INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
INTERNATIONAL CENTRE FOR SCIENCE AND HIGH TECHNOLOGY
c/o INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS 34100 TRIESTE (ITALY) VIA GRIGNANO, 9 (ADRIATICO PALACE) P.O. BOX 586 TELEPHONE 040-224571 TELEFAX 040-224575 TELEX 460449 APH I

SMR.657/4

SCHOOL ON DYNAMICAL SYSTEMS
(25 May - 5 June 1992)

Lecture on the Classical KAM Theorem

J. Pöschel
E.T.H.-Zentrum
F.I.M.
CH-8092 Zürich
Switzerland

Version 1.0, May 1992

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, *K*olmogorov, *A*rnold and *M*oser (whence the akronym '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 [2,12] for details. Thus we are going to consider hamiltonians of the form

$$H(p,q) = h(p) + f_\epsilon(p,q), \qquad f_\epsilon(p,q) = \epsilon f(p,q,\epsilon)$$

for small ϵ, where $p = (p_1,\ldots,p_n)$ are the action variables varying over some domain $D \subset \mathbf{R}^n$, while $q = (q_1,\ldots,q_n)$ are the conjugate angular variables, whose domain is the usual n-torus $\mathbf{T}^n$ obtained from $\mathbf{R}^n$ by identifying points whose components differ by integer multiples of 2π. Thus, f_ϵ has period 2π 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,

$$\dot p = -H_q(p,q), \qquad \dot q = H_p(p,q)$$

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

$$\dot p = 0, \qquad \dot q = \omega$$

with

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

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

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

Hence, every solution curve is a straight line, which, due to the identification of the q-coordinates modulo 2π, is winding around the invariant torus $\{p_0\} \times \mathbf{T}^n$ with frequencies – or winding numbers – $\omega(p_0) = (\omega_1(p_0),\ldots,\omega_n(p_0))$.

Thus, the whole phase space is *foliated* into an n-parameter family of *invariant tori* $\{p_0\} \times \mathbf{T}^n$, on each of which the flow is linear with constant *frequencies* $\omega(p_0)$. That is, each torus is a so called *Kronecker system* $(\mathbf{T}^n, \omega)$. – 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 $q_1,\ldots,q_n$. 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

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

where $\langle\cdot,\cdot\rangle$ denotes the usual skalar 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 $(\mathbf{T}^n, \omega)$ crucially depends on the arithmetical properties of the frequencies ω. There are essentially two cases.

1. – The frequencies ω are *nonresonant*, or *rationally independent*:

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

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

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

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

The prototype is $\omega = (\omega_1,\ldots,\omega_{n-m},0\ldots,0)$, with $1 \leq m \leq n-1$ trailing zeroes and nonresonant $(\omega_1,\ldots,\omega_{n-m})$. 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 $\mathbf{T}^n$.

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

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*

$$h_p\colon\ D \to \Omega \subset \mathbf{R}^n, \quad p \mapsto \omega(p) = h_p(p).$$

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* $\Omega \subset \mathbf{R}^n$, and "the frequencies ω effectively depend on the amplitudes p", as a physicist would say. It follows that nonresonant tori and resonant tori of all types all 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 – 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 people say, "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 behaviour*. – 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*.

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 [6].

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 here 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 ω are not only nonresonant, but are *strongly nonresonant* in the sense that they satisfy a *diophantine*, or *small divisor condition*: there exist constants $\alpha > 0$ and $\tau > 0$ such that

$$|\langle k, \omega\rangle| \geq \frac{\alpha}{|k|^\tau} \qquad \text{for all } 0 \neq k \in \mathbf{Z}^n,$$

where $|k| = |k_1| + \cdots + |k_n|$.

The existence of such frequencies is easy to see. Let Δ_α^τ denote the set of all $\omega \in \mathbf{R}^n$ satisfying these infinitely many conditions. Then Δ_α^τ is the complement of the open dense set

$$R_\alpha^\tau = \bigcup_{0 \neq k \in \mathbf{Z}^n} R_{\alpha,k}^\tau, \qquad R_{\alpha,k}^\tau = \left\{ \omega \in \mathbf{R}^n : |\langle k, \omega\rangle| < \frac{\alpha}{|k|^\tau} \right\}.$$

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

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

provided $\tau > n-1$. 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 $\mathbf{R}^n$, for any $\tau > n-1$. In other words, *almost every* ω in $\mathbf{R}^n$ belongs to Δ^τ, $\tau > n-1$, and thus satisfies a diophantine condition for some $\alpha > 0$.

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

$$(1) \qquad \min_{k:0\neq \max|k_j|\leq K} |\langle k, \omega\rangle| \leq \frac{|\omega|}{K^{n-1}}$$

by Dirichlet's pigeon hole argument. And for $\tau = n-1$, the set Δ^{n-1} has measure zero, but Hausdorff dimension n - see [15] for references. So there are continuum many diophantine frequencies to the exponent $n-1$, although they form a set of measure zero.

From now on, we will fix $\tau > n-1$ once and for all 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_ϵ, no matter how small ϵ. The reason is that the parameter α 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 ϵ, only those Kronecker tori with frequencies ω in Δ_a,

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

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

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

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

whose frequencies belong to Δ_α and also have distance $\geq \alpha$ to the boundary of Ω. These, like Δ_α, are *Cantor sets*: they are closed, perfect and *nowhere dense*, hence of first Baire category. But they also have *large measure*:

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

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

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

The Classical KAM Theorem [1,7,8]. *Suppose the integrable hamiltonian h is nondegenerate, such that the frequency map h_p is a diffeomorphism $D \to \Omega$, and $H = h + f_\epsilon$ is real analytic on $\bar{D} \times \mathbf{T}^n$. Then there exists a constant $\delta > 0$ such that for*

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

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

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

Is is an immediate and important consequence of the KAM theorem that small perturbations of nondegenerate hamiltonians are *not ergodic*, as the collection of Kronecker tori forms 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.

