M P E J

Mathematical Physics Electronic Journal

ISSN 1086-6655
Volume 10, 2004

Paper 3
Received: Feb 17, 2004, Revised: Feb 26, 2004, Accepted: Mar 1, 2004
Editor: R. de la Llave

The Kolmogorov–Arnold–Moser theorem

Dietmar A. Salamon
ETH-Zürich

26 February 2004

Abstract

This paper gives a self contained proof of the perturbation theorem for invariant tori in Hamiltonian systems by Kolmogorov, Arnold, and Moser with sharp differentiability hypotheses. The proof follows an idea outlined by Moser in [16] and, as byproducts, gives rise to uniqueness and regularity theorems for invariant tori. [1]

1 Introduction

KAM theory is concerned with the existence of invariant tori for Hamiltonian differential equations. More precisely, we consider the following situation. Let $H(x, y)$ be a smooth real valued function of $2n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n$. We assume that H is of period 1 in each of the variables $x_1, \ldots, x_n$ and y is restricted to an open domain $G \subset \mathbb{R}^n$ so that H is defined on $\mathbb{T}^n \times G$ where $\mathbb{T}^n :=$

[1] The present paper was written in 1986 while I was a postdoc at ETH Zürich. I didn't publish it at the time because the results are well known and the paper is of expository nature. The paper is reproduced here with the following changes: there are a few updates in the introduction, a mistake in Lemma 3 and the proof of Theorem 2 has been corrected, the hypotheses of Theorem 2 have been weakened, and Lemma 5 has been moved to Section 3. The original manuscript can be found on my webpage **http://www.math.ethz.ch/ salamon/publications.html**.

$\mathbb{R}^n/\mathbb{Z}^n$ denotes the n-torus. Consider the Hamiltonian differential equation

$$\dot{x} = H_y, \qquad \dot{y} = -H_x, \tag{1.1}$$

where $H_x \in \mathbb{R}^n$ and $H_y \in \mathbb{R}^n$ denote the vectors of partial derivatives of H with respect to x_ν and y_ν. As a matter of fact, any Hamiltonian vector field on a symplectic manifold can, in a neighborhood of a Lagrangian invariant torus, be represented in the form (1.1).

By an invariant torus for (1.1) in parametrized form with prescribed frequencies $\omega_1, \ldots, \omega_n$ we mean an embedding $x = u(\xi)$, $y = v(\xi)$ of $\mathbb{T}^n$ into $\mathbb{T}^n \times G$ such that $u(\xi) - \xi$ and $v(\xi)$ are of period 1 in all variables, u represents a diffeomorphism of the torus, and the solutions of the differential equation

$$\dot{\xi} = \omega$$

are mapped to solutions of (1.1). Equivalently, u and v satisfy the degenerate, nonlinear partial differential equations

$$Du = H_y(u, v), \qquad Dv = -H_x(u, v), \tag{1.2}$$

where D denotes the first order differential operator

$$D := \sum_{\nu=1}^{n} \omega_\nu \frac{\partial}{\partial \xi_\nu}.$$

Observe that any such invariant torus consists of quasiperiodic solutions of the Hamiltonian differential equation (1.1). In fact, the solutions are periodic if all the frequencies $\omega_1, \ldots, \omega_n$ are integer multiples of a fixed number and they lie dense on the torus if the frequencies are rationally independent.

The simplest case is where the function $H = H(y)$ is independent of x so that each of the sets $y =$ constant defines an invariant torus which can be represented by the functions $u(\xi) = \xi$ and $v(\xi) = y$. The frequency vector is then given by $\omega = H_y(y)$ and will in general depend on the value of y. This situation corresponds to an integrable Hamiltonian system and must be considered as the exceptional case. In fact, it has already been shown by Poincaré that in general one cannot expect equation (1.2) to have a solution for all values of ω in an open region, even if the Hamiltonian function $H(x, y)$ is close to an integrable one. The remarkable discovery of KAM theory was that even though a solution of (1.2) cannot be expected for rational frequency vectors it does indeed exist if the frequencies $\omega_1, \ldots, \omega_n$ are rationally independent and, moreover, satisfy the Diophantine inequalities

$$j \in \mathbb{Z}^n \setminus \{0\} \qquad \Longrightarrow \qquad |\langle j, \omega \rangle| \geq \frac{1}{c_0 |j|^\tau} \tag{1.3}$$

for some constants $c_0 > 0$ and $\tau > n - 1$. Here $|j|$ denotes the Euclidean norm.

Actually, only the perturbation problem for equation (1.2) is treated in KAM theory. More precisely, in addition to (1.3) we will have to assume that the

Hamiltonian function $H(x, y)$ is sufficiently close to a function $F(x, y)$ for which a solution is known to exist. After a suitable coordinate transformation this amounts to the condition

$$F_x(x, 0) = 0, \qquad F_y(x, 0) = \omega, \tag{1.4}$$

so that the invariant torus for F can be represented by the functions $u(\xi) = \xi$ and $v(\xi) = 0$. Moreover, we assume that the unperturbed Hamiltonian function satisfies the Legendre type condition

$$\det\left(\int_{\mathbb{T}^n} F_{yy}(x, 0)\, dx\right) \neq 0. \tag{1.5}$$

Under these assumptions the theorem asserts that the invariant torus with prescribed frequencies $\omega_1, \ldots, \omega_n$ survives under sufficiently small perturbations of the Hamiltonian function and its derivatives.

The perturbation theory described above goes back over forty years to the work of Kolmogorov [10, 11], Arnold [3, 4], and Moser [13, 14], Actually, the history of the problem of finding quasiperiodic solutions of Hamiltonian differential equations is much longer. In particular, it was of great interest to Poincaré and Weierstrass especially because of its relevance for the stability theory in celestial mechanics. For a more detailed account of the historical development the interested reader is referred to [17]. Closely related developments in the theory of Hamiltonian systems like the theory of Mather and Aubry as well as connections to various other fields in mathematics are described in [19].

The present paper builds on the ideas developed by Kolmogorov, Arnold, Moser and also by Rüssmann, Zehnder, Pöschel and many others. Its purpose is to summarize some of their work and to present a complete proof of the above mentioned perturbation theorem for invariant tori. We are particularly interested in the minimal number of derivatives required for F and H. The main ideas of this proof were outlined by Moser [16] and the details were later provided by Pöschel [22, 23] in a somewhat different way than it is done here. Moser's proof may be roughly sketched as follows. The first step is to prove a theorem for analytic Hamiltonians which involves quantitative estimates (Theorem 1). For differentiable Hamiltonians the result will then be obtained by approximating $H(x, y)$ with a sequence of real analytic functions - again including quantitative estimates - and then applying Theorem 1 to each element of the sequence (Theorem 2).

A beautiful presentation of similar ideas can be found in [15] in connection with Arnold's theorem about vector fields on a torus [2]. Moser has also used these methods for proving his perturbation theorem for minimal foliations for variational problems on a torus. This result was first stated in [18] and its proof appeared in [20]. In [9] Jacobowitz has applied the same approach to Nash's embedding problem for compact Riemannian manifolds [21] in order to reduce the number of derivatives to $\ell > 2$. Later Zehnder [28] has developed a general abstract setup for small divisor problems along these lines. Other versions of such abstract implicit function theorems can be found in the work of Hamilton [6] and Hörmander [8].

In connection with his paper [16] Moser produced a set of handwritten notes in which he sketched the necessary estimates for the proofs of the theorems presented here. The present paper is to a large extent based on those notes. Incidentally, also Pöschel's paper [22] was based on Moser's notes. Both in [16] and [22] it is essential that the unperturbed Hamiltonian function F is analytic whereas, changing the proof slightly, one can dispense with the unperturbed function F and instead assume that H is a C^ℓ Hamiltonian function that satisfies (1.5) and is an *approximate solution* of (1.4) in a suitable sense. Here

$$\ell > 2\tau + 2 > 2n$$

and τ is the number appearing in (1.3). The resulting solutions u and v of (1.2) are of class C^m, where $m < \ell - 2\tau - 2$, and the function $v \circ u^{-1}$, whose graph is the invariant torus, is of class $C^{m+\tau+1}$ (Theorem 2). These are precisely the regularity requirements which we need for the solutions of (1.2) to be locally unique (Theorem 3). Combining these two results we obtain that every solution of (1.2) of class $C^{\ell+1}$ for a Hamiltonian function of class C^∞ must itself be of class C^∞ (Theorem 4). We point out that the result presented here is close to the optimal one as in [7] Herman gave a counterexample concerning the nonexistence of an invariant curve for an annulus mapping of class $C^{3-\varepsilon}$. Translated to our situation this corresponds to the case $n = 2$ with $\ell = 4 - \varepsilon$. Other counterexamples with less smoothness were given by Takens [27] and Mather [12].

Before entering into the details of the proof we shall discuss a few fundamental properties of invariant tori. If the frequencies $\omega_1, \ldots, \omega_n$ are rationally independent then it follows from (1.2) that the torus is a Lagrangian submanifold of $\mathbb{T}^n \times G$ or, equivalently,

$$u_\xi^T v_\xi = v_\xi^T u_\xi. \tag{1.6}$$

This implies that the embedding of the torus extends to a symplectic embedding $z = \phi(\zeta)$ where $z = (x, y)$, $\zeta = (\xi, \eta)$, and

$$x = u(\xi), \qquad y = v(\xi) + u_\xi^T(\xi)^{-1}\eta. \tag{1.7}$$

In this notation u and v satisfy equations (1.2) if and only if the transformed Hamiltonian function $K := H \circ \phi$ satisfies

$$K_\xi(\xi, 0) = 0, \qquad K_\eta(\xi, 0) = \omega \tag{1.8}$$

Note that the transformations of the form (1.7) with u and v satisfying (1.6) form a subgroup of the group of symplectic transformations of $\mathbb{T}^n \times \mathbb{R}^n$.

It follows from (1.6) that the transformation $z = \phi(\zeta)$ defined by (1.7) can be represented in terms of its generating function

$$S(x, \eta) := U(x) + \langle V(x), \eta \rangle$$

where the scalar function $U(x)$ and the vector function $V(x)$ are chosen as to satisfy

$$V \circ u = \mathrm{id}, \qquad U_x \circ u = v, \tag{1.9}$$

so that $z = \phi(\zeta)$ if and only if $y = S_x$ and $\xi = S_\eta$. Observe that the functions $V(x) - x$ and $U_x(x)$ are of period 1 in all variables. The invariant torus can now be represented as the graph of

$$y = U_x(x) \tag{1.10}$$

and the flow on it can be described by the differential equation

$$\dot{x} = H_y(x, U_x). \tag{1.11}$$

The requirement that equation (1.10) defines an invariant torus is equivalent to the Hamilton-Jacobi equation

$$H(x, U_x) = constant. \tag{1.12}$$

In particular, the Hamiltonian system (1.1) can be understood in terms of the characteristics of the partial differential equation (1.12). Thus we are trying to find a solution of the partial differential equation (1.12) such that U_x is of period 1 and such that there is a diffeomorphism $\xi = V(x)$ of the torus which transforms the differential equation (1.11) into $\dot{\xi} = \omega$. The latter condition can be expressed by the formula

$$V_x H_y(x, U_x) = \omega. \tag{1.13}$$

Equations (1.12) and (1.13) together are equivalent to (1.8) and hence to (1.2).

2 The analytic case

The existence proof of invariant tori for analytic Hamiltonians goes back to Kolmogorov [10]. It was his idea to solve in each step of the iteration a linearized version of equations (1.12) and (1.13). The complete argument of the proof was given by Arnold [3] for the first time and in [16] Moser provided the quantitative estimates. The latter play an essential role in this paper.

In the course of the iteration we shall need an estimate for the solutions of the degenerate, linear partial differential equation $Df = g$. Such an estimate was given by Arnold [3] and Moser [15] and was later improved by Rüssmann [24, 25]. A slightly modified proof of Rüssmann's result can be found in Pöschel [22]. For the convenience of the reader we include a proof here (Lemma 2).

This requires some preparation. We introduce the space W_r of all bounded real analytic functions $w(\xi)$ in the strip $|\mathrm{Im}\,\xi| \le r$, $\xi \in \mathbb{C}^n$, which are of period 1 in all variables. Here we denote by $|\mathrm{Im}\,\xi|$ the Euclidean norm of the vector $\mathrm{Im}\,\xi \in \mathbb{R}^n$. Moreover, we introduce the norms

$$|w|_r := \sup_{|\mathrm{Im}\,\xi| \le r} |w(\xi)|, \qquad \|w\|_r := \sup_{|v| \le r} \left(\int_{\mathbb{T}^n} |w(u+iv)|^2 \right)^{1/2},$$

for $w \in W_r$ and we denote by W_r^0 the subspace of those functions $w \in W_r$ which have mean value zero on $\operatorname{Im}\xi = 0$. Observe that every $w \in W_r$ can be represented by its Fourier series

$$w(\xi) = \sum_{j\in\mathbb{Z}^n} w_j e^{2\pi i\langle j,\xi\rangle}, \qquad w_j := \int_{\mathbb{T}^n} w(u+iv)e^{-2\pi i\langle j,u\rangle}\,du\; e^{2\pi\langle j,v\rangle}.$$

Here the formula for w_j holds for every $v \in \mathbb{R}^n$ with $|v| \le r$ and $w_{-j} = \bar{w}_j$. Note also that $w \in W_r^0$ if and only if $w_0 = 0$.

Lemma 1. *There exists a constant $c = c(n) > 0$ such that the following inequalities hold for every $w \in W_r$ and every $\xi \in \mathbb{C}^n$ with $|\mathrm{Im}\xi| \le \rho < r \le 1$.*

(i) $\|w\|_r \le |w|_r$ *and* $|w_j| \le \|w\|_r e^{-2\pi|j|r}$.

(ii) $|w(\xi)| \le \sum_{j\in\mathbb{Z}^n} |w_j| e^{-2\pi\langle j,\operatorname{Im}\xi\rangle} \le c(r-\rho)^{-n/2}\|w\|_r$.

(iii) $|w_\xi|_\rho \le (r-\rho)^{-1}|w|_r$.

