}^n \times \mathbb{T}^n$  with the standard symplectic structure

$$\nu = \sum_j dp_j \wedge dq_j.$$

The hamiltonian vectorfield  $X_H$  associated with the equations of motions then satisfies  $\nu \lrcorner X_H = -dH$ .

We assume that the number  $n$  of degrees of freedom is at least 2, since one degree of freedom systems are always integrable.

**c.** For  $\epsilon = 0$  the system is governed by the unperturbed, integrable hamiltonian  $h$ , and the equations of motion reduce to

$$\dot{p} = 0, \quad \dot{q} = \omega$$

with

$$\omega = h_p(p).$$

They are easily integrated – hence the name *integrable system* – and their general solution is

$$p(t) = p_0, \quad q(t) = q_0 + \omega(p_0)t.$$

Hence, every solution curve is a straight line, which, due to the identification of the  $q$ -coordinates modulo  $2\pi$ , is winding around the *invariant torus*

$$\mathcal{T}_{p_0} = \{p_0\} \times \mathbb{T}^n$$

with constant frequencies, or winding numbers,  $\omega(p_0) = (\omega_1(p_0), \dots, \omega_n(p_0))$ . Such tori with linear flow are also called *Kronecker tori*.

In addition these tori are *Lagrangian*. That is, the restriction of the symplectic form  $\nu$  to their tangent space vanishes, and their dimension is maximal with respect to this property.

Thus, the whole phase space is *foliated* into an  $n$ -parameter family of invariant Lagrangian tori  $\mathcal{T}_{p_0}$ , on which the flow is linear with constant frequencies  $\omega(p_0)$ . – This is the geometric picture of an integrable hamiltonian system.

It should be kept in mind that due to the introduction of action angle coordinates these solutions are related to “real world solutions” by some coordinate transformation, which is

periodic in  $q_1, \dots, q_n$ . Expanding such a transformation into Fourier series and inserting the linear solutions obtained above, the “real world solutions” are represented by series of the form

$$\sum_{k \in \mathbb{Z}^n} a_k(p_0) e^{(k, q_0) + t \langle k, \omega(p_0) \rangle}, \quad a_k \in \mathbb{R}^{2n},$$

where  $\langle \cdot, \cdot \rangle$  denotes the usual scalar product. Thus, every solution is now *quasi-periodic* in  $t$ : its frequency spectrum in general does not consist of integer multiples of a single frequency – as is the case with periodic solutions –, but rather of integer combinations of a finite number of different frequencies. In essence, the “real world solutions” are superpositions of  $n$  oscillations, each with its own frequency. Moreover, these quasi-periodic solutions occur in families, depending on the parameter  $q_0$ , which together fill an invariant embedded  $n$ -torus.

Let us return to action angle coordinates. We observe that the topological nature of the flow on each Kronecker torus crucially depends on the arithmetical properties of its frequencies  $\omega$ . There are essentially two cases.

1. – The frequencies  $\omega$  are *nonresonant*, or *rationally independent*:

$$\langle k, \omega \rangle \neq 0 \quad \text{for all } 0 \neq k \in \mathbb{Z}^n.$$

Then, on this torus, each orbit is dense, the flow is ergodic, and the torus itself is minimal.

2. – The frequencies  $\omega$  are *resonant*, or *rationally dependent*: that is, there exist integer relations

$$\langle k, \omega \rangle = 0 \quad \text{for some } 0 \neq k \in \mathbb{Z}^n.$$

The prototype is  $\omega = (\omega_1, \dots, \omega_{n-m}, 0, \dots, 0)$ , with  $1 \leq m \leq n - 1$  trailing zeroes and nonresonant  $(\omega_1, \dots, \omega_{n-m})$ . In this case the torus decomposes into an  $m$ -parameter family of invariant  $n - m$ -tori. Each orbit is dense on such a lower dimensional torus, but not in  $\mathbb{T}^n$ .

A special case arises when there exist  $m = n - 1$  independent resonant relations. Then each frequency  $\omega_1, \dots, \omega_n$  is an integer multiple of a fixed non-zero frequency  $\omega_*$ , and the whole torus is filled by *periodic* orbits with one and the same period  $2\pi/\omega_*$ .

In an integrable system the frequencies on the tori may or may not vary with the torus, depending on the nature of the *frequency map*

$$h_p: D \rightarrow \Omega, \quad p \mapsto \omega(p) = h_p(p),$$

where  $\Omega \subset \mathbb{R}^n$ . We now make the assumption that this system is *nondegenerate* in the sense that

$$\det h_{pp} = \det \frac{\partial \omega}{\partial p} \neq 0$$

on  $D$ . Then  $h_p$  is an open map, even a local diffeomorphism between  $D$  and some open *frequency domain*  $\Omega \subset \mathbb{R}^n$ , and “the frequencies  $\omega$  effectively depend on the amplitudes  $p$ ”, as a physicist would say. It follows that nonresonant tori and resonant tori of all types each form dense subsets in phase space. Indeed, the resonant ones sit among the nonresonant ones like the rational numbers among the irrational numbers.

This “frequency-amplitude-modulation” is a genuinely nonlinear phenomenon. By contrast, in a linear system the frequencies are the same all over the phase space. As we will see, this is essential for the stability results of the KAM theory. As it is said, “the nonlinearities have a stabilizing effect”.

**d.** Now we consider the perturbed hamiltonian. The objective is to prove the persistence of invariant tori for small  $\epsilon \neq 0$ .

The first result in this direction goes back to Poincaré and is of a negative nature. He observed that the *resonant* tori are in general *destroyed* by an arbitrarily small perturbation. In particular, out of a torus with an  $n - 1$ -parameter family of periodic orbits, usually only *finitely* many periodic orbits survive a perturbation, while the others disintegrate and give way to *chaotic behavior*. – So in a nondegenerate system a *dense* set of tori is usually destroyed. This, in particular, implies that a generic hamiltonian system is *not integrable* [6,9,27].

Incidentally, it would not help to drop the nondegeneracy assumption to avoid resonant tori. If  $h$  is too degenerate, the motion may even become ergodic on each energy surface, thus destroying all tori [7].

A dense set of tori being destroyed there seems to be no hope for other tori to survive. Indeed, until the fifties it was a common belief that arbitrarily small perturbations can turn an integrable system into an ergodic one on each energy surface. In the twenties there even appeared an – erroneous – proof of this “ergodic hypothesis” by Fermi.

But in 1954 Kolmogorov observed that the converse is true – the *majority* of tori survives. He proved the persistence of those Kronecker systems, whose frequencies  $\omega$  are not only nonresonant, but are *strongly nonresonant* in the sense that there exist constants  $\alpha > 0$  and  $\tau > 0$  such that

$$|\langle k, \omega \rangle| \geq \frac{\alpha}{|k|^\tau} \quad \text{for all } 0 \neq k \in \mathbb{Z}^n, \quad (1)$$

where  $|k| = |k_1| + \dots + |k_n|$ . Such a condition is called a *diophantine* or *small divisor condition*.

The existence of such frequencies is easy to see. Let  $\Delta_\alpha^\tau$  denote the set of all  $\omega \in \mathbb{R}^n$  satisfying these infinitely many conditions with fixed  $\alpha$  and  $\tau$ . Then  $\Delta_\alpha^\tau$  is the complement of the open dense set  $R_\alpha^\tau = \bigcup_{0 \neq k \in \mathbb{Z}^n} R_{\alpha,k}^\tau$ , where

$$R_{\alpha,k}^\tau = \{ \omega \in \mathbb{R}^n : |\langle k, \omega \rangle| < \alpha / |k|^\tau \}.$$

Obviously, for any bounded domain  $\Omega \subset \mathbb{R}^n$ , we have the Lebesgue measure estimate  $m(R_{\alpha,k}^\tau \cap \Omega) = O(\alpha/|k|^{\tau+1})$ , and thus

$$m(R_\alpha^\tau \cap \Omega) \leq \sum_k m(R_{\alpha,k}^\tau \cap \Omega) = O(\alpha),$$

provided that  $\tau > n-1$ . Hence,  $R^\tau = \bigcap_{\alpha>0} R_\alpha^\tau$  is a set of measure zero, and its complement

$$\Delta^\tau = \bigcup_{\alpha>0} \Delta_\alpha^\tau$$

is a set of *full measure* in  $\mathbb{R}^n$ , for any  $\tau > n-1$ . In other words, *almost every*  $\omega$  in  $\mathbb{R}^n$  belongs to  $\Delta^\tau$ ,  $\tau > n-1$ , which is the set of all  $\omega$  in  $\mathbb{R}^n$  satisfying (1) for some  $\alpha > 0$  while  $\tau$  is fixed..

As an aside we remark that  $\Delta^\tau = \emptyset$  for  $\tau < n-1$ , because for every nonresonant  $\omega$ ,

$$\min_{0 \neq |k|_\infty \leq K} |\langle k, \omega \rangle| \leq \frac{|\omega|}{K^{n-1}} \quad (2)$$

by Dirichlet's pigeon hole argument. And for  $\tau = n-1$ , the set  $\Delta^{n-1}$  has measure zero, but Hausdorff dimension  $n$  – see [17] for references. So there are continuum many diophantine frequencies to the exponent  $n-1$ , although they form a set of measure zero. – But here we will fix  $\tau > n-1$  and drop it from the notation, letting  $\Delta_\alpha = \Delta_\alpha^\tau$ .

**e.** But although almost all frequencies are strongly nonresonant for any fixed  $\tau > n-1$ , it is *not true* that almost all tori survive a given perturbation  $f_\epsilon$ , no matter how small  $\epsilon$ . The reason is that the parameter  $\alpha$  in the nonresonance condition limits the size of the perturbation through the condition

$$\epsilon \ll \alpha^2.$$

Conversely, under a given small perturbation of size  $\epsilon$ , only those Kronecker tori with frequencies  $\omega$  in  $\Delta_\alpha$  with

$$\alpha \gg \sqrt{\epsilon},$$

do survive. Thus, we can not allow  $\alpha$  to vary, but have to *fix* it in advance.

