June 1989
Commun. Math. Phys. **127** (1990) 351–393

Small Divisors with Spatial Structure in Infinite Dimensional Hamiltonian Systems

Jürgen Pöschel

1 Basic concepts and notions

The purpose of this paper is to present a perturbation theory for integrable hamiltonian systems of the Kolmogorov-Arnold-Moser type that comprises the classical result for general perturbations in the finite dimensional case [12,1,18,33,2,20], finite chains of weakly coupled oscillators [36,39], finite dimensional systems with short range interactions [37], systems of infinitely many oscillators with finite range couplings [10,36], and a few more infinite dimensional systems with varying kinds of couplings and localizations. Indeed, our work was initiated and inspired by the progress in this area due to Bellissard, Fröhlich, Spencer, Vittot and Wayne (in alphabetical order) and grew out of an attempt to obtain a unified approach to their results.

The key idea is to consider perturbations not as a single chunk but rather as composites of smaller pieces reflecting an underlying *spatial structure*. The allowable size of these pieces is determined by *weights* associated with their supports. These weights also determine all other quantitative aspects of the theory such as the shape of domains and the small divisor conditions. The validity of those nonresonance conditions is tied to some distribution property of the spatial structure with respect to the weight and cardinality of its components.

Spatial structures are characterized by a single structure property, and weight functions by the properties of monotonicity and subadditivity — see (3) and (4) respectively. These are simple concepts, and switching from one weighted spatial structure to another allows one to study various kinds of perturbations without labouring through the tedious KAM-proof again and again.

Our approach may also give some hint why the KAM-theory fails for certain models. In one way or another such a failure is tied to the failure of the small divisors to obey the growth conditions imposed on them in terms of approximation functions. In a most remarkable achievement Yoccoz [41] was able to show that such growth conditions are indeed necessary in the simplest of these stability problems, the Siegel center problem in the complex plane [33].

Our approach also fails in another way, and naturally so. The perturbations are required to be "sufficiently localized" allowing for "sufficiently localized" nonresonant invariant tori of *maximal* dimension. It apparently does *not* apply to nonlinear partial differential equations such as the nonlinear wave equation, where there is no such localization at all. A different approach is appropriate here, aiming to find invariant tori of *finite* dimension in infinite dimensional systems. This crucially reduces the restrictions posed on the small divisors. We refer to [13,40,27] for details.

Our point of departure is a collection of an arbitrary number of harmonic oscillators occupying the sites λ of some lattice Λ or a subset thereof. The shape, size or dimension of this lattice are of no concern.

The configuration of an individual oscillator is described by a single pair of angle-action coordinates $\varphi_\lambda, I_\lambda$ for ease of notation. Its motion is described by a single frequency ω_λ. The hamiltonian of such a system is

$$N = e + \sum_{\lambda \in \Lambda} \omega_\lambda I_\lambda = e + \langle \omega, I \rangle ,$$

and its equations of motion are

$$\begin{aligned} \dot{\varphi} &= \omega \\ \dot{I} &= 0 \end{aligned}$$

in usual vector notation. The underlying phase space is

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

where $\mathbb{T}$ denotes the standard one-torus obtained from the real line by identifying points modulo 2π.

As a further simplification the frequencies ω are regarded as parameters varying freely over some subset $\mathcal{O}$ of the parameter space $\mathbb{R}^\Lambda$. This is tantamount to imposing a "nondegeneracy" or "anisochronicity" condition upon the unperturbed system, and given such a condition those frequencies may always be introduced as parameters. This has the advantage that it suffices to consider hamiltonians N that are just linear in I.

We are going to study hamiltonians that are — in an appropriate sense — small perturbations of the integrable hamiltonian N. Our aim is to prove the persistence of the invariant torus

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

of maximal dimension together with its constant vectorfield ω.

The crucial assumption is that the perturbation decomposes into a series of smaller pieces which involve only finitely many lattice sites each. Precisely, we consider hamiltonians of the form

$$H = N + P, \qquad P = \sum_{A \in \mathcal{S}} P_A,$$

where $\mathcal{S}$ is a family of finite subsets A of Λ on which the individual perturbations P_A "live". That is to say, P_A does not depend on the configuration of any oscillator outside of A.

This family $\mathcal{S}$ is not totally arbitrary. Rather, $\mathcal{S}$ has to be a *spatial structure* on Λ characterized by the property that the union of any two sets in $\mathcal{S}$ is again in $\mathcal{S}$, if they intersect:

$$A, B \in \mathcal{S}, \quad A \cap B \neq \emptyset \quad \Rightarrow \quad A \cup B \in \mathcal{S}. \tag{3}$$

This property is necessary and sufficient for the spatial structure to be preserved under Poisson brackets. Of course, not all terms in the given spatial expansion of the perturbation P need to be present.

The main ingredient of our perturbation theory is a nonnegative *weight function*

$$[\,\cdot\,]: \quad A \mapsto [A]$$

defined on $\mathcal{S} \cap \mathcal{S} = \{ A \cap B : A, B \in \mathcal{S} \}$. The weight of a subset may reflect its size, its location or something else. This, however, is immaterial for the perturbation theory itself. Here only the properties of monotonicity and subadditivity are required:

$$\begin{aligned} A \subseteq B \quad &\Rightarrow \quad [A] \leq [B] \\ A \cap B \neq \emptyset \quad &\Rightarrow \quad [A \cup B] + [A \cap B] \leq [A] + [B] \end{aligned} \tag{4}$$

for all A, B in $\mathcal{S}$. All other quantitative aspects are expressed in terms of this weight function.

In a crucial fashion the weight function determines the nonresonance conditions for the small divisors arising in this theory. They are the usual ones, namely

$$\langle k, \omega \rangle = \sum_{\lambda \in \Lambda} k_\lambda \omega_\lambda ,$$

where due to the spatial structure of the perturbation k runs over all nonzero integer vectors in $\mathbb{Z}^\Lambda$ whose support

$$\operatorname{supp} k = \{ \lambda : k_\lambda \neq 0 \}$$

is a *finite* set.