Proof. The first inequality in (i) is obvious and the second is a consequence of the following estimate for $v := -r|j|^{-1}j$:

$$|w_j| \le \int_{\mathbb{T}^n} |w(u+iv)|\,du\; e^{2\pi\langle j,v\rangle} \le \|w\|_r e^{-2\pi|j|r}.$$

The first inequality in (ii) is again obvious. To establish the second inequality in (ii) fix a vector $\xi \in \mathbb{C}^n$ with $|\operatorname{Im}\xi| \le \rho$ and consider the set $J_0 \subset \mathbb{Z}^n$ of those integer vectors $j \in \mathbb{Z}^n$ that satisfy

$$2\langle j, \operatorname{Im}\xi\rangle < -|j|\rho.$$

We will use the identity

$$\int_{\mathbb{T}^n} |w(u+iv)|^2\,du = \sum_{j\in\mathbb{Z}^n} |w_j|^2 e^{-4\pi\langle j,v\rangle}$$

and obtain with $\mu := r/\rho > 1$ that

$$\begin{aligned}
\sum_{J_0} |w_j| e^{-2\pi\langle j,\operatorname{Im}\xi\rangle} &\le \sum_{J_0} |w_j| e^{-2\pi\langle j,\mu\operatorname{Im}\xi\rangle} e^{-\pi|j|(r-\rho)} \\
&\le \left(\sum_{J_0} |w_j|^2 e^{-4\pi\langle j,\mu\operatorname{Im}\xi\rangle}\right)^{1/2} \left(\sum_{J_0} e^{-2\pi|j|(r-\rho)}\right)^{1/2} \\
&\le \frac{c_1^{1/2}}{(r-\rho)^{n/2}}\|w\|_r,
\end{aligned}$$

where

$$c_1 = c_1(n) := \sup_{0<\lambda\le 1} \sum_{j\in\mathbb{Z}^n} \lambda^n e^{-2\pi|j|\lambda} < \infty.$$

Now let $e_1, \ldots, e_s$ be a collection of unit vectors in $\mathbb{R}^n$ such that, for every $x \in \mathbb{R}^n$, there exists a $\sigma \in \{1, \ldots, s\}$ with $\langle x, e_\sigma \rangle \geq |x|/2$. Then every integer vector outside J_0 belongs to one of the sets J_σ of all those $j \in \mathbb{Z}^n$ that satisfy

$$2\langle j, \operatorname{Im} \xi \rangle \geq -|j|\rho, \qquad 2\langle j, e_\sigma \rangle \geq |j|.$$

Moreover, for $\sigma = 1, \ldots, s$ we obtain

$$\begin{aligned}
\sum_{J_\sigma} |w_j| e^{-2\pi\langle j, \operatorname{Im} \xi \rangle} &\leq \sum_{J_\sigma} |w_j| e^{\pi|j|\rho} \\
&\leq \left(\sum_{J_\sigma} |w_j|^2 e^{2\pi|j|r} \right)^{1/2} \left(\sum_{J_\sigma} e^{-2\pi|j|(r-\rho)} \right)^{1/2} \\
&\leq \left(\sum_{\mathbb{Z}^n} |w_j|^2 e^{4\pi\langle j, r e_\sigma \rangle} \right)^{1/2} \left(\sum_{\mathbb{Z}^n} e^{-2\pi|j|(r-\rho)} \right)^{1/2} \\
&\leq \frac{c_1^{1/2}}{(r-\rho)^{n/2}} \|w\|_r.
\end{aligned}$$

Hence (ii) holds with $c := c_1^{1/2}(1+s)$, where both c_1 and s depend on n only.

Assertion (iii) follows from Cauchy's integral formula

$$|w_\xi(\xi)| = \frac{1}{2\pi i} \int_\Gamma \lambda^{-2} w \left(\xi + \lambda \frac{\overline{w_\xi(\xi)}}{|w_\xi(\xi)|} \right) d\lambda,$$

with $\Gamma := \{\lambda \in \mathbb{C} \,|\, |\lambda| = r - \rho\}$. This proves the lemma. $\square$

It might be interesting to compare assertion (ii) of the previous lemma with similar results in the literature working with $\max|\operatorname{Im} \xi_\nu|$ and $\sum|j_\nu|$ instead of the Euclidean norms. In that case the L^2-norm on the larger strip $|\operatorname{Im} \xi_\nu| \leq r$ is equivalent to the square root of $\sum_j |w_j|^2 e^{4\pi r \sum|j_\nu|}$, but an analogous statement does not seem to hold in the case of the Euclidean norm. Therefore the proof of Lemma 1 (ii) is a bit more delicate than might be expected.

We are now in a position to prove the desired inequality for the solutions of $Df = g$. This requires estimating a series with *small divisors*. The observation that the inequality (2.1) below holds with the L^2-norm $\|g\|_r$ on the right hand side, instead of the sup-norm $|g|_r$, was pointed out to the author by Jürgen Moser. Another minor difference to analogous results in the literature lies in the aforementioned use of the Euclidean norm for $\operatorname{Im} \xi \in \mathbb{R}^n$.

Lemma 2 (Moser, Rüssmann). *Let $n \geq 2$, $\tau > n-1$, and $c_0 > 0$ be given. Then there exists a constant $c > 0$ such that the following holds for every vector $\omega \in \mathbb{R}^n$ that satisfies (1.3). If $g \in W_r^0$ with $0 < r \leq 1$ then the equation*

$$Df = g$$

has a unique solution $f \in W^0_\rho$, $\rho < r$, and this solution satisfies the inequality

$$|f|_\rho \leq \frac{c}{(r-\rho)^\tau}\|g\|_r. \tag{2.1}$$

Proof. Representing the functions $f, g \in W^0_r$ by their Fourier series we obtain that the equation $Df = g$ is equivalent to

$$f_j = \frac{g_j}{2\pi i\langle j,\omega\rangle}, \qquad 0 \neq j \in \mathbb{Z}^n.$$

This proves uniqueness. To establish existence and the inequality (2.1), we first single out the subset $J_0 \subset \mathbb{Z}^n$ of all those vectors $j \neq 0$ that satisfy $|\langle j,\omega\rangle|^{-1} \leq 2c_0$ and define

$$f^0(\xi) := \sum_{j\in J_0} f_j e^{2\pi i\langle j,\xi\rangle}.$$

Then the following inequality holds for $|\mathrm{Im}\,\xi| \leq \rho < r$:

$$\begin{aligned}
|f^0(\xi)| &\leq \frac{1}{2\pi}\sum_{j\in J_0}\frac{|g_j|}{|\langle j,\omega\rangle|}e^{-2\pi\langle j,\mathrm{Im}\,\xi\rangle} \\
&\leq \frac{c_0}{\pi}\sum_{j\in J_0}|g_j|e^{-2\pi\langle j,\mathrm{Im}\,\xi\rangle} \\
&\leq \frac{c_0c_1}{(r-\rho)^\tau}\|g\|_r.
\end{aligned}$$

Here we have used the inequality $\tau > n-1 \geq n/2$ and chosen $c_1 = c_1(n) > 0$ so that $c := c_1\pi$ is the constant of Lemma 1 (ii).

The more delicate part of the estimate concerns the integer vectors in $\mathbb{Z}^n \setminus J_0$ and is based on the observation that only a few of the divisors $\langle j,\omega\rangle$ are actually small. This fact was used e.g. by Siegel [26], Arnold [3], Moser [15], Rüssmann [25], and Pöschel [22].

Fix a number $K \geq 1$ and, for $\nu = 1, 2, 3, \ldots$, denote by $J(\nu, K)$ the set of all integer vectors $j \in \mathbb{Z}^n$ that satisfy the inequality

$$\frac{1}{2^{\nu+1}c_0} \leq |\langle j,\omega\rangle| < \frac{1}{2^\nu c_0}, \qquad 0 < |j| \leq K.$$

In order to estimate the number of points in $J(\nu, K)$ we assume without loss of generality that $|\omega_\nu| \leq |\omega_n|$ for all ν and define $\bar{j} := (j_1, \ldots, j_{n-1}) \in \mathbb{Z}^{n-1}$ for $j \in \mathbb{Z}^n$. Fixing $\bar{j} \neq 0$ and choosing j_n so as to minimize $|\langle j,\omega\rangle|$ we then obtain $|j_n| \leq |j_1| + \cdots + |j_{n-1}| + 1 \leq 2\sqrt{n-1}|\bar{j}|$ which implies $|j| \leq 2\sqrt{n}|\bar{j}|$. Therefore it follows from (1.3) that

$$|\langle j,\omega\rangle| \geq \frac{1}{c_0\,(2\sqrt{n}|\bar{j}|)^\tau}$$

for every $j \in \mathbb{Z}^n$ with $\bar{j} \neq 0$. It follows also from (1.3) that $|\omega_n| \geq 1/c_0$ and therefore any two integer vectors $j, j' \in J(\nu, K)$ with $j \neq j'$ must satisfy $\bar{j} \neq \bar{j}'$. Hence, by what we have just observed,

$$|\bar{j} - \bar{j}'|^{-\tau} \leq c_0 (2\sqrt{n})^\tau |\langle j - j', \omega\rangle| \leq 2(2\sqrt{n})^\tau 2^{-\nu} \leq (4\sqrt{n})^\tau 2^{-\nu}.$$

Thus the distance

$$|\bar{j} - \bar{j}'| \geq \frac{2^{\nu/\tau}}{4\sqrt{n}}$$

gets very large for large ν. This shows that the number of points in $J(\nu, K)$ can be estimated by

$$\#J(\nu, K) \leq c_2 K^{n-1} 2^{-\nu(n-1)/\tau}$$

for some constant $c_2 = c_2(n) > 0$. Moreover, we obtain from (1.3) that $J(\nu, K) = \emptyset$ for $2^{\nu/\tau} \geq K$. Denote by $J(K)$ the set of all integer vectors $j \in \mathbb{Z}^n$ that satisfy

$$\frac{1}{|\langle j, \omega\rangle|} > 2c_0, \qquad 0 < |j| \leq K.$$

Then $J(K)$ is the union of the sets $J(\nu, K)$ for $\nu = 1, 2, 3, \ldots$ and we conclude that

$$\sum_{j \in J(K)} \frac{1}{|\langle j, \omega\rangle|} \leq 2c_0 c_2 K^{n-1} \sum_{2^{\nu/\tau} \leq K} 2^{\nu(\tau+1-n)/\tau} \leq c_0 c_3 K^\tau$$

for some constant $c_3 = c_3(n, \tau) > 0$. It is interesting to note that the size of this sum is of the same order as the size of the largest term in it.

Using Lemma 1 (i) we can now continue estimating $|f|_\rho$:

$$\begin{aligned}
|f - f^0|_\rho &\leq \frac{1}{2\pi} \sum_{j \neq J_0} \frac{|g_j|}{|\langle j, \omega\rangle|} e^{2\pi |j| \rho} \\
&\leq \frac{\|g\|_r}{2\pi} \sum_{j \neq J_0} \frac{1}{|\langle j, \omega\rangle|} e^{-2\pi |j|(r-\rho)} \\
&= \frac{\|g\|_r}{2\pi} \sum_{k=1}^\infty \sum_{\substack{j \neq J_0 \\ |j|^2 = k}} \frac{1}{|\langle j, \omega\rangle|} e^{-2\pi\sqrt{k}(r-\rho)} \\
&= \frac{\|g\|_r}{2\pi} \sum_{k=1}^\infty \sum_{j \in J(\sqrt{k})} \frac{1}{|\langle j, \omega\rangle|} \left(e^{-2\pi\sqrt{k}(r-\rho)} - e^{-2\pi\sqrt{k+1}(r-\rho)} \right) \\
&\leq \frac{c_0 c_3 \|g\|_r}{2} \sum_{k=1}^\infty k^{\tau/2} \frac{r-\rho}{\sqrt{k}} e^{-2\pi\sqrt{k}(r-\rho)} \\
&\leq \frac{c_0 c_3 \|g\|_r}{(r-\rho)^\tau} \sup_{0 < \lambda \leq 1} \sum_{k=1}^\infty (\lambda\sqrt{k})^\tau \frac{\lambda}{2\sqrt{k}} e^{-2\pi\lambda\sqrt{k}}.
\end{aligned}$$

Here the penultimate inequality uses the fact that, for $\lambda > 0$, we have

$$e^{-2\lambda\sqrt{k}} - e^{-2\lambda\sqrt{k+1}} \leq 2\lambda \left(\sqrt{k+1} - \sqrt{k} \right) e^{-2\lambda\sqrt{k}} \leq \frac{\lambda}{\sqrt{k}} e^{-2\lambda\sqrt{k}}.$$

This proves the lemma. $\square$

Theorem 1 (Kolmogorov, Arnold, Moser). *Let $n \geq 2$, $\tau > n-1$, $c_0 > 0$, $0 < \theta < 1$, and $M \geq 1$ be given. Then there are positive constants δ^* and c such that $c\delta^* \leq 1/2$ and the following holds for every $0 < r \leq 1$ and every $\omega \in \mathbb{R}^n$ that satisfies (1.3).*

Suppose $H(x,y)$ is a real analytic Hamiltonian function defined in the strip $|\mathrm{Im}\, x| \leq r$, $|y| \leq r$, which is of period 1 in the variables $x_1, \ldots, x_n$ and satisfies

$$\begin{aligned} \left| H(x,0) - \int_{\mathbb{T}^n} H(\xi,0)\, d\xi \right| &\leq \delta r^{2\tau+2}, \\ |H_y(x,0) - \omega| &\leq \delta r^{\tau+1}, \\ |H_{yy}(x,y) - Q(x,y)| &\leq \frac{c\delta}{2M}, \end{aligned} \tag{2.2}$$

for $|\mathrm{Im}\, x| \leq r$ and $|y| \leq r$, where $0 < \delta \leq \delta^*$, and $Q(x,y) \in \mathbb{C}^{n\times n}$ is a symmetric (not necessarily analytic) matrix valued function in the strip $|\mathrm{Im}\, x| \leq r$, $|y| \leq r$ and satisfies in this domain

$$|Q(z)| \leq M, \qquad \left| \left(\int_{\mathbb{T}^n} Q(x,0)\, dx \right)^{-1} \right| \leq M. \tag{2.3}$$

Then there exists a real analytic symplectic transformation $z = \phi(\zeta)$ of the form (1.7) mapping the strip $|\mathrm{Im}\, \xi| \leq \theta r$, $|\eta| \leq \theta r$ into $|\mathrm{Im}\, x| \leq r$, $|y| \leq r$, such that $u(\xi) - \xi$ and $v(\xi)$ are of period 1 in all variables and the Hamiltonian function $K := H \circ \phi$ satisfies (1.8). Moreover, ϕ and K satisfy the estimates

$$\begin{aligned} &|\phi(\zeta) - \zeta| \leq c\delta(1-\theta)r, \quad |\phi_\zeta(\zeta) - \mathbb{1}| \leq c\delta, \\ &|K_{\eta\eta}(\zeta) - Q(\zeta)| \leq \frac{c\delta}{M}, \\ &\left| v \circ u^{-1}(x) \right| \leq c\delta r^{\tau+1}, \end{aligned} \tag{2.4}$$

for $|\mathrm{Im}\, \xi| \leq \theta r$, $|\eta| \leq \theta r$, and $|\mathrm{Im}\, x| \leq \theta r$.

Proof. We will construct inductively a sequence of real analytic Hamiltonian functions $H^\nu(x,y)$ in the strips $|\mathrm{Im}\, x| \leq r_\nu$, $|y| \leq r_\nu$, where

$$r_\nu := \left(\frac{1+\theta}{2} + \frac{1-\theta}{2^{\nu+1}} \right) r, \qquad r_0 = r,$$

such that $H^0 := H$ and $H^{\nu+1} := H^\nu \circ \psi^\nu$. The symplectic transformation $z = \psi^\nu(\zeta)$ will be real analytic, will map the strip $|\mathrm{Im}\, \xi| \leq \theta r_{\nu+1}$, $|\eta| \leq \theta r_{\nu+1}$ into $|\mathrm{Im}\, x| \leq r_\nu$, $|y| \leq r_\nu$, and it will be represented in terms of its generating function

$$S^\nu(x,\eta) = U^\nu(x) + \langle V^\nu(x), \eta \rangle.$$

Following Kolmogorov [10] we will choose the real analytic functions

$$U^\nu(x) = \langle \alpha, x \rangle + a(x), \qquad V^\nu(x) = x + b(x)$$

in the strip $|\mathrm{Im}\, x| \le r_\nu$ such that $a(x)$ and $b(x)$ are of period 1 in all variables, have mean value zero over the n-torus, and satisfy the following equations for $|\mathrm{Im}\, x| \le r_\nu$:

$$\begin{aligned} Da(x) &= \int_{\mathbb{T}^n} H^\nu(\xi, 0)\, d\xi - H^\nu(x, 0), \\ \int_{\mathbb{T}^n} &\Big(H^\nu_y(\xi, 0) + H^\nu_{yy}(\xi, 0)(\alpha + a_x(\xi)) \Big)\, d\xi = \omega, \\ Db(x) &= \omega - H^\nu_y(\xi, 0) - H^\nu_{yy}(\xi, 0)(\alpha + a_x(\xi)). \end{aligned} \tag{2.5}$$

Note that the second equation in (2.5) determines α and that it is necessary in order for the right hand side of the third equation to have mean value zero so that Lemma 2 can be applied. Moreover, observe that (2.5) can be obtained from (1.12) and (1.13) by linearizing these equations around $V(x) = x$ and $U(x) = 0$ if we replace Da by $\langle a_x, H^\nu_y(x, 0)\rangle$ and Db by $b_x H^\nu_y(x, 0)$.