To state the KAM theorem, we therefore single out the subsets

$$\Omega_\alpha \subset \Omega, \quad \alpha > 0,$$

whose frequencies belong to  $\Delta_\alpha$  and also have at least distance  $\alpha$  to the boundary of  $\Omega$ . These, like  $\Delta_\alpha$ , are *Cantor sets*: they are closed, perfect and *nowhere dense*, hence of first

Baire category. But they also have *large Lebesgue measure*:

$$m(\Omega - \Omega_\alpha) = O(\alpha),$$

provided the boundary of  $\Omega$  is piecewise smooth, or at least of dimension  $n - 1$  so that the measure of a boundary layer of size  $\alpha$  is  $O(\alpha)$ .

The main theorem of Kolmogorov, Arnold and Moser can now be stated as follows.

**The Classical KAM Theorem [1,8,10].** *Suppose the integrable hamiltonian  $h$  is nondegenerate, the frequency map  $h_p$  is a diffeomorphism  $D \rightarrow \Omega$ , and  $H = h + f_\epsilon$  is real analytic on  $\bar{D} \times \mathbb{T}^n$ . Then there exists a constant  $\delta > 0$  such that for*

$$|\epsilon| < \delta\alpha^2$$

*all Kronecker tori  $(\mathbb{T}^n, \omega)$  of the unperturbed system with  $\omega \in \Omega_\alpha$  persist as Lagrangian tori, being only slightly deformed. Moreover, they depend in a Lipschitz continuous way on  $\omega$  and fill the phase  $D \times \mathbb{T}^n$  up to a set of measure  $O(\alpha)$ .*

Here, “real analytic on  $\bar{D} \times \mathbb{T}^n$ ” means that the analyticity extends to a uniform neighborhood of  $D$ .

It is an immediate and important consequence of the KAM theorem that small perturbations of nondegenerate hamiltonians are *not ergodic*, as the Kronecker tori form an invariant set, which is neither of full nor of zero measure. Thus the ergodic hypothesis of the twenties was wrong.

It has to be stated again, however, that this invariant set, although of large measure, is a *Cantor set* and thus has no interior points. It is therefore impossible to tell with finite precision whether a given initial position falls onto an invariant torus or into a gap between such tori. From a physical point of view the KAM theorem rather makes a *probabilistic* statement: with probability  $1 - O(\alpha)$  a randomly chosen orbit lies on an invariant torus and is thus perpetually stable.

**f.** We conclude with some remarks about the necessity of the assumptions of the KAM theorem.

First, neither the perturbation nor the integrable hamiltonian need to be real analytic. It suffices that they are differentiable of class  $C^l$  with

$$l > 2\tau + 2 > 2n$$

to prove the persistence of individual tori [12,15,20]. For their Lipschitz dependence some more regularity is required [16].

The nondegeneracy condition may also be relaxed. It is not necessary that the frequency map is open. Roughly speaking, it suffices that the intersection of its range with any hyperplane in  $\mathbb{R}^n$  has measure zero. For example, if it happens that

$$h_p(p) = (\omega_1(p_1), \dots, \omega_n(p_1))$$

is a function of  $p_1$  only (and thus completely degenerate), it suffices to require that

$$\det \left( \frac{\partial^j \omega_i}{\partial p_1^j} \right)_{1 \leq i, j \leq n} \neq 0.$$

For a more general statement see [18], and [19,22,26] for proofs.

Finally, the hamiltonian nature of the equations is almost indispensable. Analogous result are true for reversible systems [13,16]. But in any event the system has to be *conservative*. Any kind of dissipation immediately destroys the Cantor family of tori, although isolated ones may persist as attractors.

## 2 The KAM Theorem with Parameters

**a.** Instead of proving the classical KAM theorem directly, we are going to deduce it from another KAM theorem, which is concerned with perturbations of a family of linear hamiltonians. This is accomplished by introducing the frequencies of the Kronecker tori as independent parameters. This approach was first taken in [11].

To this end we write  $p = p_0 + I$  and expand  $h$  around  $p_0$  so that

$$h(p) = h(p_0) + \langle h_p(p_0), I \rangle + \int_0^1 (1-t) \langle h_{pp}(p_t) I, I \rangle dt,$$

where  $p_t = p_0 + tI$ . By assumption, the frequency map is a *diffeomorphism*

$$h_p: D \rightarrow \Omega, \quad p_0 \mapsto \omega = h_p(p_0).$$

Hence, instead of  $p_0 \in D$  we may introduce the frequencies  $\omega \in \Omega$  as independent parameters, determining  $p_0$  uniquely. – Incidentally, the inverse map is given as

$$g_\omega: \Omega \rightarrow D, \quad \omega \mapsto p_0 = g_\omega(\omega),$$

where  $g$  is the *Legendre transform* of  $h$ , defined by  $g(\omega) = \sup_p (\langle p, \omega \rangle - h(p))$ . See [2] for more details on Legendre transforms.

Thus we write  $h(p) = e(\omega) + \langle \omega, I \rangle + P_h(I; \omega)$  and  $f_\epsilon(p, q) = P_\epsilon(I, q; \omega)$  with

$$P_h = \int_0^1 (1-t) \langle h_{pp}(p_t)I, I \rangle dt, \quad P_\epsilon = \epsilon f_*(p_0 + I, q, \epsilon).$$

Writing also  $\theta$  instead of  $q$  for the angular variables, we obtain the family of hamiltonians  $H = N + P$  with

$$N = e(\omega) + \langle \omega, I \rangle, \quad P = P_h(I; \omega) + P_\epsilon(I, \theta; \omega).$$

They are real analytic in the coordinates  $(I, \theta)$  in  $B \times \mathbb{T}^n$ ,  $B$  some sufficiently small ball around the origin in  $\mathbb{R}^n$ , as well as in the parameters  $\omega$  in some *uniform* neighborhood of  $\Omega$ .

This family is our new starting point. For  $P = 0$  it reduces to the *normal form*  $N = e(\omega) + \langle \omega, I \rangle$ . There is an invariant Kronecker torus  $\mathcal{T}_\omega = \{0\} \times \mathbb{T}^n$  with constant vectorfield

$$X_N = \sum_j \omega_j \frac{\partial}{\partial \theta_j}$$

for each  $\omega \in \Omega$ , and all these tori are given by the family

$$\Phi_0: \mathbb{T}^n \times \Omega \rightarrow B \times \mathbb{T}^n, \quad (\theta, \omega) \mapsto (0, \theta)$$

of trivial embeddings of  $\mathbb{T}^n$  over  $\Omega$  into phase space. Moreover, each such torus clearly is Lagrangian. Our aim is to prove the persistence of a ‘subfamily’ of such Lagrangian Kronecker tori under sufficiently small perturbations  $P \neq 0$  over the Cantor set  $\Omega_\alpha \subset \Omega$  of frequency parameters  $\omega$ . – Thus, instead of proving the existence of a Cantor family of invariant tori in *one* hamiltonian system, we first prove the existence of *one* invariant torus within a Cantor family of hamiltonian systems.

This change of perspective has several advantages. — The unperturbed hamiltonian is as simple as possible, namely linear. This simplifies the KAM proof. — The frequencies are separated from the actions. This makes their rôle more transparent. For example, the Lipschitz dependence of the tori on  $\omega$  is easily established. — Generalizations such as weaker nondegeneracy conditions and extension to infinite dimensional systems are easier. Also, this approach lends itself to applications in bifurcation theory, where systems naturally depend on parameters.



**b.** To state the basic result quantitatively we need to introduce a few notations. Let

$$D_{r,s} = \{I : |I| < r\} \times \{\theta : |\operatorname{Im} \theta| < s\} \subset \mathbb{C}^n \times \mathbb{C}^n$$

and

$$O_h = \{\omega : |\omega - \Omega_\alpha| < h\} \subset \mathbb{C}^n$$

denote complex neighborhoods of the torus  $\{0\} \times \mathbb{T}^n$  and  $\Omega_\alpha$ , respectively, where  $|\cdot|$  stands for the sup-norm of real vectors. The sup-norm of functions on  $D_{r,s} \times O_h$  is denoted by  $|\cdot|_{r,s,h}$ .

We will also consider the Lipschitz constants of mappings with respect to  $\omega$ . We define

$$|\varphi|_{\mathbb{L}} = \sup_{\nu \neq \omega} \frac{|\varphi(\nu) - \varphi(\omega)|}{|\nu - \omega|},$$

where the underlying domain will be clear from the context or indicated by a subscript.

**Theorem A.** *Let  $H = N + P$ . Suppose  $P$  is real analytic on  $D_{r,s} \times O_h$  with*

$$|P|_{r,s,h} \leq \gamma \alpha r s^\nu, \quad \alpha s^\nu \leq h,$$

where  $\nu = \tau + 1$  and  $\gamma$  is a small constant depending only on  $n$  and  $\tau$ . Suppose also that  $r, s, h \leq 1$ . Then there exists a Lipschitz continuous map  $\varphi: \Omega_\alpha \rightarrow \Omega$  close to the identity and a Lipschitz continuous family of real analytic torus embeddings  $\Phi: \mathbb{T}^n \times \Omega_\alpha \rightarrow B \times \mathbb{T}^n$  close to  $\Phi_0$ , such that for each  $\omega \in \Omega_\alpha$  the embedded tori are Lagrangian and

$$X_H|_{\varphi(\omega)} \circ \Phi = D\Phi \cdot X_N.$$

Moreover,  $\Phi$  is real analytic on  $T_* = \{\theta : |\operatorname{Im} \theta| < s/2\}$  for each  $\omega$ , and

$$\begin{aligned} |W(\Phi - \Phi_0)|, \alpha s^\nu |W(\Phi - \Phi_0)|_{\mathbb{L}} &\leq \frac{c}{\alpha r s^\nu} |P|_{r,s,h}, \\ |\varphi - id|, \alpha s^\nu |\varphi - id|_{\mathbb{L}} &\leq \frac{c}{r} |P|_{r,s,h}, \end{aligned}$$

uniformly on  $T_* \times \Omega_\alpha$  and  $\Omega_\alpha$ , respectively, where  $c$  is a large constant depending only on  $n$  and  $\tau$ , and  $W = \operatorname{diag}(r^{-1}Id, s^{-1}Id)$ .