Requiring the components of P to decay rapidly — as we will do later on by way of an appropriate norm — it suffices to estimate these small divisors from below not only in terms of the *norm* of k,

$$|k| = \sum_{\lambda \in \Lambda} |k_\lambda| ,$$

but also in terms of the *weight of its support*,

$$[\![k]\!] = \min_{\operatorname{supp} k \subseteq A \in \mathcal{S}} [A] .$$

Then the nonresonance conditions read

$$|\langle k, \omega \rangle| \geq \frac{\alpha}{\Delta([\![k]\!]) \Delta(|k|)}, \qquad 0 \neq k \in \mathbb{Z}^\Lambda, \tag{5}$$

where, as usual, α is a positive parameter and Δ some fixed approximation function as described in Appendix A. One and the same approximation function is taken here in both places for simplicity, since the generalization is straightforward. Note that the right hand side of (5) is zero when $\operatorname{supp} k$ is infinite, so this case need not be excluded explicitly.

On an infinite dimensional lattice it now depends on the chosen weight function whether these nonresonance conditions can be met by some frequency vector ω or not. If the weights are to “light”, then there are none. On the other hand, if they are too “heavy”, then the components P_A have to decay very rapidly with $[A]$. The point is to strike a good balance between these two extremes.

As an illustration, let

$$\Lambda = \mathbb{Z}^d, \qquad d \geq 1,$$

the d-dimensional integer lattice. Let $\mathcal{S}$ be the spatial structure generated by the nearest neighbour sets $A_i = \left\{ j : |i - j|_\infty \leq 1 \right\}$ with $i \in \Lambda$. A useful weight function is given by

$$[A] = \sum_{i \in A} |i| ,$$

since it reflects both the size and the location of A. Other choices such as

$$[A] = \max_{i \in A} |i| , \qquad [A] = \operatorname{card} A$$

are too "light", and the small divisor conditions can not be met.

2 The Result

Let Λ be a lattice with a weighted spatial structure $\mathcal{S}$. Let

$$N = e + \langle \omega, I \rangle$$

be the unperturbed, integrable hamiltonian with frequencies ω taken from a parameter domain $\mathcal{O}$ in $\mathbb{R}^{\Lambda}$. The nature of this set is quite irrelevant for our purposes. It suffices to assume that — after fixing some approximation function Δ — for some $\alpha > 0$ there is a nonempty subset

$$\mathcal{O}_\alpha \subseteq \mathcal{O} \subseteq \mathbb{R}^{\Lambda}$$

of strongly nonresonant frequencies in the sense of (5). Indeed, the set $\mathcal{O}$ may consist of a single strongly nonresonant frequency vector.

We consider perturbations $H = N + P$ that are real analytic in the phase space variables θ, I on a complex neighbourhood

$$\mathcal{D}_{r,s}: \quad |\operatorname{Im} \theta|_\infty < r, \quad \| I \|_w < s$$

of the torus $\mathcal{T}_0$ and real analytic in the parameter ω on a complex neighbourhood

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

of the real parameter set $\mathcal{O}$. The norms are

$$|\theta|_\infty = \sup_{\lambda \in \Lambda} |\theta_\lambda| , \qquad \| I \|_w = \sum_{\lambda \in \Lambda} |I_\lambda| \, e^{w[\lambda]},$$

where $w \geq 0$ is another parameter, and the weights at the individual lattice sites are defined by

$$[\lambda] = \min_{\lambda \in A \in \mathcal{S} \cap \mathcal{S}} [A] .$$

It is important to take the minimum over the family $\mathcal{S} \cap \mathcal{S}$ in order that the estimates (21) and (22) come out right.

The perturbation itself is supposed to be given as a spatial series

$$P = \sum_{A \in \mathcal{S}} P_A(\theta_A, I_A; \omega_A), \tag{6}$$

where $\theta_A = (\theta_\lambda : \lambda \in A)$ and similarly I_A and ω_A. Its size is measured in terms of the weighted norm

$$|||P|||_{m,r,s,h} = \sum_{A \in \mathcal{S}} \|P_A\|_{r,s,h} e^{m[A]},$$

where

$$\|P_A\|_{r,s,h} = \sum_{k \in \mathbb{Z}^A} \left|P_{A,k}\right|_{s,h} e^{r|k|}.$$

This definition refers to the Fourier series expansion $P_A = \sum_k P_{A,k} e^{i\langle k,\theta\rangle}$ whose coefficients depend on I and ω. The norm $|\cdot|_{s,h}$ is the sup-norm over $|I|_w < s$ and $\mathcal{W}_h$. The triple-bar-norm reflects the idea of treating Fourier and spatial expansions on exactly the same footing.

The smallness condition of the following theorem is expressed in terms of two functions Ψ_0, Ψ_1 that are defined on the positive real axis entirely in terms of the approximation function Δ and reflect the effect of the small divisors in solving the nonlinear problem. See Appendix A for their definition.

Theorem A. *Let Λ be a lattice with a weighted spatial structure $\mathcal{S}$. Suppose that P admits a spatial expansion as in (6), is real analytic on $\mathcal{D}_{r,s} \times \mathcal{W}_h$ and satisfies the estimate*

$$s^{-1} |||P|||_{m,r,s,h} \le \frac{\alpha \varepsilon_*}{\Psi_0(\mu)\Psi_1(\rho)} \le \frac{h}{2^5}$$

for some $0 < \mu \le m - w$ and $0 < \rho < r/2$, where ε_* is an absolute positive constant. Then there exists a transformation

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

that is real analytic and symplectic for each ω and uniformly continuous in ω, such *that*

$$(N+P)\circ\mathcal{F}=\mathrm{const}+\langle\omega,I\rangle+\dots,$$

where the dots denote terms of higher order in I. Consequently, the perturbed system has a real analytic invariant torus of maximal dimension and with a vectorfield conjugate to ω for each frequency vector ω in $\mathcal{O}_\alpha$. These tori are close of order $s^{-1}|||P|||$ to the torus $\mathcal{T}_0$ with respect to the norm $|\cdot|_w$.

Our proof yields $\varepsilon_*=2^{-22}$, but no effort was undertaken to obtain an "optimal" constant.

The next theorem gives a criterion for the existence of strongly nonresonant frequencies. It is based on growth conditions on the *distribution function*

$$N_n(t)=\mathrm{card}\left\{A\in\mathcal{S}:|A|=n,\ [A]\le t\right\}$$

for $n\ge 1$ and $t\ge 0$.

Recall the definition that a point belongs to the support of a measure μ if $\mu(O)>0$ for every open neighbourhood O of this point. The topology on the parameter space $\mathbb{R}^\Lambda$ is the topology of uniform convergence.