Let us now define the *error* at the νth step of the iteration to be the smallest number $\varepsilon_\nu > 0$ that satisfies the inequalities

$$\left| H^\nu(x, 0) - \int_{\mathbb{T}^n} H^\nu(\xi, 0)\, d\xi \right| \le \varepsilon_\nu, \qquad \left| H^\nu_y(x, 0) - \omega \right| (r_\nu - r_{\nu+1})^{\tau+1} \le \varepsilon_\nu$$

for $|\mathrm{Im}\, x| \le r_\nu$. We will then show that ε_ν converges to zero according to the quadratic estimate

$$\varepsilon_{\nu+1} \le \frac{c_3}{(r_\nu - r_{\nu+1})^{2\tau+2}} \varepsilon_\nu^2. \tag{2.6}$$

Here $c_3 > 0$ is a suitable constant independent of r. It is important to notice that the geometrically growing factor in front of ε_ν^2 in (2.6) is dominated by the quadratic convergence of ε_ν. In fact, one checks easily that (2.6) implies

$$\varepsilon_\nu \le \delta_\nu r^{2\tau+2}$$

where the sequence δ_ν is defined by the recursive law

$$\begin{aligned} \delta_{\nu+1} &:= c 2^{\nu(2\tau+3)} \delta_\nu^2, \qquad & \delta_0 &:= \delta \le \delta^*, \\ \gamma_\nu &:= c 2^{(\nu+1)(2\tau+3)} \delta_\nu, \qquad & \gamma_0 &:= 2^{2\tau+3} c\delta, \end{aligned} \tag{2.7}$$

and the constant $c > 0$ is related to c_3 as in equation (2.10) below.[2] The sequence γ_ν will then satisfy

$$\gamma_{\nu+1} = \gamma_\nu^2 \tag{2.8}$$

and therefore γ_ν converges to zero whenever $\gamma_0 < 1$. For the sequence H^ν this will lead to the estimates

$$\begin{aligned} &\left| H^\nu(x, 0) - \int_{\mathbb{T}^n} H^\nu(\xi, 0)\, d\xi \right| \le \delta_\nu r^{2\tau+2}, \\ &\left| H^\nu_y(x, 0) - \omega \right| (r_\nu - r_{\nu+1})^{\tau+1} \le \delta_\nu r^{2\tau+2}, \\ &\left| H^\nu_{yy}(x, y) - Q^\nu(x, y) \right| \le 2^{-\nu} \frac{c\delta}{2M} \end{aligned} \tag{2.9}$$

[2] A small flaw in the formula (2.7) is the factor $2\tau + 3$ where one would like to have $2\tau + 2$. The number $2\tau + 3$ is only needed in Step 3 for estimating the y-component of $\psi^\nu - \mathrm{id}$.

for $|\mathrm{Im}\, x| \leq r_\nu$ and $|y| \leq r_\nu$, where

$$Q^0 := Q, \qquad Q^\nu := H_{yy}^{\nu-1}$$

for $\nu \geq 1$. In this discussion the constants are chosen explicitly as follows. Let $c_1 = c_1(n, \tau, c_0)$ be the constant of Lemma 2 and define c_2, c_3, and c by

$$c_2 := 12M^3 \left(1 + c_1 8^{\tau+1}\right)^2, \quad c_3 := 4Mc_2^2 + c_2, \quad c := \left(\frac{4}{1-\theta}\right)^{2\tau+3} c_3. \tag{2.10}$$

Next we choose $\delta^* > 0$ so small that

$$c\delta^* \leq 2^{-2\tau-4}. \tag{2.11}$$

This implies $\gamma_0 \leq 1/2$ and therefore $\gamma_{\nu+1} \leq \gamma_\nu/2$ for $\nu \geq 1$. Finally, define

$$M_\nu := \frac{M_{\nu-1}}{1 - 2^{-\nu} c\delta^*}, \qquad M_{-1} := M, \tag{2.12}$$

and note that

$$M_\nu \leq M_{\nu-1} e^{2^{1-\nu} c\delta^*} \leq M e^{4c\delta^*} \leq 2M \tag{2.13}$$

for $\nu \geq 0$. Here the last inequality follows from (2.11).

With the constants in place, we are now ready to prove by induction that H^ν satisfies (2.9). First observe that, by assumption, the inequalities (2.9) are satisfied for $\nu = 0$ provided that $\delta \leq \delta^*$. Fix an integer $\nu \geq 0$ and assume, by induction, that the real analytic functions $H^\mu(x, y)$ on the domains $|\mathrm{Im}\, x| \leq r_\mu$, $|y| \leq r_\mu$, have been constructed for $\mu = 1, \ldots, \nu$ such that (2.9) is satisfied with ν replaced by μ. Then we obtain from (2.3), (2.9), (2.12), and (2.13) by induction that

$$|H_{yy}^\nu(x, y)| \leq M_\nu, \qquad \left| \left(\int_{\mathbb{T}^n} H_{yy}^\nu(x, 0)\, dx \right)^{-1} \right| \leq M_\nu \tag{2.14}$$

for $|\mathrm{Im}\, x| \leq r_\nu$ and $|y| \leq r_\nu$. We shall now construct, in four steps, a real analytic Hamiltonian function $H^{\nu+1} = H^\nu \circ \psi^\nu$ and show that it satisfies (2.9) with ν replaced by $\nu + 1$.

Step 1. *If $H^\nu(x, y)$ satisfies (2.14) then*

$$\begin{aligned} &\left|H^\nu(x, y) - H^\nu(x, 0) - \langle H^\nu(x, 0), y\rangle\right| \leq M\, |y|^2, \\ &\left|H_y^\nu(x, y) - H_y^\nu(x, 0)\right| \leq 2M\, |y|, \\ &\left|H_y^\nu(x, y) - H_y^\nu(x, 0) - H_{yy}^\nu(x, 0) y\right| \leq \frac{2M\, |y|^2}{r_\nu - |y|} \end{aligned}$$

for $|\mathrm{Im}\, x| \leq r_\nu$ and $|y| < r_\nu$.

These estimates follow from the fact that $M_\nu \le 2M$ and from the identities

$$
\begin{aligned}
H^\nu(x,y) - H^\nu(x,0) - \langle H^\nu_y(x,0), y\rangle &= \int_0^1 \int_0^t \langle y, H^\nu_{yy}(x,sy)y\rangle \, dsdt, \\
H^\nu(x,y) - H^\nu(x,0) &= \int_0^1 H^\nu_{yy}(x,ty)y\, dt, \\
H^\nu_y(x,y) - H^\nu_y(x,0) - H^\nu_{yy}(x,0)y &= \int_0^1 \left(H^\nu_{yy}(x,ty)y - H^\nu_{yy}(x,0)y \right) \, dsdt \\
&= \int_0^1 \frac{1}{2\pi i} \int_\Gamma \frac{1}{\lambda(\lambda-1)} H^\nu_{yy}(x,\lambda ty)y \, d\lambda dt,
\end{aligned}
$$

with $\Gamma := \{\lambda \in \mathbb{C} \mid |\lambda| = r_\nu/|y| > 1\}$. This proves Step 1.

Now it follows from Lemma 2 that there exist unique solutions $a(x)$, α, $b(x)$ of equation (2.5) in the strip $|\mathrm{Im}\, x| < r_\nu$ such that $a(x)$ and $b(x)$ are of period 1 in all variables and have mean value zero over the torus.

Step 2. *The solutions $a(x)$, α, $b(x)$ of (2.5) satisfy the estimates*

$$
|a(x)| \le \frac{c_2 \varepsilon_\nu}{(r_\nu - r_{\nu+1})^\tau}, \qquad |\alpha + a_x(x)| \le \frac{c_2 \varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}}, \tag{2.15}
$$

$$
|b(x)| \le \frac{c_2 \varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+1}}, \qquad |b_x(x)| \le \frac{c_2 \varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+2}}, \tag{2.16}
$$

for $|\mathrm{Im}\, x| \le (r_\nu + r_{\nu+1})/2$.

Define $\rho_j := ((8-j)r_\nu + jr_{\nu+1})/8$ for $0 \le j \le 4$ so that $\rho_0 = r_\nu$ and $\rho_4 = (r_\nu + r_{\nu+1})/2$. Then it follows from Lemma 2 that

$$
|a|_{\rho_1} \le c_1 \left(\frac{8}{r_\nu - r_{\nu+1}} \right)^\tau \left| H^\nu(\cdot,0) - \int_{\mathbb{T}^n} H^\nu(\xi,0)\, d\xi \right|_{\rho_0} \le \frac{c_1 8^\tau \varepsilon_\nu}{(r_\nu - r_{\nu+1})^\tau}
$$

and hence, by Lemma 1 (iii),

$$
|a_x|_{\rho_2} \le \frac{8}{r_\nu - r_{\nu+1}} |a|_{\rho_1} \le \frac{c_1 8^{\tau+1} \varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}}.
$$

This estimate, together with (2.5) and (2.14), implies

$$
\begin{aligned}
|\alpha| &\le M_\nu \left(\left|\omega - H^\nu_y(\cdot,0)\right|_{\rho_2} + \left|H^\nu_{yy}(\cdot,0)a_x\right|_{\rho_2} \right) \\
&\le M_\nu \left(1 + M_\nu c_1 8^{\tau+1}\right) \frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}} \\
&\le 4M^2 \left(1 + c_1 8^{\tau+1}\right) \frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}}.
\end{aligned}
$$

Here the second inequality follows from the definition of ε_ν and the last from equation (2.13). Thus we have proved that

$$
|\alpha + a_x|_{\rho_2} \le 5M^2 \left(1 + c_1 8^{\tau+1}\right) \frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}}.
$$

Combining this with (2.10) gives (2.15). Furthermore, we obtain from (2.5) and Lemma 2 that

$$\begin{aligned} |b|_{\rho_3} &\leq c_1 \left(\frac{8}{r_\nu - r_{\nu+1}}\right)^\tau \left|\omega - H_y^\nu(\cdot,0) - H_{yy}^\nu(\cdot,0)(\alpha + a_x)\right|_{\rho_2} \\ &\leq c_1 \left(\frac{8}{r_\nu - r_{\nu+1}}\right)^\tau \left(1 + 10M^3\left(1 + c_1 8^{\tau+1}\right)\right)\frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}} \\ &\leq \frac{12M^3}{8}\left(1 + c_1 8^{\tau+1}\right)^2 \frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+1}}. \end{aligned}$$

Hence, by Lemma 1 (iii), we have

$$|b_x|_{\rho_4} \leq \frac{8}{r_\nu - r_{\nu+1}} |b|_{\rho_3} \leq 12M^3\left(1 + c_1 8^{\tau+1}\right)^2 \frac{\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+2}}.$$

Combining this with (2.10) gives (2.16). This proves Step 2.

Having constructed the functions $U(x) := U^\nu(x) := \langle\alpha, x\rangle + a(x)$ and $V(x) := V^\nu(x) := x + b(x)$ we can define the symplectic transformation $z = \psi^\nu(\zeta)$ by (1.7) and (1.9). Thus

$$z = \psi^\nu(x) \qquad \Longleftrightarrow \qquad \xi = x + b(x), \quad y = \alpha + a_x(x) + \eta + b_x(x)^T\eta.$$

That this map is well defined will be established in the proof of Step 3.

Step 3. *The transformation $z = \psi^\nu(\zeta)$ maps the strip $|\mathrm{Im}\,\xi| \leq r_{\nu+1}$, $|\eta| \leq r_{\nu+1}$ into $|\mathrm{Im}\,x| \leq (r_\nu + r_{\nu+1})/2$, $|y| \leq (r_\nu + r_{\nu+1})/2$ and satisfies the estimates*

$$|\psi^\nu(\zeta) - \zeta| \leq 2^{-\nu} c\delta \frac{r_\nu - r_{\nu+1}}{4}, \tag{2.17}$$

and

$$\left|\psi_\zeta^\nu(\zeta) - \mathbb{1}\right| \leq 2^{-\nu} c\delta, \tag{2.18}$$

for $|\mathrm{Im}\,\xi| \leq r_{\nu+1}$ and $|\eta| \leq r_{\nu+1}$.

First note that the induction hypothesis (2.9) implies $\varepsilon_\nu \leq \delta_\nu r^{2\tau+2}$ and hence

$$\begin{aligned} \frac{c_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+2}} &\leq c_2\left(\frac{2^{\nu+2}}{1-\theta}\right)^{2\tau+2} \delta_\nu \\ &= \frac{c_3}{4Mc_2 + 1}\left(\frac{4}{1-\theta}\right)^{2\tau+2} 2^{\nu(2\tau+2)}\delta_\nu \\ &= \frac{\gamma_\nu}{2^{2\tau+3}(4Mc_2+1)}\frac{1-\theta}{2^{\nu+2}} \\ &\leq \frac{2^{-\nu}c\delta}{16M^2}\frac{1-\theta}{2^{\nu+2}}. \end{aligned} \tag{2.19}$$

Here the second equality follows from the definition of c_3, the third equality from the definition of c and γ_ν, and the last inequality follows from the fact that $\gamma_\nu \le 2^{-\nu}\gamma_0 = 2^{-\nu}2^{2\tau+3}c\delta$ and $c_2 \ge 4M$. Combining (2.19) with Step 2 we obtain the following estimates for $|\mathrm{Im}\, x| \le (r_\nu + r_{\nu+1})/2$ and $\xi := V^\nu(x)$:

$$\begin{aligned} |x-\xi| = |b(x)| &\le \frac{c_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+1}} \le \frac{2^{-\nu}c\delta}{16M^2}(r_\nu - r_{\nu+1}), \\ |b_x(x)| &\le \frac{c_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+2}} \le \frac{2^{-\nu}c\delta}{16M^2}\frac{1-\theta}{2^{\nu+2}}. \end{aligned} \tag{2.20}$$

This implies that $V^\nu = \mathrm{id} + b$ has an inverse $u := (V^\nu)^{-1}$ which maps the strip $|\mathrm{Im}\,\xi| \le (r_\nu + 3r_{\nu+1})/4$ into $|\mathrm{Im}\, x| \le (r_\nu + r_{\nu+1})/2$. To see this fix a vector $\xi \in \mathbb{C}^n$ with $|\mathrm{Im}\,\xi| \le (r_\nu + 3r_{\nu+1})/4$ and apply the contraction mapping principle to the map $x \mapsto \xi - b(x)$ on the domain $|\mathrm{Im}\, x| \le (r_\nu + r_{\nu+1})/2$.

Now let $\xi, \eta \in \mathbb{C}^n$ such that $|\mathrm{Im}\,\xi| \le (r_\nu + 3r_{\nu+1})/4$ and $|\eta| \le (r_\nu + 3r_{\nu+1})/4$, let $x \in \mathbb{C}^n$ be the unique vector that satisfies $|\mathrm{Im}\, x| \le (r_\nu + r_{\nu+1})/2$ and $x + b(x) = \xi$, and define $y \in \mathbb{C}^n$ by

$$y := \alpha + a_x(x) + \eta + b_x(x)^T\eta.$$

Then $z = \psi^\nu(\zeta)$ and it follows again from Step 2 that

$$\begin{aligned} |y-\eta| &\le |\alpha + a_x| + \left|b_x^T\eta\right| \\ &\le \frac{c_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{\tau+1}} + \frac{c_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+2}}\frac{r_\nu + 3r_{\nu+1}}{4} \\ &\le \frac{c_2\varepsilon_\nu r}{(r_\nu - r_{\nu+1})^{2\tau+2}} \\ &\le \frac{2^{-\nu}c\delta}{16M^2}(r_\nu - r_{\nu+1}) \end{aligned} \tag{2.21}$$

Here the third inequality uses the fact that $(r_\nu - r_{\nu+1})^\tau \le 1/4$. The last inequality follows from (2.19) and the fact that $r_\nu - r_{\nu+1} = 2^{-\nu-2}(1-\theta)r$. It follows from (2.21) that $|y| \le (r_\nu + r_{\nu+1})/2$. The inequalities (2.20) and (2.21) together show that ψ^ν satisfies (2.17) in the domain $|\mathrm{Im}\,\xi| \le (r_\nu + 3r_{\nu+1})/4$, $|\eta| \le (r_\nu + 3r_{\nu+1})/4$. The estimate (2.18) follows from (2.17) and Cauchy's integral formula (Lemma 1 (iii)). This proves Step 3.