We can draw a stronger conclusion, however, for the special case of *two* degree of freedom systems. Here the phase space is 4-dimensional, the invariant energy levels are 3-dimensional, and the invariant 2-tori in those energy levels have an inside and an outside. Hence, even when an orbit does not lie *on* a torus, it usually is contained *inside* some torus, which will confine its motion for all time and thus ensure stability, although not quasi-periodicity. – This trapping mechanism is not available for $n \geq 3$.

The question arises, what happens in the gaps? This is an area of active research, and some relevant keywords here are Nekhoroshev estimates, Arnold diffusion, and chaotic behaviour. But we will not discuss such matters.

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 [10,13,17]. For their Lipschitz dependence some more regularity is required [14].

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 $\mathbf{R}^n$ has measure zero. For example, if it happens that

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

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

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

For a more general statement see [16]. – Written proofs, however, are not yet available, so a cautious reader may consider these statements rather as (well founded) conjectures.

Finally, the hamiltonian nature of the equations is almost indispensable. Analogous result are true for reversible systems [11,14]. 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 [9].

To this end we write $p = p_0 + I$ and expand h around p_0 according to Taylor's formula:

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

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

$$h_p\colon \; D \to \Omega, \quad p_0 \mapsto \omega = h_p(p_0).$$

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

$$g_\omega\colon \; \Omega \to D, \quad \omega \mapsto p_0 = g_\omega(\omega),$$

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

Thus we can write

$$h(p) = e(\omega) + \langle \omega, I \rangle + P_h(I; \omega)$$
$$f_\epsilon(p,q) = f_\epsilon(p_0 + I, q) \overset{\text{def}}{=} P_{f_\epsilon}(I, q; \omega),$$

where P_h denotes the quadratic term in the expansion of h. Writing now θ instead of q for the angular variables, we obtain the family of hamiltonians

$$H = N(I; \omega) + P(I, \theta; \omega)$$

with

$$N = e(\omega) + \langle \omega, I \rangle, \qquad P = P_h(I; \omega) + P_{f_\epsilon}(I, \theta; \omega),$$

which is real analytic in the coordinates (I, θ) in $B \times \mathbf{T}^n$, B some sufficiently small ball around the origin in $\mathbf{R}^n$, as well as in the parameters ω in $\bar{\Omega}$. That is, the analyticity in ω extends to a uniform neighborhood of Ω.

Let us take this family as our new starting point. For $P = 0$ it reduces to the so called *normal form* $N = e(\omega) + \langle \omega, I \rangle$, which admits the invariant torus

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

with constant vectorfield $\dot{\theta} = \omega$. The aim is to prove the persistence of this torus and its vectorfield under sufficiently small perturbations $P \ne 0$, for all ω in the Cantor set $\Omega_\alpha \subset \Omega$. – 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 seperated from the actions. This makes their rôle more transparent. For example, the Lipschitz dependence of the tori on ω is easily established. — Generalizations such as weaker nondegeneracy conditions and extension to infinitie dimensional systems are easier. Also, this approach lends itself naturally to applications in bifurcation theory, where systems naturally depend on parameters.

b. We are going to prove the persistence of $\mathcal{T}_0$ by constructing a coordinate-parameter transformation $\mathcal{F}$, which takes H back into another normal form plus higher order terms. This transformation is of the form

$$\mathcal{F}\colon\ (I,\theta;\omega)\mapsto(\Phi(I,\theta;\omega);\phi(\omega)),$$

where Φ is real analytic and symplectic for each ω and of the form

$$\Phi\colon\ (I,\theta)\mapsto(U(I,\theta),V(\theta)),$$

where we did not indicate the dependence on ω. Moreover, U is affine linear in I. Such transformation form a group G under composition.

c. To state the basic result we need to introduce some notation. Let

$$D_{r,s}=\{|I|<r\}\times\{|\mathrm{Im}\,\theta|<s\}\subset\mathbf{C}^n\times\mathbf{C}^n$$

and

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

denote complex neighborhoods of $\mathcal{T}_0$ and Ω_α, 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 need to consider the Lipschitz norm - or rather *semi* norm - of $\mathcal{F}$ with respect to ω. We define

$$|\phi|_{\mathrm{Lip}}=\sup_{v\neq\omega}\frac{|\phi(v)-\phi(\omega)|}{|v-\omega|},$$

where the underlying domain will be clear from the context. $|\Phi|_{\mathrm{Lip}}$ is defined analogously, extending the sup over I and θ.

We recall that $\Omega_\alpha\subset\Omega$ denotes the Cantor set of all ω in Ω, which have distance $\geq\alpha$ to the boundary of Ω and satisfy

$$|\langle k,\omega\rangle|\geq\frac{\alpha}{|k|^\tau}\qquad\text{for all } 0\neq k\in\mathbf{Z}^n.$$

The exponent $\tau>n-1$ is fixed once and for all.

Theorem A. *Suppose that P is real analytic on $D_{r,s}\times O_h$, and*

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

where γ is a small constant depending only on n and τ, and $\nu=\tau+n+1$. Suppose also that $r,s,h\leq 1$. Then there exists a transformation

$$\mathcal{F}\colon\ D_{r/2,s/2}\times\Omega_\alpha\to D_{r,s}\times\Omega$$

in the group G described above, which is real analytic and symplectic for each ω and Lipschitz continous in ω, such that

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

where the dots stand for higher order terms in I. Moreover,

$$|W(\Phi-id)|,\,s^\nu|W(\Phi-id)|_{\mathrm{Lip}}\leq\frac{c}{\alpha r s^\nu}|P|_{r,s,h},$$
$$|\phi-id|,\,s^\nu|\phi-id|_{\mathrm{Lip}}\leq\frac{c}{r}|P|_{r,s,h}$$

uniformly on $D_{r/2,s/2}\times\Omega_*$, where c is a large constant depending only on n and τ, and $W=\begin{pmatrix}r^{-1}Id & \\ & s^{-1}Id\end{pmatrix}$, Id the identity matrix.

c. Theorem A states that over a Cantor set Ω_α of parameter values the perturbed family of hamiltonians H is transformed back into a family of hamiltonians $\tilde{H}$ of the form 'linear normal form plus higher order terms in I'. In these new coordinates-parameters we thus have, at $I=0$,

$$\dot{I}=-\tilde{H}_\theta|_{I=0}=0,\qquad\dot{\theta}=\tilde{H}_I|_{I=0}=\omega.$$

Hence, there is an invariant torus $\mathcal{T}_0=\{0\}\times\mathbf{T}^n$ with linear flow $\dot{\theta}=\omega$ for each $\omega\in\Omega_\alpha$.