Theorem B. *Suppose there exists a constant N_0 and an approximation function Φ such that*

$$N_n(t)\le\begin{cases}0, & t<t_n\\ N_0\Phi(t), & t\ge t_n\end{cases}$$

with a sequence of real numbers t_n satisfying

$$t_n\ge n\log^\sigma n$$

for n large with some exponent $\sigma>1$. Then there exists an approximation function Δ and a probability measure μ on the parameter space $\mathbb{R}^\Lambda$ with support at any prescribed point so that

$$\mu\left(\mathbb{R}^\Lambda-\mathbb{R}^\Lambda_\alpha\right)=O(\alpha).$$

It follows that $\mathcal{O}_\alpha$ is not empty for sufficiently small α whenever the set $\mathcal{O}$ contains an interior point.

The hypotheses of this theorem, however, is admittedly somewhat awkward and abstract. Here is a more handy criterion for the important special case where Λ is the d-dimensional integer lattice $\mathbb{Z}^d$ or a subset thereof.

Nonresonance Criterion. *Suppose* $\Lambda \subseteq \mathbb{Z}^d$, *and every set in* $\mathcal{S}$ *is connected. If there exists a constant* $\sigma > 1$ *such that*

$$[A] \geq |A| \log^{\sigma} |A| \qquad \textit{for } |A| \textit{ large}$$

$$|A|_\infty \leq \exp\left(\frac{[A]}{\log^{\sigma}[A]}\right) \qquad \textit{for } [A] \textit{ large},$$

where $|A|_\infty = \max_{i \in A} |i|_\infty$, *then the conclusion of Theorem B holds.*

The proof of this criterion is short. By the first hypotheses,

$$N_n(t) = 0 \qquad \text{for } t < n \log^{\sigma} n,\ n \geq n_0,$$

for n_0 sufficiently large. Letting

$$t_n = \begin{cases} 0, & n < n_0 \\ n \log^{\sigma} n, & n \geq n_0 \end{cases}$$

we then have $N_n(t) = 0$ for $t < t_n$ and all $n \geq 1$.

Now let $A \in \mathcal{S}$ with $|A| = n$ and $[A] \leq t$, where $t \geq t_0$ is sufficiently large. By the second hypotheses, A is contained in the ball B_r of radius

$$r = e^{t/\log^{\sigma} t}$$

around the origin. The number of lattice points in this ball is $|B_r| \leq (2r + 1)^d$. Furthermore, the number of all connected sets of cardinality n containing a given point is smaller than the number of all paths of length $2n$ starting from the same point. This number is bounded by $(2d)^{2n}$. Hence,

$$\begin{aligned} N_n(t) &\leq (2d)^{2n} |B_r| \\ &\leq 3^d (2d)^{2n} r^d \\ &\leq N_0 D^n \Theta^d(t) \end{aligned}$$

for $t \geq t_0$ with

$$\Theta(t) = \exp\left(\frac{t}{1 + \log^{\sigma}(1 + t)}\right).$$

This holds for all $n \geq 1$ with constants N_0 and D depending only on the dimension d.

Multiplying N_0 by $\Theta^d(t_0)$ this estimate holds also for $0 \leq t \leq t_0$ by the monotonicity of the left hand side. Finally, D^n is bounded by a constant multiple of

$\Theta(t_n)$ for all n, and since N_n vanishes for $t < t_n$ anyhow, we may replace the latter by $\Theta(t)$. Thus,

$$N_n(t) \le N_0 \Theta^{d+1}(t), \qquad t \ge t_n,$$

for all $n \ge 1$ with a different constant N_0. Now Theorem B applies, and the criterion is proven.

3 Examples

Finite range couplings

Consider an infinite number of harmonic oscillators occupying the sites of an integer lattice $\Lambda = \mathbb{Z}^d$ with $d \ge 1$. Their frequencies are assumed to be independent, identically distributed random variables such that they may be regarded as parameters varying over some open domain $\mathcal{O}$ in the space $\mathbb{R}^\Lambda$ endowed with the topology of uniform convergence. The unperturbed hamiltonian thus reads

$$N = \sum_{i \in \mathbb{Z}^d} \omega_i I_i = \langle \omega, I \rangle .$$

We are going to study uniform finite range perturbations of this system: each oscillator is coupled to a finite number of neighbours, and the coupling law is the same throughout the lattice. Such systems arise as models of large arrays of weakly coupled “bedsprings”, or surface layers of atoms deposited on a disordered crystalline surface. The reader may refer to the introductory section of [10] for more about the physical background.

Let us first consider *nearest neighbour coupling*: each oscillator is coupled to its immediate neighbours through some unharmonic force. The hamiltonian of such a system is

$$H = \langle \omega, I \rangle + \sum_i P_{A_i}, \qquad A_i = \left\{ j : |j - i|_\infty \le 1 \right\} \tag{7}$$

with

$$P_{A_i} = O\left(\left| I_{A_i} \right|_\infty^\lambda \right)$$

uniformly in i with an exponent $\lambda > 1$ to made precise later. The perturbing terms are assumed to be real analytic on uniform θ, I and ω domains.

Of course, $\mathbb{T}^\Lambda \times \{0\}$ is an invariant torus of this system, but the point is to find nontrivial ones. Such tori were first constructed by Fröhlich, Spencer and Wayne [10] by imposing a very strong localization condition, namely

$$I_i^o \sim s e^{-|i|^{d+\delta}}, \qquad \delta > 0,$$

with s sufficiently small. A similar result was found independently by Vittot and Bellissard [36].

We are going to improve these results. To begin with it is convenient to normalize s to some fixed value, say $s = 1$, by stretching I by this amount and dividing the resulting hamiltonian by s. This preserves the symplectic structure and gives the new hamiltonian

$$H = \langle \omega, I \rangle + \varepsilon \sum_i P_{A_i}$$

with $\varepsilon = s^{\lambda - 1}$. In the following, ε will be chosen small, and this translates back into a smallness condition on s.

Let $\mathcal{S}$ be the spatial structure "generated" by the nearest neighbour sets A_i. That is, $\mathcal{S}$ is the intersection of all spatial structures containing those sets. Let $[\,\cdot\,]$ be any weight function satisfying the hypotheses of the nonresonance criterion. Finally, let $w > 0$, and recall the definitions

$$|I|_w = \sum_i |I_i|\, e^{w[i]}, \qquad [i] = \min_{i \in A \in \mathcal{S} \cap \mathcal{S}} [A].$$

As it happens, $[i]$ equals the weight of the corresponding one-element-set, since $\mathcal{S} \cap \mathcal{S}$ contains all one-point-sets.