In view of Step 3 we can define

$$H^{\nu+1} := H^\nu \circ \psi^\nu.$$

The next step establishes (2.6) along with the required estimates for $H^{\nu+1}$.

Step 4. *The inequalities in (2.9) are satisfied with ν replaced by $\nu+1$.*

We will first establish (2.6). For this define the real number h by

$$h := \int_{\mathbb{T}^n} H^\nu(\xi, 0)\, d\xi + \langle \omega, \alpha \rangle.$$

and denote $z := (x, \alpha + a_x) := \psi^\nu(\xi, 0)$, where $|\mathrm{Im}\,\xi| \le r_{\nu+1}$. Then it follows from Step 3 that $|\mathrm{Im}\,x| \le (r_\nu + r_{\nu+1})/2$ and $|y| = |\alpha + a_x| \le (r_\nu + r_{\nu+1})/2$. Moreover, it follows from (2.5) that

$$\begin{aligned} H^{\nu+1}(\xi,0) - h &= H^\nu(x, \alpha + a_x) - h \\ &= H^\nu(x, \alpha + a_x) - H^\nu(x, 0) - Da - \langle \omega, \alpha \rangle \\ &= H^\nu(x, \alpha + a_x) - H^\nu(x, 0) - \langle H_y^\nu(x,0), \alpha + a_x \rangle \\ &\quad + \langle H_y^\nu(x,0) - \omega, \alpha + a_x \rangle. \end{aligned}$$

By Step 1 and Step 2, this implies

$$\begin{aligned} \left| H^{\nu+1}(\xi,0) - h \right| &\le M \left| \alpha + a_x \right|^2 + \left| H_y^\nu(x,0) - \omega \right| \left| \alpha + a_x \right| \\ &\le \frac{Mc_2^2 + c_2}{(r_\nu - r_{\nu+1})^{2\tau+2}} \varepsilon_\nu^2 \\ &\le \frac{c_3/2}{(r_\nu - r_{\nu+1})^{2\tau+2}} \varepsilon_\nu^2, \end{aligned}$$

and hence

$$\left| H^{\nu+1}(\xi,0) - \int_{\mathbb{T}^2} H^{\nu+1}(\xi,0)\, d\xi \right| \le \frac{c_3}{(r_\nu - r_{\nu+1})^{2\tau+2}} \varepsilon_\nu^2. \tag{2.22}$$

Secondly, it follows from (2.5) that

$$\begin{aligned} H_y^{\nu+1}(\xi,0) - \omega &= (\mathbb{1} + b_x)\, H_y^\nu(x, \alpha + a_x) - \omega \\ &= H_y^\nu(x, \alpha + a_x) - H_y^\nu(x,0) - H_{yy}^\nu(x,0)(\alpha + a_x) \\ &\quad + b_x \left(H_y^\nu(x, \alpha + a_x) - H_y^\nu(x,0) \right) \\ &\quad + b_x \left(H_y^\nu(x,0) - \omega \right). \end{aligned}$$

By Step 1 and Step 2, this implies

$$\begin{aligned} \left| H_y^{\nu+1}(\xi,0) - \omega \right| &\le \frac{2M \left| \alpha + a_x \right|^2}{r_\nu - |\alpha + a_x|} + 2M \left| b_x \right| \left| \alpha + a_x \right| + \left| b_x \right| \left| H_y^\nu(x,0) - \omega \right| \\ &\le \frac{4Mc_2^2 + c_2}{(r_\nu - r_{\nu+1})^{3\tau+3}} \varepsilon_\nu^2 \\ &= \frac{1}{(r_{\nu+1} - r_{\nu+2})^{\tau+1}} \frac{c_3}{(r_\nu - r_{\nu+1})^{2\tau+2}} \varepsilon_\nu^2. \end{aligned}$$

In the second inequality we have used the fact that $|\alpha + a_x| \le (r_\nu - r_{\nu+1})/2$, by (2.21), and hence $(r_\nu - r_{\nu+1})^{\tau+1} \le (r_\nu - r_{\nu+1})/2 \le r_\nu - |\alpha + a_x|$. The estimate (2.6) now follows by combining the last inequality with (2.22).

Combining (2.6) with (2.7) and the induction hypothesis $\varepsilon_\nu \le \delta_\nu r^{2\tau+2}$ we obtain

$$\varepsilon_{\nu+1} \le \frac{c_3 \delta_\nu^2 r^{4\tau+4}}{(r_\nu - r_{\nu+1})^{2\tau+2}} = c_3 \left(\frac{2^{\nu+2}}{1-\theta} \right)^{2\tau+2} \delta_\nu^2 r^{2\tau+2} \le \delta_{\nu+1} r^{2\tau+2}.$$

This implies the first two inequalities in (2.9) with ν replaced by $\nu+1$. To prove the last inequality, suppose that $|\mathrm{Im}\,\xi| \le r_{\nu+1}$, $|\eta| \le r_{\nu+1}$ and denote $z := \psi^\nu(\zeta)$. Then Step 3 shows that $|\mathrm{Im}\,x| \le (r_\nu + r_{\nu+1})/2$, $|y| \le (r_\nu + r_{\nu+1})/2$ and, moreover, it follows from the definition of $H^{\nu+1}$ that

$$H^{\nu+1}_{yy}(\xi,\eta) = (\mathbb{1} + b_x) H^\nu_{yy}(x,y)\left(\mathbb{1} + b_x^T\right).$$

Hence, by Step 2, we obtain

$$\begin{aligned}\left|H^{\nu+1}_{yy}(\zeta) - H^\nu_{yy}(z)\right| &\le 2\,|b_x|\left|H^\nu_{yy}(z)\right| + |b_x|^2\left|H^\nu_{yy}(z)\right| \\ &\le 3\,|b_x|\left|H^\nu_{yy}(z)\right| \\ &\le \frac{6Mc_2\varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+1}} \le \frac{2^{-\nu}c\delta}{4M}.\end{aligned}$$

The last inequality follows from (2.19). Moreover,

$$H^\nu_{yy}(z) - H^\nu_{yy}(\zeta) = \frac{1}{2\pi i}\int_\Gamma \frac{1}{\lambda(\lambda-1)} H^\nu_{yy}(\zeta + \lambda(z-\zeta))\,d\lambda,$$

where

$$\Gamma := \left\{\lambda \in \mathbb{C} \mid |\lambda| = \min\left\{\frac{r_\nu - |\mathrm{Im}\,\xi|}{|x-\xi|}, \frac{r_\nu - |\eta|}{|y-\eta|}\right\} > 1\right\}.$$

This implies

$$\begin{aligned}\left|H^\nu_{yy}(z) - H^\nu_{yy}(\zeta)\right| &\le 2M\max\left\{\frac{|x-\xi|}{r_\nu - |\mathrm{Im}\,\xi| - |x-\xi|}, \frac{|y-\eta|}{r_\nu - |\eta| - |y-\eta|}\right\} \\ &\le 2M\frac{2^{-\nu}c\delta}{16M^2}\frac{r_\nu - r_{\nu+1}}{(r_\nu - r_{\nu+1})/2} = \frac{2^{-\nu}c\delta}{4M}.\end{aligned}$$

Here the second inequality follows from (2.20) and (2.21) and the fact that $|x-\xi| \le (r_\nu - r_{\nu+1})/2$ and $|y-\eta| \le (r_\nu - r_{\nu+1})/2$. This proves Step 4.

Step 4 completes the induction step and it remains to establish the uniform convergence of the sequence

$$\phi^\nu := \psi^0 \circ \psi^1 \circ \cdots \circ \psi^\nu$$

in the domain $|\mathrm{Im}\,\xi| \le \theta r$, $|\eta| \le \theta r$ along with the estimates in (2.4). First it follows from Step 3 that, if $|\mathrm{Im}\,\xi| \le r_\nu$ and $|\eta| \le r_\nu$ and $z := \psi^{\mu+1}\circ\cdots\circ\psi^{\nu-1}(\zeta)$ then $|\mathrm{Im}\,x| \le r_{\mu+1}$ and $|y| \le r_{\mu+1}$ and therefore, by (2.18),

$$\left|\psi^\mu_\zeta(\psi^{\mu+1}\circ\cdots\circ\psi^{\nu-1}(\zeta))\right| \le 1 + 2^{-\mu}c\delta.$$

This implies

$$\left|\phi^{\nu-1}_\zeta(\zeta)\right| \le (1+c\delta)\cdots(1+2^{1-\nu}c\delta) \le e^{2c\delta} \le 2$$

for $|\operatorname{Im}\xi| \le r_\nu$ and $|\eta| \le r_\nu$ and hence

$$\left|\phi^\nu(\zeta) - \phi^{\nu-1}(\zeta)\right| = \left|\phi^{\nu-1}(\psi^\nu(\zeta)) - \phi^{\nu-1}(\zeta)\right| \le 2\,|\psi^\nu(\zeta) - \zeta| \le c\delta(r_\nu - r_{\nu+1})$$

for $|\operatorname{Im}\xi| \le r_{\nu+1}$ and $|\eta| \le r_{\nu+1}$ and $\nu \ge 1$. Here the last inequality follows from (2.17). The same inequality holds for $\nu = 0$ if we define $\phi^{-1} := \mathrm{id}$. We conclude that the limit function

$$\phi := \lim_{\nu\to\infty} \phi^\nu$$

satisfies the estimate

$$|\phi(\zeta) - \zeta| \le c\delta \sum_{\nu=0}^{\infty} (r_\nu - r_{\nu+1}) = c\delta \frac{1-\theta}{2} r$$

for $|\operatorname{Im}\xi| \le r(1+\theta)/2$ and $|\eta| \le r(1+\theta)/2$. This proves the first inequality in (2.4) and the second follows from Lemma 1 (iii). Since $c\delta \le 1/2$ this second estimate also shows that ϕ is a diffeomorphism and, as a limit of symplectomorphisms of the form (1.7), it is itself a symplectomorphism of this form.

The transformed Hamiltonian function can be expressed as the limit

$$K(\zeta) := H \circ \phi(\zeta) = \lim_{\nu\to\infty} H \circ \psi^0 \circ \cdots \circ \psi^\nu(\zeta) = \lim_{\nu\to\infty} H^\nu(\zeta)$$

for $|\operatorname{Im}\xi| \le \theta r$ and $|\eta| \le \theta r$. Since the sequence

$$\frac{\delta_\nu}{(r_\nu - r_{\nu-1})^{\tau+1}} = \left(\frac{4}{(1-\theta)r}\right)^{\tau+1} \frac{\delta_\nu}{2^{\nu(\tau+1)}}$$

converges to zero, by (2.7) and (2.8), it follows from (2.9) that the limit K satisfies (1.8). Moreover,

$$\begin{aligned} |K_{\eta\eta}(\zeta) - Q(\zeta)| &= \lim_{\nu\to\infty} \left|H^\nu_{yy}(\zeta) - Q^0(\zeta)\right| \\ &\le \lim_{\nu\to\infty} \sum_{\mu=0}^{\nu} \left|H^\mu_{yy}(\zeta) - Q^\mu(\zeta)\right| \\ &\le \sum_{\mu=0}^{\infty} 2^{-\mu} \frac{c\delta}{2M} = \frac{c\delta}{M} \end{aligned}$$

for $|\operatorname{Im}\xi| \le \theta r$ and $|\eta| \le \theta r$. Here the third inequality follows from (2.9). Thus we have proved the third inequality in (2.4). To prove the estimate for $v \circ u^{-1}$ we observe that

$$v \circ u^{-1}(x) = \sum_{\nu=1}^{\infty} \left(V^0_x\right)^T \cdot \left(V^1_x\right)^T \cdots \left(V^{\nu-1}_x\right)^T \cdot U^\nu_x$$

for $|\operatorname{Im} x| \le \theta r$. Here we abbreviate

$$V^\mu_x := V^\mu_x(V^{\mu-1} \circ \cdots \circ V^0(x)).$$

This expression is well defined for $|\operatorname{Im} x| \leq \theta r$ since, by Step 2 and (2.19), we have

$$\begin{aligned} \left|V^\nu \circ \cdots \circ V^0(x) - V^{\nu-1} \circ \cdots \circ V^0(x)\right| &\leq \frac{c_2 \varepsilon_\nu}{(r_\nu - r_{\nu+1})^{2\tau+1}} \leq r_\nu - r_{\nu+1}, \\ \left|V^\nu \circ \cdots \circ V^0(x) - x\right| \leq r - r_{\nu+1} &\leq \frac{1-\theta}{2} r \leq \frac{r_{\nu+1} + r_{\nu+2}}{2} - \theta r. \end{aligned}$$

This last expression allows us to use Step 2 and (2.19) again to estimate the terms $V_x^{\nu+1} := V_x^{\nu+1}(V^\nu \circ \cdots \circ V^0(x))$ by $1 + 2^{-\nu-1} c\delta$ for $|\operatorname{Im} x| \leq \theta r$. It also follows from Step 2 and (2.19) that $|U_x(x)| \leq 2^{-\nu} c\delta r^{\tau+1}/4$. Hence we obtain

$$\begin{aligned} \left|v \circ u^{-1}(x)\right| &= \sum_{\nu=0}^{\infty} \left|V_x^0\right| \cdots \left|V_x^{\nu-1}\right| \left|U_x^\nu\right| \\ &\leq \sum_{\nu=1}^{\infty} (1 + c\delta) \cdots (1 + 2^{1-\nu} c\delta) 2^{-\nu} c\delta \frac{r^{\tau+1}}{4} \\ &\leq \frac{e^{2c\delta}}{4} \sum_{\nu=0}^{\infty} 2^{-\nu} c\delta r^{\tau+1} \\ &\leq c\delta r^{\tau+1} \end{aligned}$$

for $|\operatorname{Im} x| \leq \theta r$. This completes the proof of Theorem 1. □

3 The differentiable case

The perturbation theorem for invariant tori for differentiable Hamiltonian functions is due to Moser [13, 14]. He first proved this result in the context of invariant curves for area preserving annulus mappings which satisfy the monotone twist property [14]. This corresponds to the case of two degrees of freedom. One of the main ideas in [13, 14] was to use a smoothing operator in order to compensate for the loss of smoothness that arises in solving the linearized equation. Moreover, it was essential to observe that the error introduced by the smoothing operator would not destroy the rapid convergence of the iteration. A similar approach was also used by Nash [21] for the embedding problem of compact Riemannian manifolds. But this method required an excessive number of derivatives like for example $\ell \geq 333$ for the annulus mapping [14]. Incidentally, this number was later reduced by Rüssmann [24] to $\ell \geq 5$.

In [16] Moser proposed a different approach which is based on approximating differentiable functions by analytic ones. In this section we provide the complete arguments for this second approach as it has also been done by Pöschel in [22] in a somewhat different way. The fundamental observation is that the qualitative property of differentiability of a function can be characterized in terms of quantitative estimates for an approximating sequence of analytic functions. Moser's first proof of this result in [15] was based on a classical approximation theorem due to Jackson. A direct proof was later provided by Zehnder [28]. For the sake of completeness we will include proofs of these results that were given by Moser in a set of unpublished notes entitled *Ein Approximationssatz* (see Lemmata 3 and 4 below).