In the original family of hamiltonians H we then have an embedded invariant torus

$$\mathcal{T}_\omega = \Phi_\omega(\mathcal{T}_0), \qquad \Phi_\omega = \Phi|_{I=0,\omega}: \ \mathbf{T}^n \hookrightarrow \mathbf{R}^n \times \mathbf{T}^n,$$

not at the parameter ω, but at the slightly shifted parameter

$$\tilde{\omega} = \phi(\omega).$$

In other words, for every parameter value $\tilde{\omega}$ in the slightly deformed Cantor set

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

the hamiltonian $H|_{\tilde{\omega}}$ admits an embedded Kronecker torus $\mathcal{T}_\omega$ with frequencies $\omega = \phi^{-1}(\tilde{\omega})$ in Ω_α. This torus is close to the unperturbed torus $\mathcal{T}_0$, since Φ is close to the identity.

Moreover, since ϕ is *Lipschitz* close to the identity, we can control the measure of $\tilde{\Omega}_\alpha$ to the effect that

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

The argument is the following.

It is a basic fact – recalled in Appendix B – that any Lipschitz continuous function on an arbitray closed subset of $\mathbf{R}^n$ can be extended to $\mathbf{R}^n$ without changing its Lipschitz constant. Thus we may extend every coordinate function of $\phi - id$, and we may even prescribe $\phi - id = 0$ outside of Ω, which affects the Lipschitz constant only verly little, since Ω_α has distance $\geq \alpha$ to the boundary of Ω. This way we obtain an extension $\bar{\phi}$ of ϕ to all of $\mathbf{R}^n$, which is still Lipschitz close to the identity, hence a Lipeomorphism of $\mathbf{R}^n$, and which is the identity outside Ω. Hence we have a Lipeomorphism

$$\bar{\phi}: \ \Omega \to \Omega, \qquad \bar{\phi} \mid \Omega_\alpha = \phi.$$

Now we can conclude that

$$\begin{aligned} m(\Omega - \tilde{\Omega}_\alpha) &= m(\Omega - \bar{\phi}(\Omega_\alpha)) \\ &= m(\bar{\phi}(\Omega - \Omega_\alpha)) \\ &\leq c m(\Omega - \Omega_\alpha) \\ &= O(\alpha), \end{aligned}$$

since the measure of sets mapped by Lipschitz maps is estimated in the same way as for C^1-maps.

We point out that such an estimate is not available, if ϕ were just a homeomorphism. For example, there exist homeomorphisms on the circle, which map Cantor sets of positive measure into Cantor sets of zero measure.

In the same way the coordinate functions of Φ are extended such that the extension is real analytic for each ω and Lipschitz in ω; we do not bother, however, to preserve the symplectic nature of the map under extension. We then arrive at the following conclusion.

Theorem B. *Suppose the assumptions of Theorem A are satisfied. Then there exists a Lipeomorphism $\bar{\phi}: \Omega \to \Omega$ close to the identity and a mapping*

$$\bar{\Phi}: \ \Omega \times \mathbf{T}^n \to \mathbf{R}^n \times \mathbf{T}^n$$

close to the trivial embedding $(\omega, \theta) \mapsto (0, \theta)$ such that for every parameter value

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

the hamiltonian $H|_{\tilde{\omega}}$ admits an invariant Kronecker torus $\mathcal{T}_\omega = \bar{\Phi}(\{\omega\} \times \mathbb{T}^n)$, where $\omega = \bar{\phi}^{-1}(\tilde{\omega})$. Moreover, the estimates for $\bar{\phi}$ and $\bar{\Phi}$ are the same as for ϕ and $\Phi|_{I=0}$ in Theorem A, though with a different constand c, and

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

where the implicit constant depends only on Ω.

We will see at the end of section 5 that the map ϕ actually can be assigned ω-derivatives of every order on the Cantor set Ω_α. This may be formalized by introducing the intrinsically defined notion of a differentiable function on an arbitrary closed set [20,21]. 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 ϕ to a C^∞-function $\bar{\phi}$ on Ω. The same applies to Φ and leads to the notion of smooth foliations of invariant tori over Cantor sets [14].

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

$$P = P_h + P_{f_\epsilon}$$

is real analytic on $B \times \mathbf{T}^n \times \tilde{\Omega}$, B some small ball around the origin in $\mathbf{R}^n$. 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

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

d some constant depending on H. To meet the smallness condition of Theorems A and B, we choose $r = \sqrt{\epsilon}$ and arrive at the condition

$$\epsilon \leq \tilde{\gamma}\alpha r s^\nu = \tilde{\gamma}\alpha s^\nu \cdot \sqrt{\epsilon}, \qquad \tilde{\gamma} = \frac{\gamma}{2d},$$

which amounts to

$$\epsilon \leq \alpha^2 \tilde{\gamma}^2 s^{2\nu} = \alpha^2 \delta, \qquad \delta = \tilde{\gamma} s^{2\nu}.$$