Pick any initial position I^o with $|I^o|_w = 1$ and expand H in a ball of radius 1 around it. By assumption,

$$\left\| P_{A_i} \right\|_{r,1} \le C \max_{j \in A_i} e^{-\lambda w[j]}$$

uniformly in i for some $r > 0$ and therefore

$$\begin{aligned} |||P|||_{m,r,1} &= \sum_i \left\| P_{A_i} \right\|_{r,1} e^{m[A_i]} \\ &\le C \sum_i \max_{j \in A_i} e^{m[A_i] - \lambda w[j]}. \end{aligned}$$

Now, if

$$\lambda > \lambda_* = \operatorname*{ess\,sup}_{j \in A_i} \frac{[A_i]}{[j]} \ge 1, \tag{8}$$

then there are $m > w$ and $\delta > 0$ such that $\lambda w > (m+\delta)\lambda_*$ and hence $\lambda w\,[j] \geq (m+\delta)\,[A_i]$ for almost all i and all $j \in A_i$. Consequently,

$$|||P|||_{m,r,1} \leq C + C\sum_i e^{-\delta[A_i]}$$

for some $m > w$ and some $\delta > 0$. The infinite sum converges, since

$$[A_i] \geq \log|i|\log^\sigma \log|i|$$

for large $|i|$ by the hypotheses of the nonresonance criterion.

Thus, if $\lambda > \lambda_*$, then the KAM-theorem applies for sufficiently small ε. For every frequency vector ω in a subset of $\mathcal{O}$ of positive measure, there exists a real analytic invariant torus which is localized like $I_i^o \sim e^{-w[i]}$. The measure of the set of “bad” frequencies in $\mathcal{O}$ is of order ε with respect to a large class of probability measures as described in Lemma 1.

Here are now various choices of weight functions. The example of [10] is recovered by choosing

$$[A] = \max_{i\in A}|i|^{d+\delta}, \qquad \delta > 0$$

where $|\cdot| = |\cdot|_\infty$. One obtains $\lambda_* = 1$ and $I_i^o \sim e^{-|i|^{d+\delta}}$ with $w = 1$. Note that for $\delta \leq 0$ the nonresonance criterion does not apply.

To get around this hyperexponential localization one may choose

$$[A] = \sum_{i\in A}|i|,$$

for which the nonresonance criterion is checked in a moment. Then $I_i^o \sim e^{-|i|}$ at the expense of having $\lambda_* = 2d+1$. But this may be further improved on by taking

$$[A] = \sum_{|i-A|\leq t}|i|$$

with some $t \geq 0$. This indeed defines a weight function for all $t \geq 0$, as one easily verifies. As the “thickness” t increases, the limit exponent λ_* goes down to 1, while still $I_i^o \sim e^{-|i|}$ by choosing the parameter w in such a way that $w\,[i] \sim |i|$ for large $|i|$.

Finally, one may indeed take

$$[A] = \sum_{|i-A|\leq t}\log^\gamma(1+|i|), \qquad \gamma > 1.$$

Now the tori are localized like $I_i^o \sim e^{-\log^\gamma |i|}$, while again $\lambda_* \downarrow 1$ as t increases. — A somewhat related result was obtained by Vittot [35].

Obviously, the preceding example is not limited to nearest neighbour sets. Everything works the same if they are replaced by connected neighbouring sets of arbitrary shape and size as long as the latter is uniformly bounded. On the other hand, our theory does *not* seem to encompass *short range couplings*. These are perturbations of the form

$$P = \sum_i \sum_{l=1}^{\infty} P_{A_{il}}, \qquad A_{il} = \{ j : |j - i| \le l \}$$

with

$$P_{A_{il}} = O\Big(|I_{A_{il}}|_\infty^\lambda \, e^{-\mu l} \Big) .$$

Estimating as above, one is lead to require that

$$\min_{j \in A_{il}} \lambda w[j] + \mu l > m[A_{il}]$$

for almost all i and l with fixed numbers $m > w \ge 0$. But this prevents the weight function to meet the first requirement of the nonresonance criterion, since $|A_{il}| \sim l^d$, unless λ is allowed to depend on l.

It remains to verify the applicability of the nonresonance criterion to our various choices of weight functions. It suffices to do this for the "lightest" example,

$$[A] = \sum_{i \in A} \log^\gamma (1 + |i|), \qquad \gamma > 1,$$

since any "heavier" weight function satisfies its hypotheses *a fortiori*.

First, let $B_n = \{ i : |i| \le n \}$. Then

$$[B_n] \sim \sum_{k=1}^{n} k^{d-1} \log^\gamma (1 + k) \sim n^d \log^\gamma n \sim |B_n| \log^\gamma |B_n| ,$$

where the tilde means that either side is bounded by a constant multiple of the other side independently of n. Among all sets of the same cardinality, B_n has the lowest weight, so if $|B_{n+1}| > |A| \ge |B_n|$, then

$$[A] \ge [B_n] \sim [B_n] \log^\gamma [B_n] \sim |A| \log^\gamma |A| .$$

Secondly, let $i \in A$. Since $t / \log^\sigma t$ is eventually monotonically increasing, we have

$$\exp\left(\frac{[A]}{\log^{\sigma}[A]}\right) \geq \exp\left(\frac{[i]}{\log^{\sigma}[i]}\right) \geq \exp\left(\frac{\log^{\gamma}|i|}{\log^{\sigma}\log^{\gamma}|i|}\right) \geq |i|$$

for $|i|$ sufficiently large and hence $[A]$ sufficiently large no matter how $\sigma > 1$ is chosen. It follows that the nonresonance criterion applies.

Arbitrary couplings

The spatial structure of short range couplings consists of connected sets only. For comparison, consider now the case of *arbitrary couplings*, where $\mathcal{S}$ consists of *all finite subsets* of the lattice $\Lambda = \mathbb{Z}^d$. We claim that here the weight function with

$$[A] = 1 + \sum_{i \in A} \log^{\gamma}(1 + |i|), \qquad \gamma > 2$$

satisfies the hypotheses of Theorem B. A related observation was made in [36].