Let us first recall that $C^\mu(\mathbb{R}^n)$ for $0<\mu<1$ denotes the space of of bounded Hölder continuous functions $f:\mathbb{R}^n\to\mathbb{R}$ with the norm

$$|f|_{C^\mu} := \sup_{0<|x-y|<1} \frac{|f(x)-f(y)|}{|x-y|^\mu} + \sup_{x\in\mathbb{R}^n} |f(x)|\,.$$

If $\mu=0$ then $|f|_{C^\mu}$ denotes the sup-norm. For $\ell=k+\mu$ with $k\in\mathbb{N}$ and $0\le\mu<1$ we denote by $C^\ell(\mathbb{R}^n)$ the space of functions $f:\mathbb{R}^n\to\mathbb{R}$ with Hölder continuous partial derivatives $\partial^\alpha f\in C^\mu(\mathbb{R}^n)$ for all multi-indices $\alpha=(\alpha_1,\dots,\alpha_n)\in\mathbb{N}^n$ with $|\alpha|:=\alpha_1+\cdots+\alpha_n\le k$. We define the norm

$$|f|_{C^\ell} := \sum_{|\alpha|\le\ell} |\partial^\alpha f|_{C^\mu}$$

for $\mu:=\ell-[\ell]<1$. Given an integer $k\ge 0$ and a C^k function $f:\mathbb{R}^n\to\mathbb{R}$ denote by $P_{f,k}:\mathbb{R}^n\times\mathbb{C}^n\to\mathbb{C}$ the Taylor polynomial of f up to order k. Thus

$$P_{f,k}(x;y) := \sum_{|\alpha|\le k} \frac{1}{\alpha!}\partial^\alpha f(x) y^\alpha$$

for $x\in\mathbb{R}^n$ and $y\in\mathbb{C}^n$. Here the sum runs over all multi-indices α with $|\alpha|\le k$. We abbreviate $\alpha! := \alpha_1!\cdots\alpha_n!$ and $y^\alpha := y_1^{\alpha_1}\cdots y_n^{\alpha_n}$ for $y=(y_1,\dots,y_n)\in\mathbb{C}^n$.

Lemma 3 (Jackson, Moser, Zehnder). *There is a family of convolution operators*

$$S_r f(x) = r^{-n}\int_{\mathbb{R}^n} K(r^{-1}(x-y))f(y)\,dy, \qquad 0<r\le 1, \tag{3.1}$$

from $C^0(\mathbb{R}^n)$ into the space of entire functions on $\mathbb{C}^n$ with the following property. For every $\ell\ge 0$, there exists a constant $c=c(\ell,n)>0$ such that, for every $f\in C^\ell(\mathbb{R}^n)$, every multi-index $\alpha\in\mathbb{N}^n$ with $|\alpha|\le\ell$, and every $x\in\mathbb{C}^n$, we have

$$|\mathrm{Im}\,x|\le r \quad\Longrightarrow\quad \left|\partial^\alpha S_r f(x) - P_{\partial^\alpha f,[\ell]-|\alpha|}(\mathrm{Re}\,x; i\mathrm{Im}\,x)\right| \le c\,|f|_{C^\ell}\, r^{\ell-|\alpha|}.$$

Moreover, $K(\mathbb{R}^n)\subset\mathbb{R}$ so that $S_r f$ is real analytic whenever f is real valued.

Proof. Let

$$K(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \widehat{K}(\xi)e^{i\langle x,\xi\rangle}\,d\xi, \qquad x\in\mathbb{C}^n,$$

be an entire function whose Fourier transform

$$\widehat{K}(\xi) = \int_{\mathbb{R}^n} K(x)e^{-i\langle x,\xi\rangle}\,dx, \qquad \xi\in\mathbb{R}^n,$$

is a smooth function with compact support, contained in the ball $|\xi| \le a$, that satisfies $\widehat{K}(\xi) = \widehat{K}(-\xi)$ and

$$\partial^\alpha \widehat{K}(0) = \begin{cases} 1, & \text{if } \alpha = 0, \\ 0, & \text{if } \alpha \ne 0. \end{cases}$$

Then $K : \mathbb{C}^n \to \mathbb{R}$ is a real analytic function with the property

$$\int_{\mathbb{R}^n} (u+iv)^\alpha \partial^\beta K(u+iv)\, du = \begin{cases} (-1)^{|\alpha|}\alpha!, & \text{if } \alpha = \beta, \\ 0, & \text{if } \alpha \ne \beta, \end{cases} \tag{3.2}$$

for $v \in \mathbb{R}^n$ and multi-indices $\alpha, \beta \in \mathbb{N}^n$. The integral is always well defined since, for every $m > 0$ and every $p > 0$, there exists a constant $c_1 = c_1(m,p) > 0$ such that

$$|\beta| \le m \qquad \Longrightarrow \qquad \left|\partial^\beta K(u+iv)\right| \le c_1 \left(1+|u|\right)^{-p} e^{a|v|} \tag{3.3}$$

for all $\beta \in \mathbb{N}^n$ and $u, v \in \mathbb{R}^n$. Moreover, it follows from Cauchy's integral formula that the expression on the left hand side of (3.2) is independent of v. This proves (3.2) in the case $\beta = 0$ and in general it follows from partial integration.

We prove that any such function K satisfies the requirements of the lemma. The proof is based on the observation that, by (3.2), the convolution operator (3.1) acts as the identity on polynomials, i.e.

$$S_r p = p$$

for every polynomial $p : \mathbb{R}^n \to \mathbb{R}$. We will make use of this fact in case of the Taylor polynomial

$$p_k(x;y) := P_{f,k}(x;y) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} \partial^\alpha f(x) y^\alpha$$

of f with $k := [\ell]$. The difference $f - p_k$ can be expressed in the form

$$\begin{aligned} & f(x+y) - p_k(x;y) \\ &= \int_0^1 \int_0^{s_1} \cdots \int_0^{s_{k-1}} \sum_{|\alpha|=k} \frac{k!}{\alpha!} \Big(\partial^\alpha f(x+s_k y) - \partial^\alpha f(x)\Big) y^\alpha \, ds_k ds_{k-1} \cdots ds_1 \\ &= \int_0^1 k(1-t)^{k-1} \sum_{|\alpha|=k} \frac{1}{\alpha!} \Big(\partial^\alpha f(x+ty) - \partial^\alpha f(x)\Big) y^\alpha \, dt. \end{aligned}$$

for $x, y \in \mathbb{R}^n$. This gives rise to the well known estimate

$$|f(x+y) - p_k(x;y)| \le c_2 \, |f|_{C^\ell} \, |y|^\ell \tag{3.4}$$

for $x, y \in \mathbb{R}^n$, $k := [\ell]$, and a suitable constant $c_2 = c_2(n, \ell) > 0$.

Now let $x = u + iv$ with $u, v \in \mathbb{R}^n$. Then

$$\begin{aligned} S_r f(x) &= r^{-n} \int_{\mathbb{R}^n} K\left(r^{-1}(u-y) + ir^{-1}v\right) f(y)\, dy \\ &= \int_{\mathbb{R}^n} K\left(ir^{-1}v - \eta\right) f(u + r\eta)\, d\eta. \end{aligned}$$

Moreover, it follows from (3.2) that

$$\begin{aligned} p_k(u;iv) &= r^{-n}\int_{\mathbb{R}^n} K(r^{-1}(iv-y))p_k(u;y)\,dy \\ &= \int_{\mathbb{R}^n} K\left(ir^{-1}v-\eta\right)p_k(u;r\eta)\,d\eta. \end{aligned}$$

Hence, by (3.4), we have

$$\begin{aligned} |S_rf(u+iv)-p_k(u;iv)| &\le \int_{\mathbb{R}^n}\left|K\left(ir^{-1}v-\eta\right)\right||f(u+r\eta)-p_k(u;r\eta)|\,d\eta \\ &\le c_2\,|f|_{C^\ell}\,r^\ell\int_{\mathbb{R}^n}\left|K\left(ir^{-1}v-\eta\right)\right||\eta|^\ell\,d\eta \\ &\le c_1c_2e^a\,|f|_{C^\ell}\,r^\ell\int_{\mathbb{R}^n}(1+|\eta|)^{\ell-p}\,d\eta. \end{aligned}$$

The last inequality follows from (3.3) with $\beta=0$, $\left|r^{-1}v\right|\le 1$, and $p>\ell+n$ so that the integral on the right is finite. This proves the lemma for $\alpha=0$. For $\alpha\ne 0$ the result follows from the fact that S_r commutes with ∂^α. □

The converse statement of Lemma 3 holds only if ℓ is not an integer. A classical version of this converse result is due to Bernstein and relates the differentiability properties of a periodic function to quantitative estimates for an approximating sequence of trigonometric polynomials [1].

Lemma 4 (Bernstein, Moser). *Let $\ell\ge 0$ and n be a positive integer. Then there exists a constant $c=c(\ell,n)>0$ with the following significance. If $f:\mathbb{R}^n\to\mathbb{R}$ is the limit of a sequence of real analytic functions $f_\nu(x)$ in the strips $|\mathrm{Im}\,x|\le r_\nu:=2^{-\nu}r_0$ such that $0<r_0\le 1$ and*

$$f_0=0,\qquad |f_\nu(x)-f_{\nu-1}(x)|\le Ar_\nu^\ell$$

for $\nu\ge 1$ and $|\mathrm{Im}\,x|\le r_\nu$, then $f\in C^s(\mathbb{R}^n)$ for every $s\le\ell$ which is not an integer and, moreover,

$$|f|_{C^s}\le\frac{cA}{\mu(1-\mu)}r_0^{\ell-s},\qquad 0<\mu:=s-[s]<1.$$

Proof. It is enough to consider the case $\ell=s$. Moreover, once the result has been established for $0<\ell<1$ it follows for $\ell>1$ by Cauchy's estimate, or else by repeatedly applying Lemma 1 (iii). Therefore we may assume that $0<\mu=s=\ell<1$.

Let us now define

$$g_\nu:=f_\nu-f_{\nu-1}.$$

Then $f=\sum_{\nu=1}^\infty g_\nu$ satisfies the estimate

$$|f|_{C^0}\le A\sum_{\nu=1}^\infty r_\nu^\mu=\frac{2^{-\mu}Ar_0^\mu}{1-2^{-\mu}}\le\frac{2A}{\mu}r_0^\mu.$$

Here we have used the inequality $1-2^{-\mu} \geq \mu/2$ for $0<\mu<1$. For $x, y \in \mathbb{R}^n$ with $r_0 < |x-y| \leq 1$ this implies

$$|f(x)-f(y)| \leq \frac{4A}{\mu} r_0^\mu \leq \frac{4A}{\mu(1-\mu)} |x-y|^\mu .$$

In the case $0<|x-y|<r_0$ there is an integer $N \geq 0$ such that

$$2^{-N-1} r_0 \leq |x-y| < 2^{-N} r_0 .$$

Moreover, it follows from Lemma 1 (iii) that

$$|g_{\nu x}(u)| \leq A r_\nu^{\mu-1}$$

for every $u \in \mathbb{R}^n$. and hence

$$|g_\nu(x)-g_\nu(y)| \leq A r_\nu^{\mu-1} |x-y| .$$

We shall use this estimate for $\nu = 1, \ldots, N$. For $\nu > N$ we use the trivial estimate

$$|g_\nu(x)-g_\nu(y)| \leq 2 A r_\nu^\mu .$$

Taking into account the inequalities $1-2^{-\mu} \geq \mu/2$ and $2^{1-\mu}-1 \geq (1-\mu)/2$ for $0<\mu<1$ we conclude that

$$\begin{aligned} |f(x)-f(y)| &\leq \sum_{\nu=1}^\infty |g_\nu(x)-g_\nu(y)| \\ &\leq A|x-y| \sum_{\nu=1}^N \left(2^\nu r_0^{-1}\right)^{1-\mu} + 2A \sum_{\nu=N+1}^\infty \left(2^{-\nu} r_0\right)^\mu \\ &\leq \frac{2A}{2^{1-\mu}-1} |x-y| \left(2^N r_0^{-1}\right)^{1-\mu} + \frac{2A}{1-2^{-\mu}} \left(2^{-N-1} r_0\right)^\mu \\ &\leq \frac{4A}{1-\mu} |x-y|^\mu + \frac{4A}{\mu} |x-y|^\mu \\ &= \frac{4A}{\mu(1-\mu)} |x-y|^\mu . \end{aligned}$$

This proves the lemma. □

The approximation result of Lemma 3 can be used to prove the following interpolation and product estimates. Denote by $C^s(\mathbb{T}^n, \mathbb{R}^k)$ the space of all functions $w \in C^s(\mathbb{R}^n, \mathbb{R}^k)$ that are of period 1 in all variables, and by $C_0^s(\mathbb{T}^n, \mathbb{R}^k)$ the space of all functions $w \in C^s(\mathbb{T}^n, \mathbb{R}^k)$ with mean value zero. For $k=1$ we abbreviate $C^s(\mathbb{T}^n) := C^s(\mathbb{T}^n, \mathbb{R})$ and $C_0^s(\mathbb{T}^n) := C_0^s(\mathbb{T}^n, \mathbb{R})$.

Lemma 5. *For every $n \in \mathbb{N}$ and every $\ell > 0$ there is a constant $c = c(\ell, n) > 0$ such that the following inequalities hold for all $f, g \in C^\ell(\mathbb{T}^n)$:*

$$|f|_{C^m}^{\ell-k} \leq c\, |f|_{C^k}^{\ell-m} |f|_{C^\ell}^{m-k}, \qquad k \leq m \leq \ell,$$

$$|fg|_{C^s} \leq c\left(|f|_{C^0} |g|_{C^s} + |f|_{C^s} |g|_{C^0}\right), \qquad 0 \leq s \leq \ell.$$

Proof. If $S_r : C^0(\mathbb{T}^n) \to C^\infty(\mathbb{T}^n)$ denotes the smoothing operator of Lemma 3 then one checks easily that

$$|f - S_r f|_{C^m} \le c_1 r^{\ell-m} |f|_{C^\ell}, \qquad |S_r f|_{C^m} \le c_1 r^{k-m} |f|_{C^k}$$

for all $f \in C^\ell(\mathbb{T}^n)$, $0 < r \le 1$, $k \le m \le \ell$, and a suitable constant $c_1 = c_1(\ell, n) > 0$. Choosing $r > 0$ so as to satisfy

$$r^{\ell-k} = \frac{|f|_{C^k}}{|f|_{C^\ell}}$$

we obtain

$$\begin{aligned} |f|_{C^m} &\le c_1 \left(r^{\ell-m} |f|_{C^\ell} + r^{k-m} |f|_{C^k} \right) \\ &= 2c_1 |f|_{C^k}^{(\ell-m)/(\ell-k)} |f|_{C^\ell}^{(m-k)/(\ell-k)}. \end{aligned}$$

This proves the first estimate. Moreover, the second follows from the first since

$$\partial^\beta (fg) = \sum_{\alpha \le \beta} \binom{\beta}{\alpha} (\partial^\alpha f) \left(\partial^{\beta-\alpha} g\right)$$

where $\alpha \le \beta$ if and only if $\alpha_\nu \le \beta_\nu$ for all ν and

$$\binom{\beta}{\alpha} := \binom{\beta_1}{\alpha_1} \cdots \binom{\beta_n}{\alpha_n}.$$

This proves the lemma. □

The next theorem is the main result of this paper. It is Moser's perturbation theorem for invariant tori of differentiable Hamiltonian systems. We have phrased it in the form that the existence of an *approximate invariant torus* implies the existence of a true invariant torus nearby. It then follows that every invariant torus of class $C^{\ell+1}$ satisfying (1.5) with a frequency vector satisfying (1.3) persists under C^ℓ small perturbations of the Hamiltonian function. In this formulation it is also an obvious consequence that every invariant torus of class $C^{\ell+1}$ satisfying (1.5) gives rise to nearby invariant tori for nearby frequency vectors that satisfy the Diophantine inequalities (1.3).