By slight abuse of notation we wrote  $X_N$  also for the analogous constant vectorfield on  $\mathbb{T}^n$  alone. – The theorem states that for each  $\omega$  in  $\Omega_\alpha$  there is an embedded invariant Kronecker torus  $\mathcal{T}_\omega = \Phi(\mathbb{T}^n; \omega)$  with frequencies  $\omega$  for the hamiltonian vectorfield  $X_H$

at the slightly *shifted* parameter value  $\tilde{\omega} = \varphi(\omega)$ . Conversely, for each  $\tilde{\omega}$  in the slightly deformed Cantor set

$$\tilde{\Omega}_\alpha = \varphi(\Omega_\alpha) \subset \Omega,$$

the vectorfield  $X_H|_{\tilde{\omega}}$  admits an invariant Kronecker torus  $\mathcal{T}_\omega$  with frequencies  $\omega = \varphi^{-1}(\tilde{\omega})$ . Each such torus is Lagrangian and close to the corresponding unperturbed torus.

**c.** The Lipschitz estimates allow us to control the measure of  $\tilde{\Omega}_\alpha$  and its complement. To this end we first extend  $\varphi$  to a lipeomorphism  $\bar{\varphi}$  of  $\Omega$ .

**Proposition 1.** *There exists an extension of  $\varphi$  to a lipeomorphism  $\bar{\varphi}: \Omega \rightarrow \Omega$  with*

$$|\bar{\varphi} - id|_{L,\Omega} \leq \max(|\varphi - id|_{L,\Omega_\alpha}, \alpha^{-1} |\varphi - id|_{\Omega_\alpha}).$$

*Proof.* Let  $\psi$  be any coordinate function of  $\varphi - id$  defined on  $\Omega_\alpha$ . Define  $\psi$  on  $\Omega^c = \mathbb{R}^n - \Omega$  to be zero. Then

$$|\psi|_{L,\Omega_\alpha \cup \Omega^c} \leq \max(|\varphi - id|_{L,\Omega_\alpha}, \alpha^{-1} |\varphi - id|_{\Omega_\alpha}) < 1.$$

According to Appendix B we can extend  $\psi$  to a function  $\bar{\psi}$  on all of  $\mathbb{R}^n$  preserving its Lipschitz constant. Doing this with every coordinate of  $\varphi$  we obtain an extension  $\bar{\varphi}$  of  $\varphi$  such that  $\bar{\varphi} = id$  on  $\Omega^c$  and

$$|\bar{\varphi} - id|_{L,\mathbb{R}^n} = |\psi|_{L,\Omega_\alpha \cup \Omega^c} < 1.$$

Hence  $\bar{\varphi}$  is a lipeomorphism on  $\mathbb{R}^n$ . Since it is the identity outside of  $\Omega$ , it is also a lipeomorphism of  $\Omega$ , extending  $\varphi$ . ■

**Proposition 2.** *For  $\tilde{\Omega}_\alpha = \varphi(\Omega_\alpha)$  the estimate*

$$m(\Omega - \tilde{\Omega}_\alpha) = O(\alpha)$$

*holds, where the implicit constant depends only on  $\Omega$*

*Proof.* Let  $\bar{\varphi}$  be the extension of Proposition 1. Then

$$\begin{aligned} m(\Omega - \tilde{\Omega}_\alpha) &= m(\Omega - \varphi(\Omega_\alpha)) \\ &= m(\Omega - \bar{\varphi}(\Omega_\alpha)) \\ &= m(\bar{\varphi}(\Omega - \Omega_\alpha)) \\ &\leq |\bar{\varphi}|_{L,\Omega} m(\Omega - \Omega_\alpha) \\ &= O(\alpha). \quad \blacksquare \end{aligned}$$

The Cantor family of torus embeddings may also be extended to a family over  $\Omega$ . This extension can even be chosen so that the additional embedded tori are still Lagrangian, though of course not invariant.

**Proposition 3.** *There exists an extension of  $\Phi$  to a Lipschitz continuous family of real analytic torus embeddings*

$$\bar{\Phi}: \mathbb{T}^n \times \Omega \rightarrow B \times \mathbb{T}^n$$

such that each embedded torus is Lagrangian, and the estimates for  $\bar{\Phi}$  are the same as for  $\Phi$ , though with a different constant  $c$ .

d. We thus arrive at the following conclusion.

**Theorem B.** *Suppose the assumptions of Theorem A are satisfied. Then there exist a lipeomorphism  $\bar{\varphi}: \Omega \rightarrow \Omega$  close to the identity and a family of torus embeddings*

$$\bar{\Phi}: \mathbb{T}^n \times \Omega \rightarrow \mathbb{R}^n \times \mathbb{T}^n$$

close to  $\Phi_0$  such that for every parameter value

$$\tilde{\omega} \in \tilde{\Omega}_\alpha = \bar{\varphi}(\Omega_\alpha)$$

the hamiltonian vectorfield  $X_H|_{\tilde{\omega}}$  admits an invariant Lagrangian Kronecker torus  $\mathcal{T}_\omega = \bar{\Phi}(\mathbb{T}^n; \omega)$ , where  $\omega = \bar{\varphi}^{-1}(\tilde{\omega})$ . Moreover, the estimates for  $\bar{\varphi}$  and  $\bar{\Phi}$  are the same as for  $\varphi$  and  $\Phi$  in Theorem A, though with a different constant  $c$ , and

$$m(\Omega - \tilde{\Omega}_\alpha) = O(\alpha),$$

where the implicit constant depends only on  $\Omega$ .

We will see at the end of section 5 that the map  $\varphi$  actually can be assigned  $\omega$ -derivatives of every order on the Cantor set  $\Omega_\alpha$ . This may be formalized by introducing the intrinsically defined notion of a differentiable function on an arbitrary closed set [24,25]. The point is that – due to the Whitney extension theorem – such functions can be extended to functions on the whole space with the same differentiability properties. The upshot is that there even exists an extension of  $\varphi$  to a  $C^\infty$ -function  $\bar{\varphi}$  on  $\Omega$ . The same applies to  $\Phi$  and leads to the notion of smooth foliations of invariant tori over Cantor sets [16].

**d.** We now prove the classical KAM theorem. Introducing the frequencies as parameters we wrote the hamiltonian as  $H = N + P$ , where

$$P = P_h + P_\epsilon$$

is real analytic on  $B \times \mathbb{T}^n \times \bar{\Omega}$ ,  $B$  some small ball around the origin in  $\mathbb{R}^n$ . Thus we can fix some small  $h$  and  $s$ , with  $s^\nu \leq h$ , so that  $P$  is real analytic on the complex domain  $D_{r,s} \times O_h$  for all small  $r$ , and so that

$$|P|_{r,s,h} \leq |P_h|_{r,s,h} + |P_{f_\epsilon}|_{r,s,h} \leq Mr^2 + F\epsilon,$$

where  $M$  is a bound on the hessian of  $h$  and  $F = \sup_{p,q,\epsilon} |f_*(p, q, \epsilon)|$ .

To meet the smallness condition of Theorems A and B, we choose  $r$  by  $Mr^2 = F\epsilon$  and arrive at the condition

$$2F\epsilon \leq \gamma\alpha r s^\nu = \gamma\alpha s^\nu \sqrt{\frac{F\epsilon}{M}},$$

or

$$\epsilon \leq \frac{\gamma^2 \alpha^2 s^{2\nu}}{4FM} = \delta \alpha^2, \quad \delta = \frac{\gamma^2 s^{2\nu}}{4FM}.$$

So there is a  $\delta$  depending on  $n$ ,  $\tau$  and  $H$  such that Theorems A and B apply for  $\epsilon \leq \delta \alpha^2$ .

By construction, an orbit  $(I(t), \theta(t))$  for the hamiltonian  $H$  at the parameter value  $\tilde{\omega}$  translates into an orbit  $(p(t), q(t)) = (p_0(\tilde{\omega}) + I(t), \theta(t))$  for this hamiltonian in  $p, q$ -coordinates. It therefore follows with Theorem B that the mapping

$$\Psi: \mathbb{T}^n \times \Omega \rightarrow D \times \mathbb{T}^n,$$

which is a composition of  $\bar{\Phi}$  and  $\bar{\varphi}$  with

$$\mathbb{E}: B \times \mathbb{T}^n \times \Omega \rightarrow D \times \mathbb{T}^n, \quad (I, \theta; \omega) \mapsto (h_p^{-1}(\omega) + I, \theta),$$

is an embedding of an invariant Lagrangian Kronecker torus  $(\mathbb{T}^n, \omega)$  for every  $\omega \in \Omega_\alpha$ . Moreover,  $\Psi$  is *Lipschitz close* to the real analytic unperturbed embedding

$$\Psi_0: \mathbb{T}^n \times \Omega \rightarrow D \times \mathbb{T}^n, \quad (\theta; \omega) \mapsto (h_p^{-1}(\omega), \theta).$$

It follows that the measure of the complement of all those tori in the phase space is bounded by a constant times the measure of  $\Omega_\alpha \times \mathbb{T}^n$ , hence is  $O(\alpha)$ . This finishes the proof of the classical KAM theorem.

### 3 Outline of the Proof of Theorem A

**a.** We prove Theorem A by a rapidly converging iteration procedure that was proposed by Kolmogorov [8]. At each step of this scheme a hamiltonian

$$H_j = N_j + P_j$$

is considered, which is a small perturbation of  $N_j = e_j + \langle \omega, I \rangle$ . A coordinate and parameter transformation  $\mathcal{F}_j$  is constructed such that

$$H_j \circ \mathcal{F}_j = N_{j+1} + P_{j+1}$$

with another normal form  $N_{j+1}$  and a much smaller error term  $P_{j+1}$ . Namely,

$$|P_{j+1}| \leq C |P_j|^\kappa$$

with some exponent  $\kappa > 1$ . Repetition of this process leads to a sequence of transformations  $\mathcal{F}_0, \mathcal{F}_1, \dots$ , whose products  $\mathcal{F}^j = \mathcal{F}_0 \circ \mathcal{F}_1 \circ \dots \circ \mathcal{F}_{j-1}$  converge to an embedding of an invariant Kronecker torus.