For the proof, let again $B_m = \{ i : |i| \leq m \}$. As in the preceding example, $[B_m] \sim |B_m| \log^{\gamma} |B_m|$, and B_m has the lowest weight among all sets with the same number of elements. It follows that

$$N_n(t) = 0 \qquad \text{for} \quad t \leq t_n \sim n \log^{\gamma} n\,.$$

Next, let $t \geq 0$ be arbitrary, and consider the collection of all sets A with n elements and weight not bigger than t. Picking any element from A with weight $0 \leq t - s \leq t$, the remaining $n - 1$ elements have total weight not bigger than s. This leads to the estimate

$$N_n(t) \leq \frac{1}{n} \int_0^t W(t-s)\, dN_{n-1}(s),$$

where W is any continuous function bounding N_1 from above. Integrating by parts and assuming that $W(0) = 0$ the role of W and N_{n-1} can be interchanged. And proceeding by induction, we obtain

$$N_n(t) \leq \frac{1}{n!} \int\limits_{t_1 + \cdots + t_n \leq t} dW(t_1) \cdots dW(t_n).$$

Now let $d = 1$ for simplicity and choose $W(t) = we^{t^{\mu}} - w$ with $\mu = \gamma^{-1}$ and a suitable constant $w \geq 1$. Then

$$N_n(t) \leq \frac{w^n}{n!} \int\limits_{t_1 + \cdots + t_n \leq t} \exp(t_1^{\mu} + \cdots + t_n^{\mu})\, dt_1^{\mu} \ldots dt_n^{\mu}.$$

On the domain of integration, $t_1^\mu + \cdots + t_n^\mu \le n^{1-\mu}(t_1 + \cdots + t_n)^\mu \le n^{1-\mu} t^\mu$, while the integral of $dt_1^\mu \ldots dt_n^\mu$ over $[0, t]^n$ is bounded by $t^{\mu n}$. Hence,

$$N_n(t) \le \frac{w^n t^{\mu n}}{n!} \exp(n^{1-\mu} t^\mu) \le \exp(w t^\mu) \exp(n^{1-\mu} t^\mu).$$

Finally, to eliminate the dependence on n for large n, recall that $t \ge t_n \sim n \log^\gamma n$. Hence, $n \sim t_n / \log^\gamma t_n \le t / \log^\gamma t$ for $t \ge t_n$ and so

$$n^{1-\mu} t^\mu \lesssim \frac{t}{\log^\sigma t}, \qquad \sigma = \gamma(1-\mu) = \gamma - 1 > 1.$$

This shows that $N_n(t)$ is bounded from above by a constant multiple of a fixed approximation function independently of n as required by Theorem B.

This result applies for example to the hamiltonian

$$H = \langle \omega, I \rangle + \varepsilon \sum_{i,j \in \mathbb{Z}^d} P_{ij}(I) \cos(\theta_i - \theta_j),$$

where more precisely P_{ij} depends only on I_i and I_j. If

$$\sup_{|I|_w < 1} \left| P_{ij}(I) \right| \le c e^{-m \log^\gamma |i| - m \log^\gamma |j|}$$

with $m > w \ge 0$ and $\gamma > 2$, then Theorem A applies for sufficiently small ε, and there exist invariant tori localized like $I_i^o \sim e^{-\log^\gamma |i|}$ [35].

Hierarchical systems

Another interesting class of examples is provided by *hierarchical systems*, as was pointed out to the author by Jürg Fröhlich. At the lowest level they consist of many unrelated small scale systems. At the next level a weak force couples a few of them at a time leading to a collection of unrelated systems at a somewhat larger scale. Then a weaker force couples a few of those at a time, and so on. The universe with its hierarchy of solar systems, star clusters, galaxies, clusters of galaxies, and so on may serve as a model for this kind of system.

In terms of spatial structures a hierarchical system $\mathcal{S}$ is characterized by the *hierarchical property*:

$$A \cap B \neq \emptyset \quad \Rightarrow \quad A \subseteq B \quad \text{or} \quad B \subseteq A$$

for all A, B in $\mathcal{S}$. This property has the very pleasant effect that *any* nonnegative, monotone function on $\mathcal{S} \cap \mathcal{S}$ is a weight function on $\mathcal{S}$, since automatically

$$[A \cup B] + [A \cap B] = [A] + [B]$$

whenever A and B intersect. Thus, the class of weight functions is considerably enlarged.

For example, let Λ be a countable collection of points for which a distance function is defined. Let $[A]$ be zero, if A has no proper components, otherwise let $[A]$ be the minimal distance between the maximal components which A consists of. This is a *bona fide* weight function, if the diameter of each set A in the structure is smaller than its distance to any other disjoint set in the structure. Moreover, any nonnegative, monotone increasing function of $[A]$ is also a weight function. It then depends on how thinly this universe is populated for which functions of $[A]$ the hypotheses of Theorem B can be verified, and this in turn determines the admissable strength of the coupling forces as a function of $[A]$. But for lack of a genuine application we will not expand this further.

Finite chains of oscillators

The theory of spatial structures is also helpful in studying hamiltonian systems with a finite, but large number of degrees of freedom. The point of interest is the dependence of the smallness condition of the classical KAM-theorem on that number. Rigorous results in this direction are due to Vittot [35] and Wayne [37]. There are also quite a number of numerical studies of that question which are referenced in their papers.

To have a specific example in mind, consider a chain of N identical, weakly coupled oscillators with hamiltonian

$$H = \frac{1}{2} \sum_{i=1}^{N} I_i^2 + \varepsilon \sum_{i=1}^{N-1} \cos(\theta_i - \theta_{i+1}).$$

Wayne showed, among others, that it suffices to choose $\varepsilon \sim N^{-a}$ to obtain invariant tori, with an exponent a in the hundreds, whereas the general theory requires ε to depend *exponentially* on N. The results of Vittot are similar but less comparable, since he uses different, stronger norms. We are going to recover and improve Wayne's result.

Let

$$\Lambda = [1, N] \subset \mathbb{Z},$$

the integer interval from 1 to N, and let $\mathcal{J}$ be the spatial structure consisting of all subintervals A of Λ containing at least two points. As usual, we consider nearly integrable hamiltonians of the form

$$H = \langle \omega, I \rangle + \sum_{A \in \mathcal{J}} P_A,$$

which depend on the frequencies ω as parameters varying over a domain $\mathcal{O}$ whose size is assumed to be independent of N. The hamiltonian above is brought into this form by expanding H in a ball $|I|_1 < 1$ around every initial position I^o in $[1, 2]^N$, say, and having $\omega(I^o) = I^o$. In this case, $\mathcal{O} = [1, 2]^N$.