Theorem 2 (Moser). *Let $n \ge 2$, $\tau > n - 1$, $c_0 > 0$, $m > 0$, $\ell > 2\tau + 2 + m$, $M \ge 1$, and $\rho > 0$ be given. Then there are positive constant ε^* and c such that the following holds for every vector $\omega \in \mathbb{R}^n$ that satisfies (1.3) and every open set $G \subset \mathbb{R}^n$ that contains the ball $B_\rho(0)$.*

Suppose $H \in C^\ell(\mathbb{T}^n \times G)$ satisfies

$$|H|_{C^\ell} \le M, \qquad \left| \left(\int_{\mathbb{T}^n} H_{yy}(\xi, 0)\, d\xi \right)^{-1} \right| \le M, \tag{3.5}$$

and

$$\left| H(x,0) - \int_{\mathbb{T}^n} H(\xi,0)\, d\xi \right| + |H_y(x,0) - \omega|\, \varepsilon^{\tau+1} < M\varepsilon^{\ell} \tag{3.6}$$

for every $x \in \mathbb{R}^n$ and some constant $0 < \varepsilon \le \varepsilon^*$. Then there is a solution

$$x = u(\xi), \qquad y = v(\xi)$$

of (1.2) such that $u(\xi) - \xi$ and $v(\xi)$ are of period 1 in all variables. Moreover, $u \in C^s(\mathbb{R}^n, \mathbb{R}^n)$ and $v \circ u^{-1} \in C^{s+\tau}(\mathbb{R}^n, G)$ for every $s \le m+1$ such that $s \notin \mathbb{N}$ and $s + \tau \notin \mathbb{N}$ and

$$\begin{aligned} |u - \mathrm{id}|_{C^s} &\le \frac{c}{\mu(1-\mu)} \varepsilon^{m+1-s}, \quad 0 < s \le m+1, \\ \left| v \circ u^{-1} \right|_{C^s} &\le \frac{c}{\mu(1-\mu)} \varepsilon^{m+\tau+1-s}, \quad 0 < s \le m+\tau+1, \end{aligned} \tag{3.7}$$

where $0 < \mu := s - [s] < 1$.

Proof. By Lemma 3, we can approximate $H(x,y)$ by a sequence of real analytic functions $H^\nu(x,y)$ for $\nu = 0,1,2,\dots$ in the strips

$$|\mathrm{Im}\, x| \le r_{\nu-1}, \qquad |\mathrm{Im}\, y| \le r_{\nu-1}, \qquad r_\nu := 2^{-\nu}\varepsilon,$$

around $\mathrm{Re}\, x \in \mathbb{T}^n$, $|\mathrm{Re}\, y| \le \rho$, such that

$$\begin{aligned} \left| H^\nu(z) - \sum_{|\alpha| \le \ell} \partial^\alpha H(\mathrm{Re}\, z) \frac{(i\mathrm{Im}\, z)^\alpha}{\alpha!} \right| &\le c_1 \, |H|_{C^\ell} \, r_\nu^{\ell}, \\ \left| H_y^\nu(z) - \sum_{|\alpha| \le \ell} \partial^\alpha H_y(\mathrm{Re}\, z) \frac{(i\mathrm{Im}\, z)^\alpha}{\alpha!} \right| &\le c_1 \, |H|_{C^\ell} \, r_\nu^{\ell-1}, \\ \left| H_{yy}^\nu(z) - \sum_{|\alpha| \le \ell} \partial^\alpha H_{yy}(\mathrm{Re}\, z) \frac{(i\mathrm{Im}\, z)^\alpha}{\alpha!} \right| &\le c_1 \, |H|_{C^\ell} \, r_\nu^{\ell-2}, \end{aligned} \tag{3.8}$$

for $|\mathrm{Im}\, x| \le r_{\nu-1}$, $|\mathrm{Im}\, y| \le r_{\nu-1}$, and $|\mathrm{Re}\, y| \le \rho$. Here the constant $c_1 = c_1(\ell, n) > 0$ is chosen appropriately (Lemma 3) and $\varepsilon > 0$ is the number appearing in (3.6).

Fix the constant $\theta := 1/\sqrt{2}$. By induction, we shall construct a sequence of real analytic symplectic transformations $z = \phi^\nu(\zeta)$ of the form

$$x = u^\nu(\xi), \qquad y = v^\nu(\xi) + \left(u_\xi^\nu\right)^T (\xi)^{-1} \eta, \tag{3.9}$$

such that $u^\nu(\xi) - \xi$ and $v^\nu(\xi)$ are of period 1 in all variables, ϕ^ν maps the strip $|\mathrm{Im}\, \xi| \le \theta r_{\nu+1}$, $|\eta| \le \theta r_{\nu+1}$ into $|\mathrm{Im}\, x| \le r_\nu$, $|\mathrm{Im}\, y| \le r_\nu$, $|\mathrm{Re}\, y| \le \rho$, and the transformed Hamiltonian function

$$K^\nu := H^\nu \circ \phi^\nu$$

satisfies (1.8).

For $\nu = 0$ we shall use the smallness assumption on $H^0 - H$ and condition (3.6) in order to verify that the real analytic Hamiltonian function $H^0(x, y)$ satisfies the assumptions of Theorem 1 in the strip of size $r = \theta\varepsilon$ with $\delta = \varepsilon^m$ (Step 1). Having established the existence of $\phi^{\nu-1}$ for some $\nu \geq 1$, we use the smallness condition on $H^\nu - H$ and $H - H^{\nu-1}$ in order to verify that $H^\nu \circ \phi^{\nu-1}$ satisfies the assumptions of Theorem 1 with $r = \theta r_\nu$ and $\delta = r_\nu^m$ (Step 2). This guarantees the existence of a symplectic transformation $z = \psi^\nu(\zeta)$ of the form (1.7) from the strip $|\mathrm{Im}\,\xi| \leq r_{\nu+1}$, $|\eta| \leq r_{\nu+1}$ to $|\mathrm{Im}\,x| \leq \theta r_\nu$, $|y| \leq \theta r_\nu$ such that $\psi^\nu(\xi, 0) - (\xi, 0)$ is of period 1 and $K^\nu := H^\nu \circ \phi^\nu$ satisfies (1.8), where

$$\phi^\nu := \phi^{\nu-1} \circ \psi^\nu.$$

Moreover, Theorem 1 will yield the estimates

$$\begin{aligned}
&|\psi^\nu(\zeta) - \zeta| \leq c_2(1-\theta)r_\nu^{m+1}, \quad \left|\psi^\nu_\zeta(\zeta) - \mathbb{1}\right| \leq c_2 r_\nu^m, \\
&\left|K^\nu_{\eta\eta}(\zeta) - Q^\nu(\zeta)\right| \leq \frac{c_2 r_\nu^m}{2M}, \\
&|U^\nu_x(x)| \leq \theta c_2 r_\nu^{m+\tau+1},
\end{aligned} \tag{3.10}$$

for $|\mathrm{Im}\,\xi| \leq r_{\nu+1}$, $|\eta| \leq r_{\nu+1}$, and $|\mathrm{Im}\,x| \leq r_{\nu+1}$, where $Q^\nu := K^{\nu-1}_{\eta\eta}$ for $\nu \geq 1$,

$$Q^0(z) := \sum_{|\alpha| \leq \ell-2} \partial^\alpha H_{yy}(\mathrm{Re}\,z)\frac{(i\mathrm{Im}\,z)^\alpha}{\alpha!},$$

and

$$S^\nu(x, \eta) = U^\nu(x) + \langle V^\nu(x), \eta\rangle$$

is the generating function for ψ^ν. In Step 3 we will show that (3.10) implies the inequalities

$$\begin{aligned}
&\left|\phi^\nu(\zeta) - \phi^{\nu-1}(\zeta)\right| \leq 2c_2(1-\theta)r_\nu^{m+1}, \quad |\mathrm{Im}\,\xi| \leq r_{\nu+1},\ |\eta| \leq r_{\nu+1}, \\
&\left|\phi^\nu_\zeta(\zeta) - \phi^{\nu-1}_\zeta(\zeta)\right| \leq 4c_2 r_\nu^m, \quad |\mathrm{Im}\,\xi| \leq \theta r_{\nu+1},\ |\eta| \leq \theta r_{\nu+1}, \\
&\left|v^\nu \circ (u^\nu)^{-1}(x) - v^{\nu-1} \circ (u^{\nu-1})^{-1}(x)\right| \leq c_2 r_\nu^{m+\tau+1}, \quad |\mathrm{Im}\,x| \leq \theta r_{\nu+1}.
\end{aligned} \tag{3.11}$$

We denote by $c_2 = c_2(n, \tau, c_0, \theta, 2M) \geq 4M$ and $\delta^* = \delta^*(n, \tau, c_0, \theta, 2M) > 0$ the constants of Theorem 1 with $\theta := 1/\sqrt{2}$, by $c_3 > 0$ the constant of Lemma 4 with ℓ replaced by $m + \tau + 1$ and $m + 1$, and by $c_4 = c_4(\ell, n) > 0$ the constant of Lemma 5. The constants $c_5 > 0$, $c > 0$, $\gamma > 0$, and $\varepsilon^* > 0$ will be chosen as to satisfy $\varepsilon^* \leq \rho$ and

$$\varepsilon^{*m} \leq \delta^*, \qquad 2n\varepsilon^* \leq \log 2, \tag{3.12}$$

$$(4c_1 + (\ell+1)c_4)M\varepsilon^{*\gamma} \leq (\theta/2)^\ell, \qquad \gamma := \ell - 2\tau - 2 - m, \tag{3.13}$$

$$c_5\varepsilon^{*m} \leq 1 - \theta \leq \log 2, \qquad c_5 := \frac{4c_2}{1 - 2^{-m}}, \tag{3.14}$$

$$c := \left(\frac{2}{\theta}\right)^{m+\tau+1} c_2 c_3. \tag{3.15}$$

Given $r_\nu := 2^{-\nu}\varepsilon$ with $0 < \varepsilon \le \varepsilon^*$ we define the numbers $M_\nu > 0$ recursively by

$$M_\nu := \frac{M_{\nu-1}}{1 - c_2 r_\nu^m}, \qquad M_0 := M, \tag{3.16}$$

and observe that, by (3.14), we have

$$M_\nu \le M_{\nu-1} e^{2c_2\varepsilon_\nu^m} \le M e^{c_5 \varepsilon^{*m}} \le 2M. \tag{3.17}$$

Step 1. *There is a symplectic transformation $z = \psi^0(\zeta)$ of the form (1.7) from the strip $|\mathrm{Im}\,\xi| \le r_1$, $|\eta| \le r_1$ to $|\mathrm{Im}\,x| \le \theta r_0$, $|y| \le \theta r_0$ such that $\psi^0(\xi,0) - (\xi,0)$ is of period 1 and $K^0 := H^0 \circ \psi^0$ satisfies (1.8). Moreover, K^0 and ψ^0 satisfy the estimates in (3.10) for $\nu = 0$.*

First abbreviate

$$h(x) := H(x,0) - \int_{\mathbb{T}^n} H(\xi,0)\,d\xi, \qquad x \in \mathbb{R}^n.$$

Then $|h|_{C^\ell} \le M$ and $|h|_{C^0} \le M\varepsilon^\ell$, by (3.6). Hence, by Lemma 5, we have

$$|h|_{C^k} \le c_4 \,|h|_{C^\ell}^{k/\ell}\,|h|_{C^0}^{1-k/\ell} \le c_4 M \varepsilon^{\ell-k}$$

for $0 \le k \le \ell$. Now

$$\begin{aligned} H^0(x,0) - \int_{\mathbb{T}^n} H^0(\xi,0)\big)\,d\xi \;&=\; H^0(x,0) - \sum_{|\alpha|\le\ell} \partial_x^\alpha H(\mathrm{Re}\,x,0)\frac{(i\mathrm{Im}\,x)^\alpha}{\alpha!} \\ &\quad + \int_{\mathbb{T}^n}\big(H(\xi,0) - H^0(\xi,0)\big)\,d\xi \\ &\quad + \sum_{|\alpha|\le\ell} \partial^\alpha h(\mathrm{Re}\,x)\frac{(i\mathrm{Im}\,x)^\alpha}{\alpha!}. \end{aligned}$$

If $|\mathrm{Im}\,x| \le \theta r_0 = \theta\varepsilon$ then it follows from (3.8) that

$$\begin{aligned} \left| H^0(x,0) - \int_{\mathbb{T}^n} H^0(\xi,0)\,d\xi \right| \;&\le\; 2c_1\,|H|_{C^\ell}\,\varepsilon^\ell + \sum_{k=0}^{\ell} |h|_{C^k}\,\varepsilon^k \\ &\le\; (2c_1 + (\ell+1)c_4) M \varepsilon^\gamma \varepsilon^m \varepsilon^{2\tau+2} \\ &\le\; \varepsilon^m (\theta\varepsilon)^{2\tau+2}. \end{aligned}$$

The second inequality follows from $\ell = \gamma + m + 2\tau + 2$ and the last from (3.13).

Second, consider the vector valued function

$$f(x) := H_y(x,0) - \omega, \qquad x \in \mathbb{R}^n.$$

It satisfies $|f|_{C^{\ell-\tau-1}} \le M$ and $|f|_{C^0} \le M\varepsilon^{\ell-\tau-1}$, by (3.6). Hence, by Lemma 5, we have

$$|f|_{C^k} \le c_4 \, |f|_{C^{\ell-\tau-1}}^{k/(\ell-\tau-1)} \, |f|_{C^0}^{1-k/(\ell-\tau-1)} \le c_4 M \varepsilon^{\ell-\tau-1-k}.$$

for $0 \le k \le \ell-\tau-1$. The estimate $|f|_{C^k} \le c_4 M\varepsilon^{\ell-\tau-1-k}$ obviously continues to hold for $\ell-\tau-1 < k \le \ell-1$. Now

$$\begin{aligned} H_y^0(x,0) - \omega &= H_y^0(x,0) - \sum_{|\alpha|\le \ell-1} \partial_x^\alpha H_y(\mathrm{Re}\, x, 0) \frac{(i\mathrm{Im}\, x)^\alpha}{\alpha!} \\ &\quad + \sum_{|\alpha|\le \ell-1} \partial^\alpha f(\mathrm{Re}\, x) \frac{(i\mathrm{Im}\, x)^\alpha}{\alpha!}. \end{aligned}$$

Hence, with $|\mathrm{Im}\, x| \le \theta\varepsilon$, it follows from (3.8) that

$$\begin{aligned} \left|H_y^0(x,0) - \omega\right| &\le c_1 \, |H|_{C^\ell} \, \varepsilon^{\ell-1} + \sum_{k=0}^{\ell-1} |f|_{C^k} \, \varepsilon^k \\ &\le (c_1+\ell) M \varepsilon^{\ell-\tau-1} \\ &= (c_1+\ell) M \varepsilon^\gamma \varepsilon^m \varepsilon^{\tau+1} \\ &\le \varepsilon^m (\theta\varepsilon)^{\tau+1}. \end{aligned}$$

Here the last inequality follows from (3.13).