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

By construction, an orbit $(I(t), \theta(t))$ for the hamiltonian H at the parameter value $\tilde{\omega}$ translates into an orbit $(p(t) = p_0(\tilde{\omega}) + I(t), q(t) = \theta(t))$ for this hamiltonian in p,q-coordinates, where

$$p_0(\tilde{\omega}) = h_p^{-1}(\tilde{\omega}) = g_\omega(\tilde{\omega}).$$

It therefore follows with Theorem B that the mapping

$$\Psi\colon\ \Omega \times \mathbf{T}^n \to D \times \mathbf{T}^n,$$

given by

$$\begin{pmatrix} \omega \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} p = (h_p^{-1} \circ \tilde{\phi})(\omega) + \tilde{U}(0,\theta;\omega) \\ q = \qquad\qquad \tilde{V}(\theta;\omega) \end{pmatrix},$$

where $(\tilde{U}, \tilde{V}) = \tilde{\Phi}$, is an embedding of an invariant Kronecker torus $(\mathbf{T}^n, \omega)$ for every $\omega \in \Omega_\alpha$. Moreover, Ψ is *Lipschitz close* to the real analytic unperturbed embedding

$$\Psi_0\colon\ \Omega \times \mathbf{T}^n \to D \times \mathbf{T}^n, \qquad \begin{pmatrix} \omega \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} h_p^{-1}(\omega) \\ \theta \end{pmatrix}.$$

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 $\Omega_\alpha \times \mathbf{T}^n$, hence is $O(\alpha)$. This finishes the proof of the clasical KAM theorem.

3. Outline of the Proof of Theorem A

a. We prove Theorem A by a rapidly converging iteration procedure that was originally proposed by Kolmogorov and implemented by Arnold and Moser [1,7,8]. At each step of this scheme a hamiltonian

$$H_j = N_j + P_j$$

is considered, which is a small perturbation of some normal form $N_j = e_j + \langle \omega, I \rangle$. A transformation $\mathcal{F}_j = (\Phi_j, \phi_j)$ in the group G described in section 2.b 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,

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

for some $\kappa > 1$. Repetition of this process leads to a sequence of transformations $\mathcal{F}_0, \mathcal{F}_1, \ldots$, whose infinite product

$$\mathcal{F} = \lim_{j\to\infty} F_0 \circ \mathcal{F}_1 \circ \cdots \circ \mathcal{F}_j$$

also belongs to G and transforms the initial hamiltonian $H = H_0$ into some normal form up to first order in I.

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 [22], or by refering to an implicit function theorem in tame Frechet spaces [4]. Recently, Salamon and Zehnder gave a proof that avoids coordinate transformations altogether and works in configuration space [18], and Eliasson described a way of using power series expansions and majorant techniques in a very tricky way [5].

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 linearizing it in I and truncating its Fourier series in θ at some suitable order. $P-R$ will be small, and we now consider the hamiltonian $\bar{H} = N + R$ instead of $H = N + P$. The reason for this approximation will become clear later – hopefully.

The transformation $\mathcal{F}$ consists of a coordinate transformation Φ and a subsequent change ϕ of the frequency parameters. Φ is obtained as the time-1-map of the flow X_F^t of a hamiltonian vectorfield X_F, that is, of the system of ode's

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

Then Φ is symplectic for each ω. 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,

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

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 ϵ, 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 now to find F such that

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

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

(2) $$\{F, N\} + \hat{N} = R, \qquad \hat{N} = N_+ - N$$

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

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

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

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

c. Let us now consider equation (2) first on a formal level. Clearly,

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

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

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

with coefficients depending on I and ω, we find

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

Thus, ∂_ω admits a basis of eigenfunctions $e^{i\langle k, \theta \rangle}$ with eigenvalues $i\,\langle k, \omega \rangle$, $k \in \mathbf{Z}^n$. That is, ∂_ω diagonalizes with respect to this basis.

If ω 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 the zeroth one R_0, which is given by the mean value of R over $\mathbf{T}^n$,

$$[R] = R_0 = \frac{1}{(2\pi)^n} \int_{\mathbf{T}^n} R\, d\theta.$$

Hence, if R is given, then we can always formally solve the equation

$$\partial_\omega F = R - [R]$$

by setting

$$F = \sum_{0 \neq k \in \mathbf{Z}^n} \frac{R_k}{i\,\langle k, \omega \rangle} e^{i\langle k, \theta \rangle}. \tag{4}$$

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

Finally, equation (2) 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. As an aside we point out a more systematic interpretation of the preceding construction. For irrational ω, the domain of ∂_ω, consisting of all formal Fourier series in θ (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, ∂_ω 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 simply given by taking the mean value, so

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

The equation

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

is then simply solved by "solving componentwise",

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

where the latter can be solved uniquely for F in $\mathcal{R}$, since ∂_ω 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 (4) of F, we are confronted with the well known and notorious problem of "small divisors". Even if ω is nonresonant, infinitely many of the divisors $\langle k, \omega \rangle$ become arbitrarily small in view of (1), threatening to make the series (4) divergent.

This divergence is avoided, if ω is strongly nonresonant. To formulate this key lemma, let $\mathcal{A}^s$ denote the space of all analytic functions u defined in the complex strip $\{\theta : \sup_j |\mathrm{Im}\,\theta_j| < s\} \subset \mathbf{C}^n$ 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 $0 \neq k \in \mathbf{Z}^n$.

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

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

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

$$|u|_{s-\sigma} \le \frac{c}{\alpha \sigma^{\tau+n}} |v|_s,$$

where the constant c depends only on n and τ.

Proof. Expanding u and v into Fourier series, the unique formal solution u with $[u] = 0$ is

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

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

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

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

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

The infinite sum is now easily estimates by $c\sigma^{-\tau-n}$. ∎

The lemma is actually true with σ^τ in place of $\sigma^{\tau+n}$, but the proof is more involved [15,13].