The point is a matching choice of the approximation and weight functions to determine those α as functions of N for which the set of "good" frequencies is not empty. Choosing the weight function

$$[A] = f\,|A| - f, \qquad f > 2e - 1$$

and the approximation function

$$\Delta(t) = \frac{D(t)}{1+t}, \qquad D(t) = \left(1 + \frac{t}{N+1}\right)^{N+1},$$

one roughly needs (this estimate is not optimal)

$$\alpha \sim \frac{1}{N \log N}$$

to obtain nonresonant frequencies in $\mathcal{O}$ filling a set of positive measure. The details are given in Appenix E.

There is, however, a catch. The Ψ-functions now depend exponentially on N, and the only way to beat this is by having the parameters μ and ρ sufficiently *large*. More precisely, the function Ψ_1 for Δ is the same as the function Ψ_0 for D, and

$$\Psi_0(l+1) \le \left(\frac{4}{\kappa - 1}\right)^{(N+1)/\kappa^l} \le 2^{4N/\kappa^l}$$

for integers l by the remark following Lemma 7. Hence there is a uniform bound if $l \ge \log N / \log \kappa$. Accordingly, if μ and ρ are integers greater or equal $1 + \log N / \log \kappa$, then the Ψ-functions are independent of N.

We thus arrive at the following result. If the weight function $[\,\cdot\,]$ is chosen as above and

$$s^{-1}|||P|||_{\xi+2,2\xi+4,s,h} \le \frac{\varepsilon_0}{N \log N}, \qquad \xi = \frac{\log N}{\log \kappa},$$

where ε_0 is independent of N, then there exist real analytic invariant tori for frequencies ω in a set of positive measure in $\mathcal{O}$. In particular, for large N one roughly needs

$$s^{-1}\|P_A\|_{2\xi+4,s,h} \sim e^{-\xi[A]} \sim N^{-\zeta|A|}, \qquad \zeta = f/\log \kappa.$$

This recaptures Wayne's result in [37] about finite dimensional short range interactions.

We apply this result to the chain of weakly coupled oscillators. For $A = [i, i+1]$,

$$P_A = \tfrac{1}{2} I_i^2 + \varepsilon \cos(\theta_i - \theta_{i+1}), \qquad \|P_A\|_{r,\sqrt{\varepsilon}} \le \varepsilon + \varepsilon e^{2r}.$$

Since $|A| = 2$ for all terms in the given perturbation,

$$\begin{aligned} \frac{1}{\sqrt{\varepsilon}}|||P|||_{\xi+2,2\xi+4,\sqrt{\varepsilon}} &= \sqrt{\varepsilon} \sum_{|A|=2} \|P_A\|_{2\xi+4,\sqrt{\varepsilon}}\, e^{(\xi+2)[A]} \\ &\le C N^2 e^{(4+f)\xi} \sqrt{\varepsilon} \\ &= C N^{2+(4+f)/\log\kappa} \sqrt{\varepsilon}. \end{aligned}$$

Thus, one roughly needs $\varepsilon \sim N^{-48}$ to apply KAM.

Incidentally, by a slight modification of the general KAM-scheme it suffices to have $0 < \rho < r/b$ for any fixed $b > 1$, whence the '$2\xi + 4$' may be replaced by '$b\xi + 3$' in the preceding statements. This reduces the power of N to 38, but still this is certainly not optimal. In fact, numerical experiments seem to indicate that there is almost no dependence on N at all. Those observations, however, may also be due to the extreme slowness of the process of Arnold diffusion in nearly integrable systems.

The classical KAM-theorem

The theory of small divisors with spatial structure is a natural extension of the classical theory of Kolmogorov, Arnold and Moser, and the latter is recovered simply by having no structure at all.

More precisely, let

$$\Lambda = [1, n] \subset \mathbb{Z},$$

and let $\mathcal{S}$ consist of Λ alone having zero weight. Then

$$\mathcal{D}_{r,s}: \quad |\mathrm{Im}\,\theta|_\infty < r, \quad |I|_1 < s,$$

and the weighted norm of a real analytic perturbation $P = P_\Lambda$ on $\mathcal{D}_{r,s}$ reduces to

$$|||P|||_{m,r,s,h,} = \|P\|_{r,s,h} = \sum_{k\in\mathbb{Z}^n} |P_k|_{s,h}\, e^{r|k|}$$

for all m. Moreover, by standard estimates for the Fourier coefficients of analytic functions, $|P_k| \le e^{-r|k|}\,|P|_r$, where the latter stands for the supremum of P over $|\mathrm{Im}\,\theta|_\infty < r$. It follows that

$$\|P\|_{r,s,h} \le \sum_{k\in\mathbb{Z}^n} e^{-2\rho|k|}\,|P|_{r+2\rho,s,h} = \coth^n\rho\ |P|_{r+2\rho,s,h}\,,$$

whereas the estimate $|P|_{r,s,h} \le \|P\|_{r,s,h}$ is trivial.

Finally, we choose the approximation function

$$\Delta(t) = (1+t)^{-1}(1+t/\tau)^\tau, \qquad \tau > n$$

and observe that $[\![k]\!] = 0$ for all k, hence this term does not matter in the small divisor conditions. By the estimates in Appendix E,

$$\alpha^{-1}\mu(\mathbb{R}^n - \mathbb{R}^n_\alpha) \le \frac{c_o^n}{\min(1, \tau - n)}$$

for the standard gaussian probability measures on $\mathbb{R}^n$ described in Section 8 with a universal constant c_o, and

$$\Psi(\rho) \le \left(\frac{8}{\rho}\right)^\tau, \qquad 0 < \rho \le 1,$$

by Lemma 3 for $\Psi = \Psi_1$.