In the meantime a number of other proofs have been given, for example by formulating some generalized implicit function theorem suited for small divisor problems [28], or by referring to an implicit function theorem in tame Frechet spaces [4]. Recently, Salamon and Zehnder [21] gave a proof that avoids coordinate transformations altogether and works in configuration space. Also, Eliasson [5] described a way of using power series expansions and majorant techniques in a very tricky way.

But here we stick to the traditional method of proof, as it probably is the most transparent way to get to know the basic techniques. They are indeed quite flexible and robust, and not at all restricted to perturbations of integrable hamiltonian systems. As we mentioned in the beginning, these techniques rather amount to a *strategy* of how to approach a large class of perturbation problems.

**b.** To describe one cycle of this iterative scheme in more detail we now drop the subscript  $j$ .

First, the perturbation  $P$  is approximated by some hamiltonian  $R$  by truncating its Taylor series in  $I$  at first order and its Fourier series in  $\theta$  at some suitable high order  $K$ . The approximation error  $P - R$  will be small, and we now consider the hamiltonian  $\bar{H} = N + R$  instead of  $H = N + P$ . The purpose of this approximation will become clear later.

The transformation  $\mathcal{F}$  consists of a parameter dependent symplectic change of coordinates  $\Phi$  and a change  $\varphi$  of the parameters alone:

$$\mathcal{F} = (\Phi, \varphi): (I, \theta; \omega) \mapsto (\Phi(I, \theta; \omega), \varphi(\omega)).$$

Moreover, this coordinate transformation is of the form

$$\Phi: (I, \theta; \omega) \rightarrow (U(I, \theta; \omega), V(\theta; \omega)),$$

where  $U$  is *affine in  $I$* , and  $V$  is independent of  $I$ . Such transformations  $\mathcal{F}$  form a group  $\mathcal{G}$  under composition.

We obtain  $\Phi$  as the time-1-map of the flow  $X_F^t$  of a hamiltonian  $F$ , which is affine in  $I$ . That is,  $\Phi$  is the time-1-shift determined by the equations of motion

$$\dot{I} = -F_\theta(I, \theta; \omega), \quad \dot{\theta} = F_I(\theta; \omega).$$

Then  $\Phi$  is symplectic for each  $\omega$ .

To describe the transformed hamiltonian  $H \circ \Phi$  we recall that for a function  $K$ ,

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

the *Poisson bracket* of  $K$  and  $F$  evaluated at  $X_F^t$ . Indeed,

$$\begin{aligned} \left. \frac{d}{dt} K \circ X_F^t \right|_{t=0} &= \sum_{1 \leq j \leq n} K_{\theta_j} \dot{\theta}_j + K_{I_j} \dot{I}_j \\ &= \sum_{1 \leq j \leq n} K_{\theta_j} F_{I_j} - K_{I_j} F_{\theta_j} = \{K, F\}, \end{aligned}$$

and the general formula follows.

So we can use Taylor's formula to expand  $\bar{H} \circ \Phi = \bar{H} \circ X_F^t|_{t=1}$  with respect to  $t$  at 0 and write

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

This is a *linear* expression in  $R$  and  $F$  – the *linearization* of  $\bar{H} \circ \Phi$  – plus a *quadratic* integral remainder. That is, if  $R$  and  $F$  are both roughly of order  $\epsilon$ , then the integral will roughly be of order  $\epsilon^2 \ll \epsilon$  and may be ascribed to the next perturbation  $P_+ = P_{j+1}$ .

The point is to find  $F$  such that

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

is again a normal form. Equivalently, we want to solve

$$\{F, N\} + \hat{N} = R, \quad \hat{N} = N_+ - N \quad (3)$$

for  $F$  and  $\hat{N}$  when  $R$  is given. Suppose for a moment that such a solution exists. Then  $(1-t)\{N, F\} + R = (1-t)\hat{N} + tR$ , and altogether we obtain

$$H \circ \Phi = \bar{H} \circ \Phi + (P - R) \circ \Phi = N_+ + P_+$$

with  $N_+ = N + \hat{N}$  and

$$P_+ = \int_0^1 \{(1-t)\hat{N} + tR, F\} \circ X_F^t dt + (P - R) \circ X_F^1 \quad (4)$$

as the new error term.

c. Let us consider equation (3) first on a formal level. Clearly,

$$\partial_\omega F \stackrel{\text{def}}{=} \{F, N\} = \sum_j F_{\theta_j} N_{I_j} = \sum_j \omega_j F_{\theta_j}$$

is a first order partial differential operator on the torus  $\mathbb{T}^n$  with constant coefficients  $\omega$ . Expanding  $F$  into a Fourier series,

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

with coefficients depending on  $I$  and  $\omega$ , we find

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

Thus,  $\partial_\omega$  admits a basis of eigenfunctions  $e^{i\langle k, \theta \rangle}$  with eigenvalues  $i \langle k, \omega \rangle$ ,  $k \in \mathbb{Z}^n$ . In other words,  $\partial_\omega$  diagonalizes with respect to this basis.

If  $\omega$  is now *nonresonant*, then these eigenvalues are all different from zero except when  $k = 0$ . We then can solve for all Fourier coefficients  $R_k$  of the given function  $R$ , except for  $R_0$ , which is given by the mean value of  $R$  over  $\mathbb{T}^n$ ,

$$R_0 = [R] = \frac{1}{\text{vol } \mathbb{T}^n} \int_{\mathbb{T}^n} R d\theta.$$

Hence, if  $R$  is given, then we can always solve the equation  $\partial_\omega F = R - [R]$  formally by setting

$$F = \sum_{0 \neq k \in \mathbb{Z}^n} \frac{R_k}{i \langle k, \omega \rangle} e^{i\langle k, \theta \rangle}. \quad (5)$$

We are still free to add a  $\theta$ -independent function to  $F$ , but we choose to normalize  $F$  so that  $[F] = 0$ .

Finally, equation (3) is completely solved by setting

$$\hat{N} = [R].$$

Of course, this choice of  $\hat{N}$  is in no way uniquely determined, but this is in some sense the simplest one.

**d.** There is a more systematic interpretation of the preceding construction. For irrational  $\omega$ , the domain of  $\partial_\omega$ , consisting of all formal Fourier series in  $\theta$  (ignoring the other coordinates here), *splits* into two *invariant subspaces*, its nullspace  $\mathcal{N}$  consisting of all constant functions, and its range  $\mathcal{R}$ , consisting of all series with vanishing constant term. Moreover,  $\partial_\omega$  is invertible on  $\mathcal{R}$ .

Decompose  $R$  into its respective components in  $\mathcal{N}$  and  $\mathcal{R}$ ,

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

The projection onto  $\mathcal{N}$  is given by taking the mean value, so

$$R_{\mathcal{N}} = [R], \quad R_{\mathcal{R}} = R - [R].$$

The equation

$$\partial_\omega F + \hat{N} = R = R_{\mathcal{R}} + R_{\mathcal{N}}$$

is then solved by “solving componentwise”,

$$\hat{N} = R_{\mathcal{N}} = [R], \quad \partial_\omega F = R_{\mathcal{R}} = R - [R],$$

where the latter can be solved uniquely for  $F$  in  $\mathcal{R}$ , since  $\partial_\omega$  is invertible on  $\mathcal{R}$ .

This general procedure – “solve for all the terms you can solve for, and keep the rest” – is at the basis of all normal form theory. It just happens to take a particularly simple form in our case.

**e.** So far our considerations were formal. But in estimating the series representation (5) of  $F$ , we are confronted with the well known and notorious problem of “small divisors”. Even if  $\omega$  is nonresonant, infinitely many of the divisors  $\langle k, \omega \rangle$  become arbitrarily small in view of (2), threatening to make the series (5) divergent.

This divergence is avoided, if  $\omega$  is required to be strongly nonresonant. To formulate this key lemma, let  $\mathcal{A}^s$  denote the space of all analytic functions  $u$  defined in the complex



strip  $\{\theta : \sup_j |\operatorname{Im} \theta_j| < s\} \subset \mathbb{C}^n$  with bounded sup-norm  $|u|_s$  over that strip. Let

$$\mathcal{A}_0^s = \{u \in \mathcal{A}^s : [u] = 0\},$$

and recall that  $\omega \in \Delta_\alpha^\tau$  satisfies  $|\langle k, \omega \rangle| \geq \alpha/|k|^\tau$  for all  $0 \neq k \in \mathbb{Z}^n$ .

**Lemma 1.** *Suppose that  $\omega \in \Delta_\alpha^\tau$ . Then the equation*

$$\partial_\omega u = v, \quad v \in \mathcal{A}_0^s,$$

has a unique solution  $u$  in  $\bigcup_{0 < \sigma < s} \mathcal{A}_0^{s-\sigma}$ , with

$$|u|_{s-\sigma} \leq \frac{c}{\alpha \sigma^\tau} |v|_s, \quad 0 < \sigma < s,$$

where the constant  $c$  depends only on  $n$  and  $\tau$ .