We observe that ∂_ω^{-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 σ tends to zero. This phenomen is known as "loss of smoothness" affected by the solution operator ∂_ω^{-1}, and is the main culprit why small divisor problems are technically so involved. For example, during the iteration we have to let $\sigma \to 0$ in order to stay in the classes $\mathcal{A}$. But then $|\partial_\omega^{-1}| \to \infty$. 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 ω satisfies infinitely many small divisor conditions, thus restricting ω 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 ω-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 (2) we arrive at $H \circ \Phi = N_+ + P_+$ with $P_+ \ll P$ and

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

since $[R]$ is linear in I and independent of θ. 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 $\phi: \omega_+ \mapsto \omega$ by the implicit function theorem – see Appendix A. With this change of parameters,

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

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. Before giving the detailed estimates we observe that we may scale the parameter α to some convenient value, say, $\alpha = 2$, by multiplying N and P by $2/\alpha$. This rescales ω to $2\omega/\alpha$, so the domain Ω and the parameter h are scaled accordingly. From now on, we therefore consider the Cantor set $\Omega_* \subset \Omega$ of all frequencies ω in Ω, which have distance ≥ 2 to the boundary of Ω and satisfy

$$|\langle k,\omega\rangle| \geq \frac{2}{|k|^\tau} \qquad \text{for all } 0 \neq k \in \mathbf{Z}^n. \tag{5}$$

Accordingly, we set $O_h = \{|\omega - \Omega_*| < h\}$.

To avoid a flood of constants we will write

$$u < v, \qquad u \cdot\!< v,$$

if there exist positive constants $c > 1$ and $\gamma < 1$, which depend *only on* n *and* τ and could be made explicit, such that $u \leq cv$ and $u \leq \gamma v$, respectively.

b. 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*

$$\text{(a)} \qquad \epsilon \cdot\!< \eta r \sigma^\nu,$$

$$\text{(b)} \qquad \epsilon \cdot\!< hr,$$

$$\text{(c)} \qquad h \leq K^{-\tau-1},$$

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

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

in the group G *described in section 2.b such that* $H \circ \mathcal{F} = N_+ + P_+$ *with*

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

and $|N_+ - N \circ \phi|_{r,s,h/4} < \epsilon$. *Moreover,*

$$|W(\Phi - id)|, \left|W(D\Phi - Id)W^{-1}\right| < \frac{\epsilon}{r\sigma^\nu}$$

$$|\phi - id|, h\,|D\phi - Id| < \frac{\epsilon}{r}$$

uniformly on $D_{\eta r, s-5\sigma} \times O_h$ *and* $O_{h/4}$, *respectively, with the weight matrix* $W = \begin{pmatrix} r^{-1}Id & \\ & \sigma^{-1}Id \end{pmatrix}$.

c. 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 linear in I and a trigonometric polynomial in θ. 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

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

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

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

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

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

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

2. Extending the small divisor estimate. The nonresonance conditions (5) are assumed to hold on Ω_* only. But assumption (c) implies that

$$|\langle k,\omega\rangle| \geq \frac{1}{|k|^\tau} \qquad \text{for all } 0 \neq k \leq K \tag{6}$$

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

$$|\langle k, \omega - \omega_o\rangle| \leq |k| \cdot |\omega - \omega_o| \leq Kh \leq \frac{1}{K^\tau} \leq \frac{1}{|k|^\tau}$$

for $|k| \leq K$. Together with the estimate (5) for $\langle k, \omega_o\rangle$ this proves the claim.

3. Solving the linearized equation $\{F, N\} + \hat{N} = R$. We can now 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 ω in O_h because of (6) and the fact that R contains only Fourier coefficients up to order K, by truncation. Hence the estimate of Lemma 1 applies as well with $\alpha = 1$, and we obtain a real analytic function F with

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

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

$$\frac{1}{r}\,|F_\theta|_{r/2,s-3\sigma}\,,\ \frac{1}{\sigma}\,|F_I|_{r/2,s-3\sigma} < \frac{\epsilon}{r\sigma^\nu}$$

with $\nu = \tau + 1$.

4. Transforming the coordinates. The coordinate transformation Φ is obtained as the real analytic time-1-map of the flow X_F^t of the hamiltonian vectorfield X_F – that is, of the ode's

$$\dot I = -F_\theta, \qquad \dot\theta = F_I.$$

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

$$\Phi = X_F^t\big|_{t=1}:\ D_{r/4,s-4\sigma} \to D_{r/2,s-3\sigma}, \tag{7}$$

and

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

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

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

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

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

on the domain $D_{r/8,s-5\sigma} \supseteq D_{\eta r,s-5\sigma}$. This proves all the estimates for Φ. Finally, we observe that, since $|U - id| \le |F_\theta| \le \eta r$, also

$$\Phi:\ D_{\eta r,s-5\sigma} \to D_{2\eta r,s-4\sigma}.$$

5. New error term. To estimate P_+ as given in (3) we first consider $\{R, F\}$. Again, by Cauchy's estimate,

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

The same holds for $|\{\hat N, F\}|_{r/2,s-3\sigma}$. Together with (7) and $\eta < \frac{1}{8}$ we get

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

The other term in (3) is bounded by

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

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

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

$$\omega \mapsto \omega_+ = \omega + v(\omega), \qquad v = \hat N_I = [R_I]$$

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

$$|v|_{h/2} = |N_I|_{h/2} < \frac{\epsilon}{r} \le \frac{h}{4}.$$

The implicit function theorem of Appendix A applies, and there exists a real analytic inverse map

$$\phi:\ O_{h/4} \to O_{h/2}, \quad \omega_+ \mapsto \omega$$

with the estimates

$$|\phi - id|\,, h\,|D\phi - Id| < \frac{\epsilon}{r}$$

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

5. Iteration and Proof of Theorem A

a. We are now going to interate the KAM step infinitely often, choosing appropriate sequences for the parameters σ, η and so on. To motivate our choices, let us start by fixing a geometric sequence for σ, 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

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

Then we have

$$E_+ < \frac{1}{\eta}\left(E^2 + \left(\eta^2 + K^n e^{-K\sigma}\right) E\right).$$

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

$$E_+ < \frac{1}{\eta} E^2 = E^\kappa, \qquad \kappa = \frac{3}{2}.$$

That is, $E_+ \le c_1^{\kappa-1} E^\kappa$ for some constant c_1 determined by the KAM step and depending only on n and τ. Consequently,

$$c_1 E_+ \le (c_1 E)^\kappa,$$

and this scheme converges exponentially fast, if $c_1 E < 1$.