Third, it follows from the definition of Q^0 by

$$Q^0(z) := \sum_{|\alpha|\le \ell-2} \partial^\alpha H_{yy}(\mathrm{Re}\, z) \frac{(i\mathrm{Im}\, z)^\alpha}{\alpha!}$$

and (3.8) that

$$\left|H_{yy}^0(z) - Q^0(z)\right| \le c_1 \, |H|_{C^\ell} \, \varepsilon^{\ell-2} \le c_1 M \varepsilon^\gamma \varepsilon^m \le \frac{c_2 \varepsilon^m}{4M}$$

for $|\mathrm{Im} x| \le \theta\varepsilon$ and $|y| \le \theta\varepsilon$. Here the last inequality follows from the fact that $c_2 \ge 4M$ and $c_1 M \varepsilon^\gamma \le 1$, by (3.13). Moreover, since $\sum_{\alpha\in\mathbb{N}^{2n}} \varepsilon^{|\alpha|}/\alpha! = e^{2n\varepsilon}$ and $2n\varepsilon \le \log 2$, by (3.12), we have

$$\left|Q^0(z)\right| \le \sum_{|\alpha|\le \ell-2} M \frac{\varepsilon^{|\alpha|}}{\alpha!} \le M e^{2n\varepsilon} \le 2M$$

for $|\mathrm{Im}\, z| \le \varepsilon$. Since $Q^0(z) = H_{yy}(z)$ for $z \in \mathbb{R}^{2n}$ we have proved that H^0 and Q^0 satisfy the hypotheses of Theorem 1 with M replaced by $2M$ and $\delta = \varepsilon^m \le \delta^*$ and $r = \theta r_0 = \theta\varepsilon$. Hence Step 1 follows from the assertion of Theorem 1.

Observe that the inequalities in (3.11) for $\nu = 0$ follow immediately from Step 1 if we define $\phi^{-1} := \mathrm{id}$ and $\phi^0 := \phi^{-1} \circ \psi^0 = \psi^0$.

Now assume, by induction, that the transformation $z = \phi^{\nu-1}(\zeta)$ of the form (3.9) from the strip $|\mathrm{Im}\,\xi| \le \theta r_\nu$, $|\eta| \le \theta r_\nu$ into $|\mathrm{Im}\,x| \le r_{\nu-1}$, $|\mathrm{Im}\,y| \le r_{\nu-1}$, $|\mathrm{Re}\,y| \le \rho$ has been constructed such that $u^{\nu-1}(\xi) - \xi$ and $v^{\nu-1}(\xi)$ have period 1 in all variables, $K^{\nu-1} := H^{\nu-1} \circ \phi^{\nu-1}$ satisfies (1.8), and (3.11) holds with ν replaced by $\nu - 1$. Assume also that the transformations $z = \psi^\mu(\zeta)$ satisfy (3.10) for $\mu = 0, \dots, \nu - 1$. Then we obtain from (3.16) and (3.17) by induction that

$$\left|K_{\eta\eta}^{\nu-1}(\zeta)\right| \le M_{\nu-1}, \qquad \left|\left(\int_{\mathbb{T}^n} K_{\eta\eta}^{\nu-1}(\xi,0)\,d\xi\right)^{-1}\right| \le M_{\nu-1} \tag{3.18}$$

for $|\mathrm{Im}\,\xi| \le r_\nu$ and $|\eta| \le r_\nu$. Define the Hamiltonian function $\widetilde{H}$ by

$$\widetilde{H}(x,y) := H^\nu \circ \phi^{\nu-1}(x,y)$$

for $|\mathrm{Im}\,x| \le \theta r_\nu$ and $|y| \le \theta r_\nu$. This is possible because $\phi^{\nu-1}$ maps this strip into the domain of H^ν.

Step 2. *The Hamiltonian function $\widetilde{H}$ satisfies the estimates*

$$\left|\widetilde{H}(x,0) - \int_{\mathbb{T}^n} \widetilde{H}(\xi,0)\,d\xi\right| \le r_\nu^m (\theta r_\nu)^{2\tau+2},$$

$$\left|\widetilde{H}_y(x,0) - \omega\right| \le r_\nu^m (\theta r_\nu)^{\tau+1},$$

$$\left|\widetilde{H}_{yy}(x,y) - Q^\nu(x,y)\right| \le \frac{c_2 r_\nu^m}{4M}$$

for $|\mathrm{Im}\,x| \le \theta r_\nu$ and $|y| \le \theta r_\nu$. Here we abbreviate $Q^\nu := K_{\eta\eta}^{\nu-1}$.

If $|\mathrm{Im}\,x| \le \theta r_\nu$ then $\phi^{\nu-1}(x,0)$ lies in the region where the estimate (3.8) holds for both H^ν and $H^{\nu-1}$. Therefore it follows from (3.8) that

$$\begin{aligned}\left|\widetilde{H}(x,0) - \int_{\mathbb{T}^n} \widetilde{H}(\xi,0)\,d\xi\right| &\le 2 \sup_{|\mathrm{Im}\,\xi| \le \theta r_\nu} \left|H^\nu(\phi^{\nu-1}(\xi,0)) - H^{\nu-1}(\phi^{\nu-1}(\xi,0))\right| \\ &\le 2c_1 M r_\nu^\ell + 2c_1 M r_{\nu-1}^\ell \\ &\le \left(\frac{2}{\theta}\right)^\ell 4c_1 M r_\nu^\gamma r_\nu^m \,(\theta r_\nu)^{2\tau+2} \\ &\le r_\nu^m \,(\theta r_\nu)^{2\tau+2}\end{aligned}$$

for $|\mathrm{Im}\,x| \le \theta r_\nu$. Here the last inequality follows from (3.13). Now the second estimate in (3.10) with $\eta = 0$ implies that, for $|\mathrm{Im}\,\xi| \le \theta r_\nu$, we have

$$\left|u_\xi^{\nu-1}(\xi) - \mathbb{1}\right| \le \sum_{\mu=0}^{\nu-1} \left|u_\xi^\mu(\xi) - u_\xi^{\mu-1}(\xi)\right| \le 4c_2 \sum_{\mu=0}^{\infty} r_\mu^m = c_5 \varepsilon^m \le 1 - \theta. \tag{3.19}$$

Here we have used (3.14) twice. It follows from (3.19) that

$$|\operatorname{Im}\xi| \le \theta r_\nu \qquad \Longrightarrow \qquad \left|u_\xi^{\nu-1}(\xi)^{-1}\right| \le \theta^{-1}. \tag{3.20}$$

Therefore it follows from (3.8) and (3.13) that

$$\begin{aligned}
\left|\widetilde{H}_y(x,0) - \omega\right| &= \left|u_\xi^{\nu-1}(x)^{-1}\left(H_y^\nu(\phi^{\nu-1}(x,0)) - H_y^{\nu-1}(\phi^{\nu-1}(x,0))\right)\right| \\
&\le \theta^{-1}\left(c_1 r_\nu^{\ell-1} + c_1 r_{\nu-1}^{\ell-1}\right) \\
&\le \left(\frac{2}{\theta}\right)^\ell c_1 r_\nu^\gamma r_\nu^m \left(\theta r_\nu\right)^{2\tau+1} \\
&\le r_\nu^m \left(\theta r_\nu\right)^{\tau+1}
\end{aligned}$$

and

$$\begin{aligned}
\left|\widetilde{H}_{yy}(z) - Q^\nu(z)\right| &= \left|u_\xi^{\nu-1}(x)^{-1}\left(H_{yy}^\nu(\phi^{\nu-1}(z)) - H_{yy}^{\nu-1}(\phi^{\nu-1}(z))\right)u_\xi^{\nu-1}(x)^{T^{-1}}\right| \\
&\le \theta^{-2}\left(c_1 r_\nu^{\ell-2} + c_1 r_{\nu-1}^{\ell-2}\right) \\
&\le \left(\frac{2}{\theta}\right)^\ell c_1 r_\nu^\gamma r_\nu^m \left(\theta r_\nu\right)^{2\tau} \\
&\le \frac{c_2 r_\nu^m}{4M}
\end{aligned}$$

for $|\operatorname{Im} x| \le \theta r_\nu$ and $|y| \le \theta r_\nu$. The last inequality uses the fact that $c_2 \ge 4M$ and $(2/\theta)^\ell c_1 r_\nu^\gamma \le 1$, by (3.12). This proves Step 2.

It follows from Step 2 and (3.18) that the Hamiltonian function $\widetilde{H}$ satisfies the hypotheses of Theorem 1 with M replaced by $2M$, $r = \theta r_\nu$, $\delta = r_\nu^m$, and $Q = Q^\nu = K_{\eta\eta}^{\nu-1}$. Hence Theorem 1 asserts that there exists a symplectic transformation $z = \psi^\nu(\zeta)$ of the form (1.7) from the strip $|\operatorname{Im}\xi| \le r_{\nu+1} = \theta^2 r_\nu$, $|\eta| \le r_{\nu+1}$ into $|\operatorname{Im} x| \le \theta r_\nu$, $|y| \le \theta r_\nu$ such that $\psi^\nu(\xi,0) - (\xi,0)$ is of period 1 in all variables, the transformed Hamiltonian function

$$K^\nu := \widetilde{H} \circ \psi^\nu = H^\nu \circ \phi^\nu, \qquad \phi^\nu := \phi^{\nu-1} \circ \psi^\nu,$$

satisfies (1.8), and ψ^ν and K^ν satisfy (3.10). Observe that, by construction, the transformation $\phi^\nu = \phi^{\nu-1} \circ \psi^\nu$ maps the strip $|\operatorname{Im}\xi| \le r_{\nu+1}$, $|\eta| \le r_{\nu+1}$ into $|\operatorname{Im} x| \le r_{\nu-1}$, $|\operatorname{Im} y| \le r_{\nu-1}$, $|\operatorname{Re} y| \le \rho$. But we will show the following.

Step 3. *The transformation $z = \phi^\nu(\zeta)$ maps the strip $|\operatorname{Im}\xi| \le \theta r_{\nu+1}$, $|\eta| \le \theta r_{\nu+1}$ into $|\operatorname{Im} x| \le r_\nu$, $|\operatorname{Im} y| \le r_\nu$, $|\operatorname{Re} y| \le \rho$ and satisfies (3.11).*

Using the second inequality in (3.11) with ν replaced by $\mu = 0, \ldots, \nu-1$, we obtain

$$\left|\phi_\zeta^{\nu-1}(\zeta)\right| \le 1 + \sum_{\mu=0}^{\nu-1}\left|\phi_\zeta^\mu(\zeta) - \phi_\zeta^{\mu-1}(\zeta)\right| \le 1 + 4c_2\sum_{\mu=0}^{\infty} r_\mu^m = 1 + c_5\varepsilon^m \le 2$$

for $|\mathrm{Im}\,\xi| \le \theta r_\nu$ and $|\eta| \le \theta r_\nu$. Here we have used (3.14) again. Hence it follows from (3.10) that, for $|\mathrm{Im}\,\xi| \le r_{\nu+1}$ and $|\eta| \le r_{\nu+1}$, we have

$$\left|\phi^\nu(\zeta) - \phi^{\nu-1}(\zeta)\right| = \left|\phi^{\nu-1}(\psi^\nu(\zeta)) - \phi^{\nu-1}(\zeta)\right| \le 2\,|\psi^\nu(\zeta) - \zeta| \le 2c_2(1-\theta)r_\nu^{m+1}.$$

Thus we have established the first inequality in (3.11) and the second follows immediately from the first and Lemma 1 (iii). As we have seen above, this implies that

$$|\mathrm{Im}\,\xi| \le \theta r_{\nu+1}, \quad |\eta| \le \theta r_{\nu+1} \qquad \Longrightarrow \qquad \left|\phi^\nu_\zeta(\zeta)\right| \le 2$$

and hence $z := \phi^\nu(\zeta)$ satisfies

$$|\mathrm{Im}\, z| \le 2\,|\mathrm{Im}\,\zeta| \le 2\left(|\mathrm{Im}\,\xi|^2 + |\mathrm{Im}\,\eta|^2\right)^{1/2} \le 2r_{\nu+1} = r_\nu.$$

Moreover, we have seen already that $|\mathrm{Re}\, y| \le \rho$.

In order to establish the last inequality in (3.11) we make use of the identity

$$v^\nu \circ (u^\nu)^{-1}(x) - v^{\nu-1} \circ (u^{\nu-1})^{-1}(x) = \left(u^{\nu-1}_\xi(\xi)^{-1}\right)^T U^\nu_x(\xi)$$

for $x = u^{\nu-1}(\xi)$. From (3.19) we obtain that if $|\mathrm{Im}\, x| \le \theta r_{\nu+1}$ and $x =: u^{\nu-1}(\xi)$ then $|\mathrm{Im}\,\xi| \le r_{\nu+1}$. (The map $\xi \mapsto x + \xi - u^{\nu-1}(\xi)$ defines a contraction from the strip $|\mathrm{Im}\,\xi| \le r_{\nu+1}$ to itself.) This allows us to apply the inequalities (3.10) and (3.20) so that

$$\left|\left(u^{\nu-1}_\xi(\xi)^{-1}\right)^T U^\nu_x(\xi)\right| \le c_2\varepsilon_\nu^{m+\tau+1}.$$

Thus we have proved Step 3.

Step 3 finishes the induction and it remains to establish the convergence of the sequences $u^\nu(\xi)$ and $v^\nu(\xi)$ along with the estimates in (3.7). But the inequalities in (3.11) imply that

$$\begin{aligned} \left|u^\nu(\xi) - u^{\nu-1}(\xi)\right| &\le 2^{m+2}c_2 r_{\nu+1}^{m+1} \\ \left|v^\nu \circ (u^\nu)^{-1}(x) - v^{\nu-1} \circ (u^{\nu-1})^{-1}(x)\right| &\le \left(\frac{2}{\theta}\right)^{m+\tau+1} c_2\,(\theta r_{\nu+1})^{m+\tau+1} \end{aligned} \tag{3.21}$$

for $|\mathrm{Im}\,\xi| \le r_{\nu+1}$ and $|\mathrm{Im}\, x| \le \theta r_{\nu+1}$. In particular, these estimates hold for $\nu = 0$ since in that case $u^{\nu-1} = \mathrm{id}$ and $v^{\nu-1} = 0$. Hence it follows from Lemma 4 that the limit functions

$$u(\xi) := \lim_{\nu\to\infty} u^\nu(\xi), \qquad v(\xi) := \lim_{\nu\to\infty} v^\nu(\xi)$$

satisfy (3.7) with $c := (2/\theta)^{m+\tau+1}c_2c_3$. Moreover, The functions $x = u(\xi)$ and $y = v(\xi)$ satisfy (1.2) as well as the periodicity requirements (because u^ν and v^ν satisfy (1.2) with H replaced by H^ν, and all three functions u^ν, v^ν, and H^ν converge in the C^1-topology). This completes the proof of Theorem 2. □

The following C^∞ result is a simple consequence of the proof of Theorem 2. Other versions of it can be found in the work of Zehnder [28], Pöschel [22], and Bost [5].

Corollary 1. *Let $H \in C^\ell(\mathbb{T}^n \times G)$ satisfy the requirements of Theorem 2 and let $x = u(\xi)$ and $y = v(\xi)$ be the solutions of (1.2) constructed in the proof of Theorem 2. Moreover, let $\gamma > 0$ be the constant defined in (3.13) and let $\ell' > \ell$ be any number such that $m' := \ell' - 2\tau - 2 - \gamma \notin \mathbb{N}$ and $m' + \tau \notin \mathbb{N}$. Then*

$$H \in C^{\ell'} \qquad \Longrightarrow \qquad u \in C^{m'+1}, \quad v \circ u^{-1} \in C^{m'+\tau+1}$$

and, in particular,

$$H \in C^\infty \qquad \Longrightarrow \qquad u, v \in C^\infty.$$

Proof. By Lemma 3, there exists a constant c_1' such that the inequalities in (3.8) are satisfied with c_1 and ℓ replaced by c_1' and ℓ', respectively. Now one checks easily that the assertions of Step 2 in the proof of Theorem 2 continue to hold with m replaced by m' provided that ν is sufficiently large. But this implies that the inequalities (3.10), and hence also (3.11), are still satisfied, with m replaced by m', for ν sufficiently large. Hence we deduce that (3.21) also still holds with m replaced by m' for large ν. Therefore Lemma 4 implies that $u \in C^{m'+1}$ and $v \circ u^{-1} \in C^{m'+\tau+1}$. This proves the corollary. □

4 Uniqueness and regularity

In the context of area preserving annulus mappings the uniqueness of an invariant curve with a given irrational rotation number follows easily from the monotone twist property in connection with Denjoy's theory. In this section we prove a local uniqueness result for invariant tori with a given frequency vector $\omega \in \mathbb{R}^n$ which in this form seems to be new. In a more general abstract setting the uniqueness problem has been discussed by Zehnder [28] for the analytic case and our methods are closely related to Zehnder's work. We begin with the following estimate for the solutions of the differential equation $Df = g$.