*Proof.* We prove the lemma with  $\sigma^{\tau+n}$  in place of  $\sigma^\tau$  to avoid lengthy technicalities. The interested reader is referred to [17] or [15] for a proof of the sharper result. – Expanding  $u$  and  $v$  into Fourier series, the unique formal solution  $u$  with  $[u] = 0$  is

$$u = \sum_{0 \neq k \in \mathbb{Z}^n} \frac{v_k}{i \langle k, \omega \rangle} e^{i \langle k, \theta \rangle}.$$

As to the estimate we recall that the Fourier coefficients of an analytic function on  $\mathbb{T}^n$  decay exponentially fast:

$$|v_k| \leq |v|_s e^{-|k|s},$$

where  $|k| = |k_1| + \dots + |k_n|$ . See Lemma A.1 for a reminder. Together with the small divisor estimate for  $\omega$  we obtain

$$|u|_{s-\sigma} \leq \sum_{k \neq 0} \frac{|v_k|}{|\langle k, \omega \rangle|} e^{|k|(s-\sigma)} \leq \frac{|v|_s}{\alpha} \sum_{k \neq 0} |k|^\tau e^{-|k|\sigma}.$$

The infinite sum is now easily estimated by constant times  $\sigma^{-\tau-n}$ . ■

We observe that  $\partial_\omega^{-1}$  is *unbounded* as an operator in  $\mathcal{A}_0^s$ . It is *bounded* only as an operator from  $\mathcal{A}_0^s$  into the larger spaces  $\mathcal{A}_0^{s-\sigma}$ , with its bound tending to infinity as  $\sigma$  tends to zero. This phenomenon is known as “loss of smoothness” affected by the solution operator  $\partial_\omega^{-1}$ , and is the main culprit why small divisor problems are technically so involved. During the iteration we have to let  $\sigma \rightarrow 0$  in order to stay in the classes  $\mathcal{A}_0^s$ . But then the operator  $\partial_\omega^{-1}$  is getting unbounded. By the rapid convergence of the Newton scheme, however, the

error term converges to zero even faster, thus allowing to overcome this effect of the small divisors.

It is absolutely essential for Lemma 1 to be true that  $\omega$  satisfies infinitely many small divisor conditions, thus restricting  $\omega$  to a *Cantor* set with no interior points. On the other hand, we will also need to transform the frequencies and thus want them to live in *open* domains. This conflict is resolved by approximating  $P$  by a trigonometric polynomial  $R$ . Then only finitely many Fourier coefficients need to be considered at each step, and only *finitely many* small divisor conditions need to be required, which are easily satisfied on some open  $\omega$ -domain. Of course, during the iteration more and more conditions have to be satisfied, and in the end these domains will shrink to some Cantor set.

**f.** We still have to finish one cycle of the iteration. Solving (3) we arrive at  $H \circ \Phi = N_+ + P_+$ , where

$$N_+ = N + [R] = e_+(\omega) + \langle \omega + v(\omega), I \rangle,$$

since  $[R]$  is affine in  $I$  and independent of  $\theta$ . To write  $N_+$  again in normal form, we have to introduce

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

as new frequencies. Since  $v$  is small, there exists an inverse map  $\varphi: \omega_+ \mapsto \omega$  by the implicit function theorem – see Appendix A. With this change of parameters,

$$N_+ = e_+(\omega_+) + \langle \omega_+, I \rangle$$

is again in normal form. This finishes one cycle of the iteration.

**g.** The next section describes the quantitative details, and the final section its iteration.

## 4 The KAM Step

**a.** To avoid a flood of constants we will write

$$u \prec v, \quad u \cdot \prec v,$$

if there exists a positive constant  $c \geq 1$ , which depends *only on  $n$  and  $\tau$*  and could be made explicit, such that  $u \leq cv$  and  $cu \leq v$ , respectively. – Now let  $P$  be a real analytic perturbation of some normal form  $N$ .

**The KAM Step.** Suppose that  $|P|_{r,s,h} \leq \epsilon$  with

$$\epsilon \ll \alpha \eta r \sigma^\nu, \quad (\text{a})$$

$$\epsilon \ll hr, \quad (\text{b})$$

$$h \leq \frac{\alpha}{2K^\nu}, \quad (\text{c})$$

for some  $0 < \eta < 1/8$ ,  $0 < \sigma < s/5$  and  $K \geq 1$ , where  $\nu = \tau + 1$ . Then there exists a real analytic transformation

$$\mathcal{F} = (\Phi, \varphi): D_{\eta r, s-5\sigma} \times O_{h/4} \rightarrow D_{r,s} \times O_h$$

in the group  $\mathcal{G}$  such that  $H \circ \mathcal{F} = N_+ + P_+$  with

$$|P_+|_{\eta r, s-5\sigma, h/4} \ll \frac{\epsilon^2}{\alpha r \sigma^\nu} + (\eta^2 + K^n e^{-K\sigma}) \epsilon.$$

Moreover,

$$\begin{aligned} |W(\Phi - id)|, |W(D\Phi - Id)W^{-1}| &\ll \frac{\epsilon}{\alpha r \sigma^\nu}, \\ |\varphi - id|, h |D\varphi - Id| &\ll \frac{\epsilon}{r}, \end{aligned}$$

uniformly on  $D_{\eta r, s-5\sigma} \times O_h$  and  $O_{h/4}$ , with the weight matrix  $W = \text{diag}(r^{-1}Id, \sigma^{-1}Id)$ .

**b.** The proof of the KAM Step follows the lines of the preceding section and consists of six small steps. Except for the last step everything is uniform in  $O_h$ , whence we write  $|\cdot|_{r,s}$  for  $|\cdot|_{r,s,h}$  throughout.

*1. Truncation.* We approximate  $P$  by a hamiltonian  $R$ , which is affine in  $I$  and a trigonometric polynomial in  $\theta$ . To this end, let  $Q$  be the linearization of  $P$  in  $I$  at  $I = 0$ . By Taylor's formula with remainder and Cauchy's estimate – see Appendix A for a reminder –, we have

$$|Q|_{r,s} \ll \epsilon, \quad |P - Q|_{2\eta r, s} \ll \eta^2 \epsilon.$$

Then we simply truncate the Fourier series of  $Q$  at order  $K$  to obtain  $R$ . By Lemma A.2,

$$|R - Q|_{r, s-\sigma} \ll K^n e^{-K\sigma} \epsilon.$$

Since the factor  $K^n e^{-K\sigma}$  will be made small later on, we also have

$$|R|_{r, s-\sigma} \ll \epsilon.$$

See Appendix A for some remarks about this truncation of Fourier series.

2. *Extending the small divisor estimate.* The nonresonance conditions (6) are assumed to hold on the real set  $\Omega_\alpha$  only. But assumption (c) implies that

$$|\langle k, \omega \rangle| \geq \frac{\alpha}{2|k|^\tau} \quad \text{for all } 0 \neq |k| \leq K \quad (7)$$

for all  $\omega$  in the neighborhood  $O_h$  of  $\Omega_\alpha$ . Indeed, for  $\omega \in O_h$  there is some  $\omega_* \in \Omega_\alpha$  with  $|\omega - \omega_*| < h$ , hence

$$|\langle k, \omega - \omega_* \rangle| \leq |k| \cdot |\omega - \omega_*| \leq Kh \leq \frac{\alpha}{2K^\tau} \leq \frac{\alpha}{2|k|^\tau}$$

for  $|k| \leq K$ . Together with the estimate (6) for  $\langle k, \omega_* \rangle$  this proves the claim.

3. *Solving the linearized equation*  $\{F, N\} + \hat{N} = R$ . We solve this equation as described in the preceding section. We have  $\hat{N} = [R]$  and thus

$$|\hat{N}|_r \leq |R|_{r, s-\sigma} < \epsilon.$$

We can solve for  $F$  uniformly for all  $\omega$  in  $O_h$  because of (7) and the fact that  $R$  contains only Fourier coefficients up to order  $K$ , by truncation. Hence the estimate of Lemma 1 applies as well, and we obtain a real analytic function  $F$  with

$$|F|_{r, s-2\sigma} < \frac{|R|_{r, s-\sigma}}{\alpha\sigma^\tau} < \frac{\epsilon}{\alpha\sigma^\tau}.$$

With Cauchy we get  $|F_\theta|_{r, s-3\sigma} < \epsilon/\alpha\sigma^{\tau+1}$  and  $|F_I|_{r/2, s-2\sigma} < \epsilon/\alpha r\sigma^\tau$ , hence

$$\frac{1}{r} |F_\theta|, \frac{1}{\sigma} |F_I| < \frac{\epsilon}{\alpha r\sigma^\nu}$$

uniformly on  $D_{r/2, s-3\sigma}$  with  $\nu = \tau + 1$ .

4. *Transforming the coordinates.* The coordinate transformation  $\Phi$  is obtained as the real analytic time-1-map of the flow  $X_F^t$  of the hamiltonian vectorfield  $X_F$ , with equations of motions

$$\dot{I} = -F_\theta, \quad \dot{\theta} = F_I.$$

With assumption (a) and the preceding estimates we can assure that we have  $|F_\theta| \leq \eta r \leq r/8$  and  $|F_I| \leq \sigma$  on  $D_{r/2, s-3\sigma}$  uniformly in  $\omega$ . Therefore, the time-1-map is well defined on  $D_{r/4, s-4\sigma}$ , with

$$\Phi = X_F^t|_{t=1} : D_{r/4, s-4\sigma} \rightarrow D_{r/2, s-3\sigma}, \quad (8)$$

and

$$|U - id| \leq |F_\theta| < \frac{\epsilon}{\alpha\sigma^\nu}, \quad |V - id| \leq |F_I| < \frac{\epsilon}{\alpha r\sigma^{\nu-1}}$$

on that domain for  $\Phi = (U, V)$ . The Jacobian of  $\Phi$  is

$$D\Phi = \begin{pmatrix} U_I & U_\theta \\ 0 & V_\theta \end{pmatrix},$$

since  $F$  is linear in  $I$ , hence  $F_I$  and  $V$  are independent of  $I$ . By the preceding estimates and Cauchy,

$$|U_I - Id| \leq \frac{\epsilon}{\alpha r \sigma^\nu}, \quad |U_\theta| \leq \frac{\epsilon}{\alpha \sigma^{\nu+1}}, \quad |V_\theta - Id| \leq \frac{\epsilon}{\alpha r \sigma^\nu}$$

on the domain  $D_{r/8, s-5\sigma} \supseteq D_{\eta r, s-5\sigma}$ . This proves all the estimates for  $\Phi$ . Finally, we observe that  $|U - id| \leq |F_\theta| \leq \eta r$  implies that we also have  $\Phi: D_{\eta r, s-5\sigma} \rightarrow D_{2\eta r, s-4\sigma}$ .