We still have to discuss our assumptions

(d) $$\eta^2 = E$$

(e) $$K^n e^{-K\sigma} \le E$$

as well as assumptions (a–c) of the KAM step. There is no obstacle to take (d) as the *definition* of η, as this *implies* (a). The other three conditions amount to

$$\frac{\epsilon}{r} < h \le K^{-\tau-1}, \qquad K^n e^{-K\sigma} \le E = \frac{\epsilon}{r\sigma^\nu}.$$

Given the rapid convergence of E we may set up similar sequences for h and K – say, $h_+ = h^\mu$ and $K_+ = K^\mu$ for $\mu = \frac{4}{3}$ – since then $K^n e^{-K\sigma}$ will converge even faster than E. We only need to make sure that these inequalities hold for the initial values $h_0, K_0, \ldots$. But here we may simply *define*

(f) $$K_0^{-\tau-1} = h_0 = \frac{c_0\epsilon_0}{r_0},$$

where c_0 is determined by the KAM step, and *fix*

(f) $$\frac{\epsilon_0}{r_0\sigma_0^\nu} = E_0 = \gamma_0$$

to some sufficiently small constant γ_0. This will make $K_0\sigma_0$ *large* so that the second inequality is satisfied.

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

$$\sigma_{j+1} = \frac{\sigma_j}{2}, \quad s_{j+1} = s_j - 5\sigma_j, \quad \sigma_0 = \frac{s_0}{20},$$

so that $s_0 > s_1 > \cdots \to s/2$. Let

$$E_{j+1} = c_1^{\kappa-1} E_j^\kappa, \quad h_{j+1} = h_j^\mu, \quad K_{j+1} = K_j^\mu,$$

with $\mu = \frac{4}{3}$ and $E_0, \epsilon_0, h_0, K_0$ related by (f–g) and c_1 given by the KAM step. Finally, let

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

and define the complex domains

$$D_j = \{|I| < r_j\} \times \{|\mathrm{Im}\,\theta| < s_j\}, \qquad O_j = \{|\omega - \Omega_*| < h_j\}.$$

Let $H = N + P$.

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

$$|P_0|_{r_0,s_0,h_0} \le \epsilon_0 = \gamma_0 r_0 \sigma_0^{\nu},$$

where γ_0 is sufficiently small depending on n and τ. Then for each $j \ge 0$ there exists a normal form N_j and a real analytic transformation

$$\mathcal{F}^j = \mathcal{F}_0 \circ \cdots \circ \mathcal{F}_{j-1}\colon\ D_j \times O_j \to D_0 \times O_0$$

in the group G defined in section 2.b such that $H \circ \mathcal{F}^j = N_j + P_j$ with

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

Moreover, $|N_{j+1} - N_j| < \epsilon_j$ and

$$\left|\bar{W}_0(\mathcal{F}^{j+1} - \mathcal{F}^j)\right|, \left|T\mathcal{F}^{j+1} - T\mathcal{F}^j \circ \mathcal{F}_j\right| < \frac{\epsilon_j}{r_j h_j}$$

uniformly on $D_{j+1} \times O_{j+1}$, where $\bar{W}_0 = \begin{pmatrix} r_0^{-1} Id & & \\ & \sigma_0^{-1} Id & \\ & & h_0^{-1} Id \end{pmatrix}$, and $T\mathcal{F}^j = \dfrac{\partial \mathcal{F}^j}{\partial I}$ denotes the coefficient matrix of the I-linear component of $\mathcal{F}$.

Here and in the following we ignore the energy constants in the normal forms, since they serve no purpose. But note that the linear terms of N_j and N_{j+1} differ through the frequency drift caused by the map ϕ.

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 $j \ge 0$. But (a) is satisfied by the definition of η_j, and (b–c) hold by the definition of h_j and K_j and the choice of their initial values, fixing $E_0 = \gamma_0$ sufficiently small

We obtain a transformation

$$\mathcal{F}_j\colon D_{j+1} \times O_{j+1} \to D_j \times O_j$$

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

$$\begin{aligned} |P_{j+1}| &< \epsilon_j E_j + \left(\eta_j^2 + K_j^n e^{-K_j \sigma_j}\right) \epsilon_j < \epsilon_j E_j \\ &< \frac{E_j^2}{\eta_j} r_{j+1} \sigma_{j+1}^{\nu} < E_j^{\kappa} r_{j+1} \sigma_{j+1}^{\nu} = c_1^{1-k} E_{j+1} r_{j+1} \sigma_{j+1}^{\nu}. \end{aligned}$$

Fixing c_1 suitably we get $|P_{j+1}| \le \epsilon_{j+1}$. Thus, the transformation $\mathcal{F}^{j+1} = \mathcal{F}^j \circ \mathcal{F}_j = \mathcal{F}_0 \circ \cdots \circ \mathcal{F}_j$ takes H into $N_{j+1} + P_{j+1}$ with the proper estimate for P_{j+1}. Moreover, ignoring the energy constants, we have $|N_{j+1} - N_j| < |\phi_j - id| \cdot r < \epsilon_j$.