Summarizing we obtain the following version of the

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

$$s^{-1}\|H - N\|_{r,s,h} \le \alpha\varepsilon_*\rho^\tau \le \frac{h}{2^5}$$

for some $0 < \rho < r/2$, where ε_* is an absolute positive constant and $\tau > n$. Then there exists a real analytic invariant torus with a vectorfield conjugate to ω for every *frequency vector* ω *in* $\mathcal{O}_\alpha$*, which is close of order* $s^{-1}\|H - N\|$ *to the unperturbed torus. Moreover,*

$$\mu(\mathcal{O}_\alpha - \mathcal{O}) = O(\alpha)$$

with respect to any standard gaussian probability measure μ *on* $\mathbb{R}^n$*.*

Note that the small divisors enter only once in the smallness condition via the expression $\alpha\rho^\tau$, which is different from other versions of the KAM-theorem where this term is squared. This is due to regarding the frequencies as independent parameters thereby decoupling them from the hamiltonian system itself. More technically speaking, in the linearized problem there is only one small divisor equation to be solved. Of course, this is not a genuine improvement, since the square is restored when reducing the traditional versions of this theorem to the one above.

4 Other Applications and Extensions

The theory of spatial structures presented here offers a general mechanism for keeping track of the interaction of couplings of varying strengths and locations in nearly integrable hamiltonian systems. Its pivotal ingredient is the effective control of the Poisson bracket of such hamiltonians. Therefore, this theory is not limited to extending KAM-theorems to certain infinite dimensional systems. It also helps to simplify and improve Nekhoroshev type estimates such as in [39] and more generally any construction of normal forms up to a finite order provided the system exhibits some sort of spatial structure. In these applications only a finite number of coordinate changes are performed. This gives the freedom of choosing exponential functions as approximation functions, which further simplifies the estimates. Likely, even relatively small systems with oscillations of different time scales may be analyzed this way.

In this paper we chose to describe the theory in its simplest form. In particular, on the ω-parameter space we chose the topology of uniform convergence. But for many applications it is necessary to have an exponentially localized topology of the same kind as that for the action variables, given by the norm

$$|\omega|_v = \sup_{\lambda\in\Lambda} |\omega_\lambda|\, e^{v[\lambda]}$$

with a parameter $v > 0$. This requires some modifications of the small divisor conditions and of various arguments of the KAM-step. Also, an extra condition on weighted spatial structures has to be imposed, which we may call *coherence*: there exists an approximation function Θ such that

$$\max_{\lambda\in A}[\lambda] - \min_{\lambda\in A}[\lambda] \le \log\Theta([A]) \qquad \text{for } [A] \text{ large.}$$

With this provision Theorem A remains valid simply by stipulating

$$m > v + w$$

and defining Ψ_0 in terms of $\Theta^v\Delta$. Theorem B remains unchanged.

To give an example, consider a lattice of identical harmonic oscillators with nearest neighbour couplings, described by the hamiltonian

$$H = \tfrac{1}{2}\langle I, I\rangle + \sum_{i\in\mathbb{Z}^d} P_{A_i}$$

with coupling terms as in example (7). The plan is to expand the hamiltonian around a range of initial positions I^o and to introduce the associated frequencies $\omega(I^o) = I^o$ as new parameters. But choosing localized positions I^o forces the frequencies ω to approach zero at a certain rate, too, whence Theorem A has to be extended to be applicable here.

It turns out that for instance the weight functions

$$[A] = \sum_{|i-A|\le t} \log^\gamma(1+|i|), \qquad \gamma > 1$$

are coherent for all $t \ge 0$, and that it suffices to have

$$\lambda > \frac{2\lambda_*}{2-\lambda_*}, \qquad 1 < \lambda_* < 2$$

to do KAM. Here, λ_* is defined by (8) and converges to 1 as t increases to infinity. Hence, if $\lambda > 2$, then there are real analytic invariant tori localized like $I_i^o \sim e^{-\log^\gamma|i|}$ and filling a set of positive measure. This improves the results in [36].

Unfortunately, a similar approach does not seem to work for nonlinear partial differential equations such as the nonlinear wave equation, because arbitrary couplings are involved which are not coherent.

5 Outline of the Proof

Theorem A is proven by the familiar KAM-method employing a rapidly converging iteration scheme [12,1,19]. At each step of the scheme, a hamiltonian

$$H_n = N_n + P_n$$

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

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

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

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

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

Here is a more detailed description of this construction. Approximating the perturbation P in a suitable way we write

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

dropping the index n to simplify the notation. In particular, R is chosen such that its spatial expansion is finite, hence all subsequent operations are finite dimensional.

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

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

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

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

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

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

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

The point is to find F such that $N + R + \{N,F\} = N_+$ is a normal form. Equivalently, setting $N_+ = N + \hat{N}$, the linear equation

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

has to be solved for F and $\hat{N}$, when R is given. Given such a solution, we obtain $(1-t)\,\{N,F\} + R = (1-t)\hat{N} + tR$ and hence $H\circ\Phi = N_+ + P_+$ with

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

Setting up spatial expansions for F and $\hat{N}$ of the same form as that for R the linearized equation breaks up into the component equations

$$\partial F_A + \hat{N}_A = R_A,$$

where ∂ is the familiar linear partial differential operator with constant coefficients on the torus,

$$\partial = \sum_{\lambda\in\Lambda} \omega_\lambda \frac{\partial}{\partial\theta_\lambda}.$$

Their solution is well-known and straightforward. The operator ∂ is diagonizable with eigenfunctions $e^{i\langle k,\theta\rangle}$ and eigenvalues $i\,\langle k,\omega\rangle$, which in our case are zero if and only if k is zero by the nonresonance conditions. It therefore suffices to choose

$$\hat{N}_A = \langle R_A\rangle,$$

the mean value of R_A over $\mathbb{T}^A$, and to solve uniquely

$$\partial F_A = R_A - \langle R_A\rangle, \qquad \langle F_A\rangle = 0.$$

We obtain

$$F_A = \sum_{\substack{k \neq 0 \\ \operatorname{supp} k \subseteq A}} \frac{R_{A,k}}{i \langle k, \omega \rangle} e^{i \langle k, \theta \rangle}, \tag{9}$$

where $R_{A,k}$ are the Fourier coefficients of R_A.

The truncation of P will be chosen so that R is of first order in I. Hence the same is true of each of the $\hat{N}_A$ and so

$$\hat{N} = \sum_{A \in \mathcal{S}} \hat{N}_A = \hat{e} + \langle v(\omega), I \rangle .$$

It suffices to change parameters by setting

$$\omega_+ = \omega + v(\omega) \tag{10}$$

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

By the same truncation, F is of first order in I. It follows that $\Phi = X_F^t\big|_{t=1}$ has the form

$$\begin{aligned} \theta &= U(\theta_+) \\ I &= V(\theta_+) + W(\theta_+) I_+, \end{aligned}$$

where the dependence of all coefficients on ω has been suppressed. This map is composed with the inverse φ of the parameter map (10) to obtain $\mathcal{F}$.