*5. New error term.* To estimate  $P_+$  as given in (4) we first consider the term  $\{R, F\}$ . Again, by Cauchy's estimate,

$$\begin{aligned} |\{R, F\}|_{r/2, s-3\sigma} &\leq |R_I| |F_\theta| + |R_\theta| |F_I| \\ &\leq \frac{\epsilon}{r} \cdot \frac{\epsilon}{\alpha \sigma^\nu} + \frac{\epsilon}{\sigma} \cdot \frac{\epsilon}{\alpha r \sigma^{\nu-1}} \leq \frac{\epsilon^2}{\alpha r \sigma^\nu}. \end{aligned}$$

The same holds for  $|\{\hat{N}, F\}|_{r/2, s-3\sigma}$ . Together with (8) and  $\eta < \frac{1}{8}$  we get

$$\begin{aligned} \left| \int_0^1 \{(1-t)\hat{N} + tR, F\} \circ X_F^t dt \right|_{\eta r, s-5\sigma} \\ \leq \left| \{(1-t)\hat{N} + tR, F\} \right|_{r/2, s-4\sigma} \leq \frac{\epsilon^2}{\alpha r \sigma^\nu}. \end{aligned}$$

The other term in (4) is bounded by

$$\begin{aligned} |(P - R) \circ \Phi|_{\eta r, s-5\sigma} &\leq |P - R|_{2\eta r, s-4\sigma} \\ &\leq |P - Q|_{2\eta r, s-4\sigma} + |Q - R|_{2\eta r, s-4\sigma} \\ &\leq (\eta^2 + K^n e^{-K\sigma}) \epsilon. \end{aligned}$$

These two estimates together give the bound for  $|P_+|$ .

*6. Transforming the frequencies.* Finally, we have to invert the map

$$\omega \mapsto \omega_+ = \omega + v(\omega), \quad v = \hat{N}_I = [R_I]$$

to put  $N + \hat{N}$  back into a normal form  $N_+$ . With assumption (b) and Cauchy's estimate we can assure that

$$|v|_{h/2} = |\hat{N}_I|_{h/2} \leq \frac{\epsilon}{r} \leq \frac{h}{4}.$$

The implicit function theorem of Appendix A applies, and there exists a real analytic inverse map  $\varphi: O_{h/4} \rightarrow O_{h/2}$ ,  $\omega_+ \mapsto \omega$ , with the estimates

$$|\varphi - id|, h |D\varphi - Id| \leq \frac{\epsilon}{r}$$

on  $O_{h/4}$ . Setting  $N_+ = (N + \hat{N}) \circ \varphi$  the proof of the KAM Step is completed.

## 5 Iteration and Proof of Theorem A

**a.** We are going to iterate the KAM Step infinitely often, choosing appropriate sequences for the parameters  $\sigma$ ,  $\eta$  and so on. To motivate our choices, let us start by fixing a geometric sequence for  $\sigma$ , say,  $\sigma_+ = \sigma/2$ , where the plus sign indicates the corresponding parameter value for the next step. Let  $r_+ = \eta r$ , and let us consider the *weighted error terms*

$$E = \frac{\epsilon}{\alpha r \sigma^\nu}, \quad E_+ = \frac{\epsilon_+}{\alpha r_+ \sigma_+^\nu}.$$

Then we have

$$E_+ \leq \frac{E^2}{\eta} + (\eta^2 + K^n e^{-K\sigma}) \frac{E}{\eta}.$$

Suppose we can choose  $\eta$  and  $K$  so that  $\eta^2 = E$  and  $K^n e^{-K\sigma} \leq E$ . Then

$$E_+ \leq \eta^{-1} E^2 = E^\kappa, \quad \kappa = \frac{3}{2}.$$

That is,  $E_+ \leq c^{\kappa-1} E^\kappa$  for some constant  $c$  determined by the KAM Step and depending only on  $n$  and  $\tau$ . Consequently,

$$c E_+ \leq c^\kappa E^\kappa,$$

and this scheme converges exponentially fast for  $E < c^{-1}$ .

It remains to discuss our assumptions

$$\eta^2 = E \tag{d}$$

$$K^n e^{-K\sigma} \leq E \tag{e}$$

as well as assumptions (a–c) of the KAM Step. There is no obstacle to take (d) as the *definition* of  $\eta$ , as this *implies*  $E \leq \eta$  and thus (a) for  $E$  sufficiently small. The other three conditions amount to

$$\frac{\epsilon}{r} \leq h \leq \frac{\alpha}{2K^\nu}, \quad K^n e^{-K\sigma} \leq E = \frac{\epsilon}{\alpha r \sigma^\nu}.$$

They are combined into the sufficient condition

$$K^\nu \sigma^\nu e^{-K\sigma} \leq E\sigma^\nu = \frac{\epsilon}{\alpha r} \cdot < \frac{h}{\alpha} \leq \frac{1}{2K^\nu}. \quad (\text{f})$$

Now it suffices to set up geometric sequences for  $K$  and  $h$  such that, say,

$$K_+ \sigma_+ = 2K\sigma, \quad K_+^\nu h_+ \leq K^\nu h.$$

Then these inequalities hold inductively, provided they hold initially, with  $K_0\sigma_0$  sufficiently large. In particular, after fixing  $\sigma_0$ ,  $K_0$  and  $E_0$ , we may set

$$h_0 = \alpha c_0 E_0 \sigma_0^\nu$$

with a suitable constant  $c_0$ .

**b.** We are now ready to set up our parameter sequences. Let

$$\sigma_{j+1} = \frac{\sigma_j}{2}, \quad K_{j+1} = 4K_j, \quad h_{j+1} = \frac{h_j}{4^\nu},$$

where  $\sigma_0 = s_0/20$ , and  $K_0$  is chosen so large that the left hand side of (f) is smaller than its right hand side, and small enough so that  $E_0$  can be fixed to meet (a) and (d). In addition,  $K_0\sigma_0$  has to be so large that the left hand side of (f) decreases at least at an exponential rate  $\kappa$ . Thus we need  $1 \cdot < K_0\sigma_0$ . Subsequently, we fix  $h_0$  as above.

Next, let

$$E_{j+1} = c^{\kappa-1} E_j^\kappa, \quad r_{j+1} = \eta_j r_j, \quad \eta_j^2 = E_j,$$

where  $r_0$  is still free and  $c$  given by the KAM Step. Finally, we define  $s_{j+1} = s_j - 5\sigma_j$  and the complex domains

$$D_j = \{|I| < r_j\} \times \{|\text{Im } \theta| < s_j\}, \quad O_j = \{|\omega - \Omega_\alpha| < h_j\}.$$

Note that  $s_j \rightarrow s/2$  and  $r_j \rightarrow 0$ . – Now let  $H = N + P_0$ .

**Iterative Lemma.** *Suppose  $P_0$  is real analytic on  $D_0 \times O_0$  with*

$$|P_0|_{r_0, s_0, h_0} \leq \epsilon_0 \stackrel{\text{def}}{=} \alpha E_0 r_0 \sigma_0^\nu,$$

*Then for each  $j \geq 0$  there exists a normal form  $N_j$  and a real analytic transformation  $\mathcal{F}^j = \mathcal{F}_0 \circ \dots \circ \mathcal{F}_{j-1}$ :  $D_j \times O_j \rightarrow D_0 \times O_0$  in the group  $\mathcal{G}$  such that  $H \circ \mathcal{F}^j = N_j + P_j$  with*

$$|P_j|_{r_j, s_j, h_j} \leq \epsilon_j \stackrel{\text{def}}{=} \alpha E_j r_j \sigma_j^\nu.$$

Moreover,

$$|\bar{W}_0(\mathcal{F}^{j+1} - \mathcal{F}^j)| \ll \frac{\epsilon_j}{r_j h_j}$$

on  $D_{j+1} \times O_{j+1}$  with the weight matrix  $\bar{W}_0 = \text{diag}(r_0^{-1}Id, \sigma_0^{-1}Id, h_0^{-1}Id)$ .

**c. Proof.** Letting  $\mathcal{F}_0 = id$ , there is nothing to do for  $j = 0$ . To proceed by induction, we have to check the assumptions of the KAM Step for each  $j \geq 0$ . But (a) is satisfied by the definition of  $\eta_j$  and the sufficiently small choice of  $E_0$ , and (b–c) hold by the definition of  $h_j$  and  $K_j$  and the choice of their initial values.

We obtain a transformation

$$\mathcal{F}_j: D_{j+1} \times O_{j+1} \rightarrow D_j \times O_j$$

taking  $H_j = N_j + P_j$  into  $H_j \circ \mathcal{F}_j = N_{j+1} + P_{j+1}$  with

$$\begin{aligned} |P_{j+1}| &\ll \epsilon_j E_j + (\eta_j^2 + K_j^n e^{-K_j \sigma_j}) \epsilon_j \\ &\ll \epsilon_j E_j = \alpha E_j^2 r_j \sigma_j^\nu \ll \alpha \eta_j^{-1} E_j^2 r_{j+1} \sigma_{j+1}^\nu. \end{aligned}$$

Since  $\eta_j^{-1} E_j^2 = E_j^\kappa = c^{1-\kappa} E_{j+1}$ , we obtain  $|P_{j+1}| \leq \epsilon_{j+1}$  by an appropriate choice of  $c$  as required. Thus, the transformation  $\mathcal{F}^{j+1} = \mathcal{F}^j \circ \mathcal{F}_j = \mathcal{F}_0 \circ \dots \circ \mathcal{F}_j$  takes  $H$  into  $N_{j+1} + P_{j+1}$  with the proper estimate for  $P_{j+1}$ .