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

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

uniformly on $D_{j+1} \times O_{j+1}$, where $\bar{D}$ denotes the Jacobian with respect to I, θ and ω, and $\bar{W}_j = \begin{pmatrix} r_j^{-1} Id & & \\ & \sigma_j^{-1} Id & \\ & & h_j^{-1} Id \end{pmatrix}$. We then have

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

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 $\bar{D}\mathcal{F}^j = \bar{D}\mathcal{F}_0 \circ \cdots \circ \bar{D}\mathcal{F}_{j-1}$, with the Jacobians evaluated at different points. Since $\left|\bar{W}_i \bar{W}_{i+1}^{-1}\right| \le 1$ for all i, we can use a telescoping argument and the inductiv estimates for the $\mathcal{F}_i$ to obtain

$$\begin{aligned} \left|\bar{W}_0 \bar{D}\mathcal{F}^j \bar{W}_{j-1}^{-1}\right| &\le \left|\bar{W}_0 \bar{D}\mathcal{F}_0 \circ \cdots \circ \mathcal{F}_{j-1} \bar{W}_j^{-1}\right| \\ &\le \left|\bar{W}_0 \bar{D}\mathcal{F}_0 \bar{W}_0^{-1}\right| \left|\bar{W}_0 \bar{W}_1^{-1}\right| \cdot \cdots \\ &\qquad \left|\bar{W}_{j-1} \bar{D}\mathcal{F}_{j-1} \bar{W}_{j-1}^{-1}\right| \left|\bar{W}_{j-1} \bar{W}_j^{-1}\right| \\ &\le \prod_j \left(1 + \frac{c_2 \epsilon_j}{r_j h_j}\right), \end{aligned}$$

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

To prove the other estimate we observe that

$$T\mathcal{F}^{j+1} = T\left(\mathcal{F}^j \circ \mathcal{F}_j\right) = T\mathcal{F}^j \circ \mathcal{F}_j \cdot T\mathcal{F}_j,$$

since θ and ω are transformed independently of I. Therefore,

$$|T\mathcal{F}^{j+1} - T\mathcal{F}^j \circ \mathcal{F}_j| \le |T\mathcal{F}^j \circ \mathcal{F}_j|\,|T\mathcal{F}_j - Id|.$$

The first factor is uniformly bounded by a similar telescoping argument, while

$$|T\mathcal{F}_j - Id| < \frac{\epsilon_j}{r_j\sigma_j^\nu} < \frac{\epsilon_j}{r_j h_j}$$

by the estimates of the KAM step. This finishes the proof of the Iterative Lemma. ∎

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 $h_0 = c_0\epsilon_0/r_0 = c_0\gamma_0\sigma_0^\nu \le s_0^\nu \le h$ by assumption, and we can fix the constant γ in Theorem A sufficiently small so that

$$|P_0|_{r_0,s_0,h_0} \le |P|_{r,s,h} \le \epsilon \le \gamma\alpha r s \le \epsilon_0 = \gamma_0 r_0 \sigma_0^\nu,$$

recalling that we normalized α to the constant $\alpha = 2$.

By the estimate of the Iterative Lemma the maps $\mathcal{F}^j$ and subsequently their I-derivatives $T\mathcal{F}^j$ converge uniformly on the set

$$D_* \times \Omega_* = \bigcap_{j\ge 0} D_j \times O_j, \qquad D_* = \{0\} \times \{|\mathrm{Im}\,\theta| < s/2\},$$

to mappings $\mathcal{F}_*$ and $T\mathcal{F}_*$ that are real analytic in θ and uniformly continuous in ω. Moreover,

$$|\bar{W}_0(\mathcal{F}_* - id)|, |T\mathcal{F}_* - Id| < \frac{\epsilon_0}{r_0 h_0}$$

by the usual telescoping estimates.

It seems to be very unfortunate that D_* contains *no open* I-domain. But, by construction, the $\mathcal{F}^j$ are *affine linear* maps in each I-fiber over $\mathbf{T}^n \times \Omega_*$, and we have just proven the convergence of their zeroth and first order terms in I. Therefore, the $\mathcal{F}^j$ indeed converge *uniformly* on the domain $D_{r/2,s/2} \times \Omega_*$ to a map $\mathcal{F}$ in the group G that is real analytic and symplectic for each ω and uniformly continous in ω. Moreover, piecing together the above estimates,

$$|\bar{W}_0(\mathcal{F} - id)| < \frac{\epsilon_0}{r_0 h_0} \le \gamma_0,$$

hence $\mathcal{F}$ maps $D_{r/2,s/2} \times \Omega_*$ into $D_{r,s} \times \Omega$.

The estimates $|N_{j+1} - N_j| < \epsilon_j$ and $|H \circ \mathcal{F}^j - N_j| = |P_j| < \epsilon_j$ together with Cauchy's estimate now imply that these functions converge to zero together with their first I-derivative, because $\epsilon_j/r_j \to 0$. It follows that N_j tends to some normal form N_*, and $H \circ \mathcal{F}^j$ tends to N_* up to terms of first order. That is,

$$H \circ \mathcal{F} = N_* + \cdots = e_*(\omega) + \langle \omega, I\rangle + \cdots,$$

as we wanted to show. – Observe that there is *no* control over the second I-derivatives, since ϵ_j/r_j^2 *diverges*. The scheme was just designed to normalize the hamiltonian H up to first order.

Let us now look at the ω-derivatives of the $\mathcal{F}^j$. Since E_j converges to zero at a faster *exponential* rate than h_j, we have

$$\frac{\epsilon_j}{r_j h_j^\lambda} \to 0 \qquad \text{for all } \lambda \ge 0.$$

Hence, *all* ω-derivatives of the $\mathcal{F}^j$ converge uniformly on $D_* \times \Omega_*$, and we could assign ω-derivatives of any order to the limit map $\mathcal{F}$ on the Cantor set Ω_* [20]. Without making this concept precise, however, we can at least conclude that $\mathcal{F}$ is *Lipschitz continuous* in ω. Its Lipschitz norm is bounded by the limit of the bounds on the first ω-derivatives of the $\mathcal{F}^j$. The usual Cauchy estimate yields

$$|\bar{W}_0(\mathcal{F} - id)|_{\mathrm{Lip}} < \frac{\epsilon_0}{r_0 h_0^2} = \frac{1}{c_0\gamma_0\sigma_0^\nu}$$

on $D_{r/2,s/2} \times \Omega_*$ by the definition of h_0 and ϵ_0.