Lemma 6. *Let $n \geq 2$, $\tau > n - 1$, $c_0 > 0$ and $\ell > \tau$ be given. Then there exists a constant $c = c(n, \tau, c_0, \ell) > 0$ such that the following holds. If $\omega \in \mathbb{R}^n$ satisfies (1.3) and $g \in C_0^\ell(\mathbb{T}^n)$ then the equation $Df = g$ has a unique solution $f \in C_0^s(\mathbb{T}^n)$ for $s + \tau \leq \ell$, $s \notin \mathbb{N}$, and this solution satisfies the estimate*

$$|f|_{C^s} \leq \frac{c}{\mu(1-\mu)} |Df|_{C^{s+\tau}}, \qquad 0 < \mu := s - [s] < 1, \tag{4.1}$$

for every $s \leq \ell - \tau$ with $s \notin \mathbb{N}$.

Proof. By Lemma 3, there exists a sequence of real analytic functions $g_\nu(x)$ on the strip $|\mathrm{Im}\, x| \leq r_\nu := 2^{-\nu}$ which are of period 1 in all variables, have mean value zero on $\mathbb{T}^n$, converge to g on $\mathbb{R}^n$, and satisfy

$$g_0 = 0, \qquad |g_\nu - g_{\nu-1}|_{r_\nu} \leq c_1 r_\nu^{s+\tau} |g|_{C^{s+\tau}}$$

for a suitable constant $c_1 = c_1(\ell, n) > 0$. Hence it follows from Lemma 2 that the unique solution f_ν of $Df_\nu = g_\nu$ satisfies the estimate

$$|f_\nu - f_{\nu-1}|_{r_{\nu+1}} \le \frac{c_2}{(r_\nu - r_{\nu-1})^\tau} |g_\nu - g_{\nu-1}|_{r_\nu} \le c_1 c_2 2^{s+\tau} r_{\nu+1}^s |g|_{C^{s+\tau}}$$

for some constant $c_2 = c_2(n, \tau, c_0) > 0$. Hence Lemma 4 asserts that the limit function $f := \lim f_\nu$ belongs to $C_0^s(\mathbb{T}^n)$ and satisfies the estimate (4.1) with $c = 2^\ell c_1 c_2$. This proves the lemma. □

We are now in a position to prove the desired uniqueness theorem for invariant tori with a given frequency vector $\omega \in \mathbb{R}^n$.

Theorem 3 (Uniqueness). *Let $n \ge 2$, $\tau > n-1$, $c_0 > 0$, $0 < \gamma < 1$, and $M \ge 1$ be given. Then there exists a constant $\delta = \delta(n, \tau, c_0, \gamma, M) > 0$ with the following significance. Suppose $\omega \in \mathbb{R}^n$ satisfies (1.3), $G \subset \mathbb{R}^n$ is an open neighborhood of zero, and $H \in C^{\ell+2}(\mathbb{T}^n \times G)$, $\ell := \gamma + \tau + 1$, is a Hamiltonian function satisfying*

$$H_x(x,0) = 0, \qquad H_y(x,0) = \omega,$$

and

$$|H|_{C^{\ell+2}} \le M, \qquad \left| \left(\int_{\mathbb{T}^n} H_{yy}(x,0)\, dx \right)^{-1} \right| \le M. \tag{4.2}$$

If $u \in C^1(\mathbb{R}^n, \mathbb{R}^n)$ and $v \in C^1(\mathbb{R}^n, G)$ satisfy (1.2), $u(\xi) - \xi$ and $v(\xi)$ are of period 1, $v \circ u^{-1} \in C^\ell(\mathbb{T}^n, G)$, and

$$|u_\xi - 1\!\!1|_{C^0} \le \delta, \qquad \left| v \circ u^{-1} \right|_{C^\ell} \le \delta, \tag{4.3}$$

then $u_\xi \equiv 1\!\!1$ and $v \equiv 0$.

Proof. Let the functions $V(x)$ and $U(x)$ be defined by (1.9) so that (1.12) and (1.13) are satisfied. Then the real numbers $\alpha_\nu := U(x + e_\nu) - U(x)$ are independent of x for $\nu = 1, \dots, n$ and the function $a(x) = U(x) - \langle \alpha, x \rangle$ can be chosen without loss of generality to be of mean value zero over the torus $\mathbb{T}^n$. Moreover, let us define $b(x) := V(x) - x$, $h := H(x, U_x)$, and

$$R_0(x) := \int_0^1 \int_0^t \langle U_x, H_{yy}(x, sU_x) U_x \rangle \, ds dt$$
$$R_1(x) := H_y(x, U_x) - H_y(x,0) - H_{yy}(x,0) U_x + b_x \Big(H_y(x, U_x) - \omega \Big).$$

Then one checks easily that

$$Da = h - \langle \alpha, \omega \rangle - R_0$$
$$Db = -H_{yy}(x,0)(\alpha + a_x) - R_1.$$

This implies that $h - \langle \alpha, \omega \rangle$ is the mean value of R_0 and, moreover,

$$\int_{\mathbb{T}^n} H_{yy}(x,0)\alpha \, dx = -\int_{\mathbb{T}^n} \Big(H_{yy}(x,0)a_x + R_1(x) \Big) \, dx.$$

Therefore, we obtain from Lemma 5 and Lemma 6 that

$$|a_x|_{C^0} \le |a|_{C^{1+\gamma}} \le c_1 \, |R_0|_{C^\ell} \le c_2 \, |U_x|_{C^0} \, |U_x|_{C^\ell}$$

for suitable constants $c_2 > c_1 > 0$, depending only on n, τ, c_0, γ and M. Moreover, it follows from (4.2) that

$$|\alpha| \le M^2 \, |a_x|_{C^0} + M \, |R_1|_{C^0} \le c_3 \Big(|U_x|_{C^\ell} + |b_x|_{C^0} \Big) \, |U_x|_{C^0}$$

for some constant $c_3 > 0$. Finally, we obtain from (4.3) that $|b_x|_{C^0} \le \delta/(1-\delta)$ and $|U_x|_{C^\ell} \le \delta$ so that

$$|U_x|_{C^0} \le |\alpha| + |a_x|_{C^0} \le 2c_3 \Big(|U_x|_{C^\ell} + |b_x|_{C^0} \Big) \, |U_x|_{C^0} \le \frac{4\delta c_3}{1-\delta} \, |U_x|_{C^0} \, .$$

We conclude that $U_x = v \circ u^{-1}$ must vanish if $4\delta c_3 < 1 - \delta$. But this implies $Db = 0$ so that $b(x) \equiv b$ is constant ad hence $u(\xi) = \xi - b$. This completes the proof of Theorem 3. □

We close this section with a regularity theorem for invariant tori which is apparently new.

Theorem 4 (Regularity). *Let $G \subset \mathbb{R}^n$ be an open set and $\omega \in \mathbb{R}^n$ be a vector which satisfies (1.3) for some constants $c_0 > 0$ and $\tau > n - 1$. Let $H \in C^\infty(\mathbb{T}^n \times G)$ be given and suppose that $u \in C^{\ell+1}(\mathbb{R}^n, \mathbb{R}^n)$ and $v \in C^\ell(\mathbb{R}^n, G)$, $\ell > 2\tau + 2$, are solutions of (1.2) such that $u(\xi) - \xi$ and $v(\xi)$ are of period 1 and u represents a diffeomorphism of the torus $\mathbb{T}^n$. If*

$$\det \left(\int_{\mathbb{T}^n} u_\xi(\xi)^{-1} H_{yy}(u(\xi), v(\xi)) u_\xi^T(\xi)^{-1} \, d\xi \right) \ne 0$$

then $u \in C^\infty$ and $v \in C^\infty$.

Proof. Let $z = \phi(\zeta)$ denote the symplectic transformation (1.7) and choose a sequence of C^∞ smooth symplectic transformation ψ^ν of the same form which converges to ϕ in the C^ℓ-norm. Hence the Hamiltonian function $H \circ \psi^\nu$ converges to H in the C^ℓ-norm and hence satisfies the assumptions of Theorem 2 for ν sufficiently large. This implies that, for large ν, there is a symplectic transformation χ^ν of the form (1.7) such that $K^\nu := H \circ \psi^\nu \circ \chi^\nu$ satisfies (1.8). By Corollary 1, we have $\chi^\nu \in C^\infty$. We claim that

$$\phi = \psi^\nu \circ \chi^\nu$$

for ν sufficiently large and hence ϕ is C^∞ smooth.

To see this, let

$$S^\nu(x,\eta) = U^\nu(x) + \langle V^\nu(x), \eta\rangle$$

be a generating function for $\phi^{-1} \circ \psi^\nu$. Then the sequences U_x^ν and $V_x^\nu - \mathbb{1}$ converge to zero in the C^ℓ-norm. Moreover, let

$$\widehat{S}^\nu(x,\eta) = \widehat{U}^\nu(x) + \langle \widehat{V}^\nu(x), \eta\rangle$$

be a generating function for χ^ν. Then Theorem 2 asserts that $\widehat{V}_x^\nu - \mathbb{1}$ converges to zero in the C^0-norm and $\widehat{U}_x^\nu$ converges to zero in the $C^{\tau+1+\gamma}$-norm for some $\gamma > 0$. Now define

$$\widetilde{U}^\nu := U^\nu + \widehat{U}^\nu \circ V^\nu, \qquad \widetilde{V}^\nu := \widehat{V}^\nu \circ V^\nu,$$

so that $\widetilde{S}^\nu(x,\eta) := \widetilde{U}^\nu(x) + \langle \widetilde{V}^\nu(x), \eta\rangle$ is a generating function for $\phi^{-1} \circ \psi^\nu \circ \chi^\nu$. Then $\widetilde{U}_x^\nu$ converges to zero in the $C^{\tau+1+\gamma}$-norm and $\widetilde{V}_x^\nu - \mathbb{1}$ converges to zero in the C^0-norm. Therefore it follows from Theorem 3 with H replaced by $H \circ \phi$ that $\phi^{-1} \circ \psi^\nu \circ \chi^\nu = \mathrm{id}$ for ν sufficiently large. Hence $\phi = \psi^\nu \circ \chi^\nu \in C^\infty$ as claimed. This proves Theorem 4. □

The proof of Theorem 4 depends in an essential way on the fact that the Hamiltonian function H in Theorem 2 is only assumed to be of class C^ℓ with $\ell > 2\tau + 2 > 2n$ and is not required to be close to any integrable analytic function. Another important ingredient in the proof of Theorem 4 is the observation that the smoothness requirements of the uniqueness result (Theorem 3) precisely coincide with the regularity which is obtained in the existence result (Theorem 2). Finally, Theorem 4 suggests that every invariant circle of class C^ℓ with $\ell > 4$ (and with a sufficiently irrational rotation number) for a monotone twist map of class C^∞ must itself be of class C^∞.

References

[1] N.I. Achieser, *Vorlesungen über Approximationstheorie*, Akademie Verlag, Berlin, 1953.

[2] V.I. Arnol'd, Small divisors I, on mappings of a circle into itself, *Isvest. Akad. Nauk, Ser. Math.* **25** (1961), 21–86; *Trans. Amer. Math. Soc.* **46** (1965), 213–284.

[3] V.I. Arnol'd, Proof of a theorem of A.N. Kolmogorov on the invariance of quasiperiodic motions under small perturbations of the Hamiltonian, *Uspehi. Mat. Nauk* **18** (1963), 13–40; *Russian Math. Surveys* **18** (1963), 9–36.

[4] V.I. Arnol'd, Small denominators and problems of stability of motions in classical and celestial mechanics, *Uspehi. Mat. Nauk* **18** (1963), 91–196; *Russian Math. Surveys* **18** (1963), 85–193.

[5] J.-B. Bost, Tores invariants, de systèmes dynamiques Hamiltoniens [d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel,...], Séminaire Bourbaki, *Astérisque* **133-134** (1986), 113–157.

[6] R.S. Hamilton, The inverse function theorem of Nash and Moser, *Bull. Amer. Math. Soc.* **7** (1982), 65–222.

[7] M. Herman, Sur les courbes invariantes par les difféomorphismes de l'anneau I & II, *Astérisque* **103-104** (1983), 1–221, and **144** (1986).

[8] L. Hörmander, On the Nash–Moser implicit function theorem, *Annales Academia Scientiarum Fennicae, Series A.J. Mathematicae* **10** (1985), 255–259.

[9] H. Jacobowitz, Implicit function theorems and isometric embeddings, *Annals of Mathematics* **95** (1972), 191–225.

[10] A.N. Kolmogorov, On the conservation of quasiperiodic motions for a small change in the Hamiltonian function, *Doklad. Acad. Nauk USSR* **98** (1954), 527–530.

[11] A.N. Kolmogorov, The general theory of dynamical systems and classical mechanics, *Proc. Intl. Congress of Math., Amsterdam 1954*, reprinted in *Foundations of Mechanics*, by R. Abraham and J. Marsden, Appendix D, Benjamin, 1967.

[12] J. Mather, Nonexistence of invariant circles, *Ergodic Theory & Dynamical Systems* **4** (1984), 301–309.

[13] J. Moser, A new technique for the construction of solutions of nonlinear differential equations, *Proc. Nat. Acad. Sci. USA* **47** (1961), 1824–1831.

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

[15] J. Moser, A rapidly convergent iteration method and non-linear partial differential equations I & II, *Ann. Scuola Norm. Sup. Pisa* **20** (1966), 265–315, 499–535.

[16] J. Moser, On the construction of almost periodic solutions for ordinary differential equations, *Proceedings of the International Conference on Functional Analysis and Related Topics*, Tokyo, 1969, 60-67.

[17] J. Moser, *Stable and Random Motions in Dynamical Systems*, Annals of Mathematics Studies **77**, Princeton University Press, 1977.

[18] J. Moser, Minimal solutions of variational problems on a torus, *Ann. Inst. Henri Poincaré* **3** (1986), 229–272.

[19] J. Moser, Recent developments in the theory of Hamiltonian systems, *SIAM Review* **28** (1986), 459–485.

[20] J. Moser, A stability theorem for minimal foliations on a torus, *Ergodic Theory & Dynamical Systems* **8** (1988), 251–281.

[21] J. Nash, The imbedding problem for Riemannian manifolds, *Annals of Mathematics* **63** (1956), 20–63.

[22] J. Pöschel, Invariante Tori in differenzierbaren Hamiltonschen Systemen, *Bonner Mathematische Schriften* **120** (1980).

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

[24] H. Rüssmann, Kleine Nenner I: Über invariante Kurven differenzierbarer Abbildungen eines Kreisringes, *Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl.* (1970), 67–105.

[25] H. Rüssmann, On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus, *Dynamical Systems, Theory and Applications*, edited by J. Moser, LNP **38**, Springer Verlag, New York, 1975, pp 598–624.

[26] C.L. Siegel, *Vorlesungen über Himmelsmechanik*, Springer, 1956.

[27] F. Takens, A C^1 counterexample to Moser's theorem, *Indag. Math.* **33** (1971), 379–386.

[28] E. Zehnder, Generalized implicit function theorems with applications to small divisor problems I & II, *Comm. Pure Appl. Math.* **28** (1975), 91–140, and **29** (1976), 49–113.