The estimate of  $\mathcal{F}^j$  requires a bit more, though elementary work. We observe that the estimates of the KAM Step and Cauchy imply

$$|\bar{W}_j(\mathcal{F}_j - id)|, |\bar{W}_j(\bar{D}\mathcal{F}_j - Id)\bar{W}_j^{-1}| \ll \max\left(\frac{\epsilon_j}{\alpha r_j \sigma_j^\nu}, \frac{\epsilon_j}{r_j h_j}\right) \ll \frac{\epsilon_j}{r_j h_j}$$

on  $D_{j+1} \times O_{j+1}$ , where  $\bar{D}$  denotes the Jacobian with respect to  $I, \theta$  and  $\omega$ , and  $\bar{W}_j = \text{diag}(r_j^{-1}Id, \sigma_j^{-1}Id, h_j^{-1}Id)$ . We then have

$$\begin{aligned} |\bar{W}_0(\mathcal{F}^{j+1} - \mathcal{F}^j)| &= |\bar{W}_0(\mathcal{F}^j \circ \mathcal{F}_j - \mathcal{F}^j)| \\ &\leq |\bar{W}_0 \bar{D}\mathcal{F}_j \bar{W}_j^{-1}| |\bar{W}_j(\mathcal{F}_j - id)| \\ &\ll |\bar{W}_j(\mathcal{F}_j - id)| \\ &\ll \frac{\epsilon_j}{r_j h_j} \end{aligned}$$

provided we can uniformly bound the first factor in the second row on the domain  $D_j \times O_j$ .

But by induction we have  $\bar{D}\mathcal{F}^j = \bar{D}\mathcal{F}_0 \circ \dots \circ \bar{D}\mathcal{F}_{j-1}$ , with the Jacobians evaluated at different points, which we do not indicate. Since  $|\bar{W}_i \bar{W}_{i+1}^{-1}| \leq 1$  for all  $i$ , we can use a



telescoping argument and the inductive estimates for the  $\mathcal{F}_i$  to obtain

$$\begin{aligned} |\bar{W}_0 \bar{D} \mathcal{F}^j \bar{W}_j^{-1}| &\leq |\bar{W}_0 \bar{D} \mathcal{F}_0 \circ \cdots \circ \mathcal{F}_{j-1} \bar{W}_j^{-1}| \\ &\leq |\bar{W}_0 \bar{D} \mathcal{F}_0 \bar{W}_0^{-1}| |\bar{W}_0 \bar{W}_1^{-1}| \times \cdots \\ &\quad \times |\bar{W}_{j-1} \bar{D} \mathcal{F}_{j-1} \bar{W}_{j-1}^{-1}| |\bar{W}_{j-1} \bar{W}_j^{-1}| \\ &\leq \prod_j \left( 1 + \frac{c_1 \epsilon_j}{r_j h_j} \right), \end{aligned}$$

which is uniformly bounded and small, since  $\epsilon_j/r_j h_j$  converges rapidly to zero. ■

**d.** We can now prove Theorem A by applying the Iterative Lemma to  $H = N + P$ , letting  $P_0 = P$  and  $r_0 = r$ ,  $s_0 = s$ . We have  $h_0 \ll \alpha \sigma_0^\nu \leq \alpha s^\nu \leq h$  by construction and assumption, and

$$|P_0|_{r_0, s_0, h_0} \leq |P|_{r, s, h} \leq \epsilon \leq \gamma \alpha r s \leq \epsilon_0 = \alpha E_0 r_0 \sigma_0^\nu,$$

by fixing the constant  $\gamma$  in Theorem A sufficiently small.

By the estimates of the Iterative Lemma the maps  $\mathcal{F}^j$  converge uniformly on

$$\bigcap_{j \geq 0} D_j \times O_j = T_* \times \Omega_\alpha, \quad T_* = \{0\} \times \{|\operatorname{Im} \theta| < s/2\},$$

to a map  $\mathcal{F}$  consisting of a family of embeddings  $\Phi: \mathbb{T}^n \times \Omega_\alpha \rightarrow D \times \mathbb{T}^n$  and a parameter transformation  $\varphi: \Omega_\alpha \rightarrow \Omega$ , which are real analytic on  $\mathbb{T}^n$  and uniformly continuous on  $\Omega_\alpha$ . Moreover,

$$|\bar{W}_0(\mathcal{F} - id)| \ll \frac{\epsilon_0}{r_0 h_0}$$

uniformly on  $T_* \times \Omega_\alpha$  by the usual telescoping argument.

From the estimate  $|H \circ \mathcal{F}^j - N_j| \ll \epsilon_j$  on  $D_j \times O_j$  we obtain

$$|W_j(J(D\Phi^j)^t \nabla H \circ \mathcal{F}^j - J \nabla N_j)| \ll \frac{\epsilon_j}{r_j \sigma_j}$$

on  $T_* \times O_j$  with  $W_j = \operatorname{diag}(r_j^{-1} Id, \sigma_j^{-1} Id)$ . The symplectic nature of the map  $\Phi^j$  and the uniform estimate of  $\bar{W}_0 \bar{D} \mathcal{F}^j \bar{W}_j^{-1}$  above then imply

$$|X_H \circ \mathcal{F}^j - D\Phi^j \cdot X_N| \ll \frac{\epsilon_j}{r_j \sigma_j}$$

on  $T_* \times \Omega_\alpha$  for all  $j$ , where  $X_N$  is the hamiltonian vectorfield of  $N = \langle \omega, I \rangle$ . Going to the limit we obtain

$$X_H \circ \mathcal{F} = D\Phi \cdot X_N.$$

Thus,  $\Phi$  is an embedding of the Kronecker torus  $(\mathbb{T}^n, \omega)$  as an invariant torus of the hamiltonian vectorfield  $X_H$  at the parameter  $\varphi(\omega)$ . Moreover, this torus is Lagrangian, since

$$\Phi^* \nu = \lim_{j \rightarrow \infty} (\Phi^j)^* \nu|_{I=0} = \lim_{j \rightarrow \infty} \nu|_{I=0} = 0$$

by the symplecticity of the  $\Phi^j$ .

**e.** Let us now look at the  $\omega$ -derivatives of the  $\mathcal{F}^j$ . Since  $E_j$  converges to zero at an exponential rate, we have

$$\frac{\epsilon_j}{r_j h_j^l} \rightarrow 0 \quad \text{for all } l \geq 0.$$

Hence, all  $\omega$ -derivatives of the  $\mathcal{F}^j$  converge uniformly on  $T_* \times \Omega_\alpha$ , and we could assign  $\omega$ -derivatives of any order to the limit map  $\mathcal{F}$  on the Cantor set  $\Omega_\alpha$  [24]. Without making this concept precise, however, we can at least conclude that  $\mathcal{F}$  is *Lipschitz continuous* in  $\omega$ . Its Lipschitz norm is bounded by the limit of the bounds on the first  $\omega$ -derivatives of the  $\mathcal{F}^j$ . The usual Cauchy estimate yields

$$|\bar{W}_0(\mathcal{F} - id)|_L < \frac{\epsilon_0}{r_0 h_0^2}$$

on  $T_* \times \Omega_\alpha$ .

**f.** We finally look at the estimate of  $\mathcal{F}$ . So far it does not reflect the actual size  $\epsilon$  of the perturbation, since we fixed  $E_0$  and thus  $\epsilon_0$  independently of  $\epsilon$ . But we observe that everything is still all right if in all the estimates for  $P_j$ ,  $\mathcal{F}_j$  and  $\mathcal{F}^j$ , the  $\epsilon_j$  are scaled down by the linear factor

$$\frac{\epsilon}{\epsilon_0} = \alpha E_0 r_0 \sigma_0^\nu.$$

Scaling down our estimates of  $\mathcal{F}$  by this factor we can finally extract our estimates of  $\Phi$  and  $\varphi$  as stated in Theorem A, since  $\alpha \sigma_0^\nu \ll h_0$ . This finishes the proof.

## A Some Facts about Analytic Functions

**a.** First we recall a variant of the Cauchy estimate, which is used over and over. Let  $D$  be an open domain in  $\mathbb{C}^n$ , let  $D_r = \{z : |z - D| < r\}$  be the neighborhood of radius  $r$  around  $D$ , and let  $F$  be an analytic function on  $D_r$  with bounded sup-norm  $|f|_r$ . Then

$$|f_{z_j}|_{r-\rho} \leq \frac{1}{\rho} |f|_r$$

for all  $0 < \rho < r$  and  $1 \leq j \leq n$ . This follows immediately from the Cauchy estimate for one complex variable.

**b.** Next we give the estimate for the Fourier coefficients of an analytic function  $v$  on  $\mathbb{T}^n$  used in the proof of Lemma 1. Recall that  $\mathcal{A}^s$  denotes the space of all functions on  $\mathbb{T}^n$  bounded and analytic in the strip  $\{\theta : |\operatorname{Im} \theta| < s\}$ .

**Lemma A.1.** *If  $v \in \mathcal{A}^s$ , then  $v = \sum_k v_k e^{i\langle k, \theta \rangle}$  with*

$$|v_k| \leq |v|_s e^{-|k|s}$$

for all  $k \in \mathbb{Z}^n$ .

*Proof.* The Fourier coefficients  $v_k$  of  $v$  are given by

$$v_k = \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} v(\theta) e^{-i\langle k, \theta \rangle} d\theta.$$

Since the integral of an analytic function over a closed contractible loop in any of the coordinate planes is zero, and since  $v$  is  $2\pi$ -periodic in each argument also in the complex neighborhood, the path of integration may be shifted into the complex, so that

$$v_k = \frac{1}{(2\pi)^n} \int_{\mathbb{T}^n} v(\theta - i\varphi) e^{-i\langle k, \theta - i\varphi \rangle} d\theta$$

for any constant real vector  $\varphi$  with  $|\varphi| < s$ . Choosing  $\varphi = (s - \sigma)(e_1, \dots, e_n)$  with  $0 < \sigma < s$  and  $e_j = \operatorname{sgn} k_j$ ,  $1 \leq j \leq n$ , we obtain

$$|v_k| \leq |v|_s e^{-|k|(s-\sigma)}$$

for all  $\sigma > 0$ . Letting  $\sigma \rightarrow 0$  the lemma follows. ■

We can now also estimate very roughly the remainder, when we truncate the Fourier series of  $v$  at order  $K$  to obtain  $T_K v = \sum_{|k| \leq K} v_k e^{i\langle k, \theta \rangle}$ .