We finally look at the estimates of $\mathcal{F}$. So far they do not reflect the actual size ϵ of the perturbation, since we fixed ϵ_0 independently of ϵ. But we observe that everything is still alright if in all the estimates for P_j, $\mathcal{F}_j$ and $\mathcal{F}^j$, the ϵ_j are scaled down by the linear factor

$$\frac{\epsilon}{\epsilon_0} = \frac{\epsilon}{\gamma_0 r_0 \sigma_0^\nu} = \frac{c\epsilon}{r s^\nu}.$$

Scaling down our estimates of $\mathcal{F}$ by this factor we can finally extract our estimates of Φ and ϕ as stated in Theorem A. 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 $\mathbf{C}^n$, 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 rom the Cauchy estimate for one complex variable.

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

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

$$|v_k| \leq |v|_s e^{-|k|s}, \qquad k \in \mathbf{Z}^n.$$

Proof. The Fourier coefficients v_k of v are given by

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

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

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

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

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

for all $\sigma > 0$. Letting $\sigma \to 0$ the lemma follows. ∎

We can now also estimate very roughly the remainder, when we truncate the Fourier series of v at order K.

Lemma A.2. *Let $T_K v = \sum_{|k| \leq K} v_k e^{i\langle k,\theta\rangle}$. If $v \in \mathcal{A}^s$, then*

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

where 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 ψ_K. For instance, one could take $\psi_K(x) = \psi(x/K)$, where ψ 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,

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

amounts to a convolution of v with a real analytic approximation of the identity $\hat{\psi}_K$, as $K \to \infty$. Such smoothing operators have many interesting properties. For more details, see for example [22].

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 Ω of $\mathbf{R}^n$. 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 real analytic from O_h into $\mathbf{C}^n$. If*

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

on O_h, then f has a real analytic inverse ϕ on $O_{h/4}$. Moreover,

$$|\phi - id|, \quad \frac{h}{4}\,|D\phi - Id| \le \delta$$

on this domain.

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

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

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

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

by Cauchy's inequality and so

$$|u - v| \le |Df - I|\,|u - v| \le \theta\,|u - v|$$

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

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

$$\begin{aligned} |D\phi - I|_\eta &= \left|(Df)^{-1} \circ \phi - I\right|_\eta \\ &\le \left|(Df)^{-1} - I\right|_{2\eta} \\ &\le \left(1 - |Df - I|_{2\eta}\right)^{-1} - 1 \\ &\le \frac{1}{1 - \delta/2\eta} - 1 \\ &\le \frac{\delta}{\eta} \end{aligned}$$

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

B. Lipschitz Functions

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

$$|u|_{\mathrm{Lip},B} = \sup_{\substack{x,y \in B \\ x \ne y}} \frac{|u(x) - u(y)|}{|x - y|},$$

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

$$U|_B = u, \qquad |U|_{\mathrm{Lip},\mathbf{R}^n} = |u|_{\mathrm{Lip},B},$$

In fact, B could be *any* point set.

Indeed, U is simply given by

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

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

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

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

$$U(x) - U(y) \le \lambda\,|x - y|.$$

Interchanging x and y we obtain

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

whence $|U|_{\mathrm{Lip},\mathbb{R}^n} \leq |u|_{\mathrm{Lip},B}$. We leave it 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'apres Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...)', Séminaire Bourbaki, *Astérisque* **133–134**, 1986, 113–157.

[5] L.H. ELIASSON, 'Compensation of signs in a small divisor problem', Preprint.

[6] A.B. KATOK, 'Ergodic properties of degenerate integrable Hamiltonian systems', *Math. USSR Izv.* **7**, 1973, 185–214.

[7] 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

[8] J. MOSER, 'On invariant curves of area preserving mappings of an annulus', *Nachr. Akad. Wiss. Gött., Math. Phys. Kl.*, 1962, 1–20.

[9] J. MOSER, 'Convergent series expansions for quasi-periodic motions', *Math. Ann.* **169**, 1967, 136–176.

[10] J. MOSER, 'On the continuation of almost periodic solutions for ordinary differential equations', in *Proc. Int. Conf. Func. Anal. and Rel. Topics*, Tokyo, 1969, pp. 60–67.

[11] J. MOSER, *Stable and Random Motions in Dynamical Systems*, Princeton University Press, 1973.

[12] J. MOSER AND E. ZEHNDER, *Lectures on Hamiltonian Systems*, Lecture Notes, ITCP, 1991.

[13] J. PÖSCHEL, 'Über invariante Tori in differenzierbaren Hamiltonschen Systemen', *Bonn. Math. Schr.* **120**, 1980, 1–103.

[14] J. PÖSCHEL, 'Integrability of Hamiltonian systems on Cantor sets', *Comm. Pure Appl. Math.* **35**, 1982, 653–695.

[15] H. RÜSSMANN, 'On optimal estimates for the solution of linear partial differential equations of first order with constant coefficients on the torus', in *Dynamical Systems, Theory and Applications*, J. MOSER (ed.), Lecture Notes in Physics 38, Springer, 1975, pp. 598–624.

[16] H. RÜSSMANN, 'Non-degeneracy in the perturbation theory of integrable dynamical systems', in *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.

[17] D. SALAMON, 'The Kolmogorov-Arnold-Moser theorem', Preprint, ETH-Zürich, 1986.

[18] D. SALAMON AND E. ZEHNDER, 'KAM theory in configuration space', *Comment. Math. Helv.* **64**, 1989, 84–132.

[19] C.L. SIEGEL AND J.K. MOSER, *Lectures on Celestial Mechanics*, Springer, 1971.

[20] E. STEIN, *Singular Integrals and Differentiability Properties of Functions*, Princeton, 1970.

[21] H. WHITNEY, 'Analytic extensions of differentiable functions defined in closed sets', *Trans. A.M.S.* **36**, 1934, 63–89.

[22] E. ZEHNDER, 'Generalized implicit function theorems with applications to some small divisor problemss I', *Comm. Pure Appl. Math.* **28**, 1975, 91–140. *II*, **29**, 1976, 49–111