**Lemma A.2.** *If  $v \in \mathcal{A}^s$ , then*

$$|v - T_K v|_{s-\sigma} \leq C K^n e^{-K\sigma} |v|_s, \quad 0 \leq \sigma \leq s,$$

where the constant  $C$  only depends on  $n$ .

*Proof.* With Lemma A.1,

$$\begin{aligned} |v - T_K v|_{s-\sigma} &\leq \sum_{|k|>K} |v_k| e^{|k|(s-\sigma)} \\ &\leq |v|_s \sum_{|k|>K} e^{-|k|\sigma} \leq |v|_s \sum_{l>K} 4^n l^{n-1} e^{-l\sigma}, \end{aligned}$$

by summing first over all  $k$  with  $|k| = l$ , whose number is bounded by  $4^n l^{n-1}$ . The last sum is then easily bounded by a constant times  $K^n e^{-K\sigma}$ . ■

There are much more efficient ways to approximate a periodic function  $v$  by trigonometric polynomials. The above crude way amounts to multiplying the Fourier transform  $\hat{v}$  of  $v$  with a *discontinuous* cut off function. Instead, one should multiply  $\hat{v}$  with a *smooth* cut off function  $\psi_K$ . For instance, one could take  $\psi_K(x) = \psi(x/K)$ , where  $\psi$  is a fixed function, which is 1 on the ball  $|x| \leq \frac{1}{2}$ , 0 outside the ball  $|x| \geq 1$ , and between 0 and 1 otherwise. Transforming back,

$$(\hat{v}\psi_K)^\wedge = v * \hat{\psi}_K$$

amounts to a convolution of  $v$  with a real analytic approximation of the identity  $\hat{\psi}_K$ , as  $K \rightarrow \infty$ . Such smoothing operators have many interesting properties. For more details, see for example [28].

**c.** We finally formulate a special version of the implicit function theorem for analytic maps, which we need to invert the frequency map during the KAM Step. Recall that  $O_h$  is an open complex neighborhood of radius  $h$  of some subset  $\Omega$  of  $\mathbb{R}^n$ . In the following,  $|\cdot|$  denotes the sup-norm for vectors and maps, and the induced operator-norm for Jacobians.

**Lemma A.3.** *Suppose  $f$  is a real analytic map from  $O_h$  into  $\mathbb{C}^n$ . If  $|f - id| \leq \delta \leq h/4$  on  $O_h$ , then  $f$  has a real analytic inverse  $\varphi$  on  $O_{h/4}$ . Moreover,*

$$|\varphi - id|, \quad \frac{h}{4} |D\varphi - Id| \leq \delta$$

*on this domain.*

*Proof.* Let  $\eta = h/4$ . Let  $u, v$  be two points in  $O_{2\eta}$  with  $f(u) = f(v)$ . Then we have

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

hence  $|u - v| \leq 2\delta \leq 2\eta$ . It follows that the segment  $(1-s)u + sv$ ,  $0 \leq s \leq 1$ , is strictly contained in  $O_{3\eta}$ . Along this segment,

$$\theta = \max |Df - I| < \delta/\eta \leq 1$$

by Cauchy's inequality and so

$$|u - v| \leq |Df - I| |u - v| \leq \theta |u - v|$$

by the mean value theorem. It follows that  $u = v$ . Thus,  $f$  is one-to-one on the domain  $O_{2\eta}$ .

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

$$\begin{aligned} |D\varphi - I|_\eta &\leq |(Df)^{-1} - I|_{2\eta} \\ &\leq \frac{1}{1 - |Df - I|_{2\eta}} - 1 \leq \frac{1}{1 - \delta/2\eta} - 1 \leq \frac{\delta}{\eta} \end{aligned}$$

by applying Cauchy to the domain  $O_{2\eta}$ . ■

## B Lipschitz Functions

Let  $B \subset \mathbb{R}^n$  be a closed set. We prove the basic fact – used in section 2 – that a Lipschitz continuous function  $u: B \rightarrow \mathbb{R}$  can be extended to a Lipschitz continuous function  $U: \mathbb{R}^n \rightarrow \mathbb{R}$  without affecting its Lipschitz constant

$$|u|_{L,B} = \sup_{\substack{x,y \in B \\ x \neq y}} \frac{|u(x) - u(y)|}{|x - y|},$$

where on  $\mathbb{R}^n$  we may take any norm  $|\cdot|$ . That is, we have

$$U|_B = u, \quad |U|_{L,\mathbb{R}^n} = |u|_{L,B},$$

Incidentally, for this extension  $B$  could be *any* point set.

Indeed,  $U$  is given by

$$U(x) = \sup_{z \in B} (u(z) - \lambda |z - x|), \quad x \in \mathbb{R}^n,$$

where  $\lambda = |u|_{L,B}$ . By the triangle inequality,

$$(u(z) - \lambda |z - y|) \geq (u(z) - \lambda |z - x|) - \lambda |x - y|.$$

Taking suprema over  $z$  we obtain  $U(y) \geq U(x) - \lambda |x - y|$ , or equivalently

$$U(x) - U(y) \leq \lambda |x - y|.$$

Interchanging  $x$  and  $y$  we obtain

$$|U(x) - U(y)| \leq \lambda |x - y|,$$

whence  $|U|_{L, \mathbb{R}^n} \leq |u|_{L, B}$ . We leave it to the reader to check that  $U = u$  on  $B$ .

## References

- [1] V.I. ARNOL'D, *Proof of a theorem by A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian*, Russian Math. Surveys **18** (1963), 9–36.
- [2] V.I. ARNOL'D, *Mathematical Methods of Classical Mechanics*, Springer, 1978.
- [3] V.I. ARNOL'D (ed.), *Dynamical Systems III*, Encyclopaedia of Mathematical Sciences Volume 3, Springer, 1988.
- [4] J.-B. BOST, *Tore invariants des systèmes dynamiques Hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...)*, Séminaire Bourbaki, Astérisque **133–134** (1986), 113–157.
- [5] L.H. ELIASSON, *Absolutely convergent series expansions for quasi periodic motions*, Math. Phys. Electron. J. **2** (1996), Paper 4, 33 p..
- [6] C. GENECAND, *Transversal homoclinic points near elliptic fixed points of area-preserving diffeomorphisms of the plane*, Dynamics Reported, New Series, Volume 2, Springer-Verlag, 1993, pp. 1–30.
- [7] A.B. KATOK, *Ergodic properties of degenerate integrable Hamiltonian systems*, Math. USSR Izv. **7** (1973), 185–214.
- [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), 527–530, (Russian). English translation in: Lectures Notes in Physics 93, Springer, 1979.
- [9] L. MARKUS AND K. MEYER, *Generic Hamiltonian systems are neither integrable nor ergodic*, Mem. Am. Math. soc. **144** (1974).
- [10] J. MOSER, *On invariant curves of area preserving mappings of an annulus*, Nachr. Akad. Wiss. Gött., Math. Phys. Kl. (1962), 1–20.
- [11] J. MOSER, *Convergent series expansions for quasi-periodic motions*, Math. Ann. **169** (1967), 136–176.
- [12] J. MOSER, *On the continuation of almost periodic solutions for ordinary differential equations*, Proc. Int. Conf. Func. Anal. and Rel. Topics, Tokyo, 1969, pp. 60–67.
- [13] J. MOSER, *Stable and Random Motions in Dynamical Systems*, Princeton University Press, 1973.
- [14] J. MOSER AND E. ZEHNDER, *Lectures on Hamiltonian Systems*, Lecture Notes, ITCF, 1991.

- [15] J. PÖSCHEL, *Über invariante Tori in differenzierbaren Hamiltonschen Systemen*, Bonn. Math. Schr. **120** (1980), 1–103.
- [16] J. PÖSCHEL, *Integrability of Hamiltonian systems on Cantor sets*, Comm. Pure Appl. Math. **35** (1982), 653–695.
- [17] H. RÜSSMANN, *On optimal estimates for the solution of linear partial differential equations of first order with constant coefficients on the torus*, Dynamical Systems, Theory and Applications, J. MOSER (ed.), Lecture Notes in Physics 38, Springer, 1975, pp. 598–624.
- [18] H. RÜSSMANN, *Non-degeneracy in the perturbation theory of integrable dynamical systems*, Number Theory and Dynamical Systems, M.M. DODSON, J.A.G. VICKERS (eds.), London Math. Soc. Lecture Note Series 134, Cambridge University Press, 1989, pp. 5–18.
- [19] H. RÜSSMANN, *Invariant tori in the perturbation theory of weakly nondegenerate integrable Hamiltonian systems*, Preprint Mainz (1998).
- [20] D. SALAMON, *The Kolmogorov-Arnold-Moser theorem*, Preprint, ETH-Zürich, 1986.
- [21] D. SALAMON AND E. ZEHNDER, *KAM theory in configuration space*, Comment. Math. Helv. **64** (1989), 84–132.
- [22] M. SEVRYUK, *KAM-stable Hamiltonians*, J. Dyn. Control Syst. **1** (1995), 351–366.
- [23] C.L. SIEGEL AND J.K. MOSER, *Lectures on Celestial Mechanics*, Springer, 1971.
- [24] E. STEIN, *Singular Integrals and Differentiability Properties of Functions*, Princeton, 1970.
- [25] H. WHITNEY, *Analytic extensions of differentiable functions defined in closed sets*, Trans. A.M.S. **36** (1934), 63–89.
- [26] JUNXIANG XU, JIANGONG YOU AND QINGJIU QIU, *Invariant tori for nearly integrable Hamiltonian systems with degeneracy*, Math. Z. **226** (1997), 375–387.
- [27] E. ZEHNDER, *Homoclinic points near elliptic fixed points*, Comm. Pure Appl. Math. **26** (1973), 131–182.
- [28] E. ZEHNDER, *Generalized implicit function theorems with applications to some small divisor problems I*, Comm. Pure Appl. Math. **28** (1975), 91–140, *II*, **29**, 1976, 49–111.

JÜRGEN PÖSCHEL  
 Mathematisches Institut A, Universität Stuttgart  
 Pfaffenwaldring 57, 70569 Stuttgart, Germany

j@poschel.de