Such symplectic transformations form a group under composition. So, if $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_n$ belong to this group, then so does $\mathcal{F}^n = \mathcal{F}_0 \circ \mathcal{F}_1 \circ \cdots \circ \mathcal{F}_n$ and the limit transformation $\mathcal{F}$ for $n \to \infty$.

For the mere existence of an invariant torus it would actually suffice to construct the embedding

$$\mathcal{F}|_{I=0} : \quad \mathcal{T}_0 \to \mathcal{P}.$$

This can be done via a new approach introduced by Salamon and Zehnder [32] that works in configuration space. For the traditional transformation method employed here, however, it is important to have control also over the normal bundle of the torus $\mathcal{T}_0$ which is transformed by

$$T\mathcal{F} = \frac{\partial I}{\partial I_+} = W. \tag{11}$$

This bundle transformation requires some special care in the convergence proof below in Section 7.

6 The KAM-Step

Before plunging into the details of the KAM-construction we observe that it suffices to consider some normalized value of α, say

$$\bar{\alpha} = 2.$$

Indeed, stretching the time scale by the factor $2/\alpha$ the hamiltonians H and N are scaled by the same amount, and so are the frequencies ω. By a similar scaling of the action-variables I the radius s may also be normalized to some convenient value. We will not do this here.

The set up

Consider a hamiltonian of the form

$$H = N + P, \qquad P = \sum_{A \in \mathcal{S}} P_A.$$

Assume that P is real analytic on the complex domain

$$\mathcal{D}_{r,s} \times \mathcal{W}_h\colon \quad |\mathrm{Im}\,\theta|_\infty < r, \ \ \| I \|_w < s, \ \ |\omega - \mathcal{O}_*|_\infty < h,$$

where $\mathcal{O}_*$ is a closed subset of the parameter space $\mathbb{R}^\Lambda$ consisting of points ω that satisfy

$$|\langle k, \omega \rangle| \ge \frac{\bar{\alpha}}{\Delta([\![k]\!])\Delta(|k|)}, \qquad 0 \ne k \in \mathbb{Z}^\Lambda,$$

where $\bar{\alpha} = 2$. Moreover, assume that for some $m > w$,

$$|||H - N|||_{m,r,s,h} = |||P|||_{m,r,s,h} \le \varepsilon$$

is sufficiently small. The precise condition will be given later in the course of the iteration.

Unless stated otherwise the following estimates are uniform with respect to ω. Therefore the index h is usually dropped.

Truncating the perturbation

Let μ and ρ be two small and T a large positive parameter to be chosen during the iteration process. The Fourier series of the A-component P_A of the perturbation is truncated at order $\langle A\rangle$ which is the smallest nonnegative number satisfying

$$\mu\,[A] + \rho\langle A\rangle \geq T. \tag{12}$$

Thus, the larger $[A]$ the more Fourier coefficients are discarded. If $[A]$ is sufficiently large the whole A-component is dropped. The upshot is that for the remaining perturbation Q one has

$$|||P - Q|||_{m-\mu,r-\rho,s} \leq e^{-T}\,|||P|||_{m,r,s}.$$

Next, each Fourier coefficient of Q is linearized with respect to y at the origin. Denoting the result of this truncation process by R we obtain

$$|||P - R|||_{m-\mu,r-\rho,\alpha s} \leq \left(e^{-T} + \frac{\alpha^2}{1-\alpha}\right)|||P|||_{m,r,s} \tag{13}$$

for $0 < \mu < m$, $0 < \rho < r$ and $0 < \alpha < 1$. Moreover, the estimate

$$|||R|||_{m,r,s} \leq 2|||P|||_{m,r,s}$$

obviously holds.

Extending the small divisor estimate

We claim that, if

$$h \leq \min_{A\in\mathcal{S}} \frac{1}{\Delta([A])\cdot\langle A\rangle\Delta(\langle A\rangle)} \tag{14}$$

with $\langle A\rangle$ as in the previous section, then the estimates

$$|\langle k,\omega\rangle| \geq \frac{1}{\Delta([\![k]\!])\Delta(|k|)}, \qquad \mu[\![k]\!] + \rho\,|k| \leq T, \quad k \neq 0$$

hold uniformly in ω on the complex neighbourhood $\mathcal{W}_h$ of the set $\mathcal{O}_*$.

The proof is simple. Given ω in $\mathcal{W}_h$ there exists an ω_* in $\mathcal{O}_*$ such that $|\omega-\omega_*|_\infty < h$. Given k there exists an A in $\mathcal{S}$ containing the support of k such that $[\![k]\!] = [A]$. It follows that $|k| \leq \langle A\rangle$ and hence

$$\begin{aligned}|\langle k,\omega\rangle - \langle k,\omega_*\rangle| &\leq |k|_1\,|\omega-\omega_*|_\infty \leq \langle A\rangle h \\ &\leq \frac{1}{\Delta([A])\Delta(\langle A\rangle)} \leq \frac{1}{\Delta([\![k]\!])\Delta(|k|)}\end{aligned}$$

by the monotonicity of Δ. The claim follows from the estimate for $\langle k, \omega_* \rangle$.

Solving the linearized equation

The linearized equation $\{F, N\} + \hat{N} = R$ is broken up into its spatial components, $\partial F_A + \hat{N}_A = R_A$, and solved for F_A and $\hat{N}_A$ as described in Section 5. Clearly, $\hat{N}_A$ is given by the mean value of R_A over $\mathbb{T}^A$, and $\|\hat{N}_A\|_{r,s} \le \|R_A\|_{r,s}$. Hence,

$$|||\hat{N}|||_{m,r,s} \le |||R|||_{m,r,s}$$

by putting pieces together.

The normalized Fourier series expansion of F_A is given by (9). By the extended small divisor estimate,

$$\begin{aligned} \|F_A\|_{r-\rho,s} &\le \sum_k \Delta([\![k]\!])\Delta(|k|) \left|R_{A,k}\right|_s e^{(r-\rho)|k|} \\ &\le \Delta([A])\Gamma_0(\rho)\|R_A\|_{r,s}, \end{aligned}$$

where $\Gamma_0(\rho) = \sup_{t \ge 0} \Delta(t)e^{-\rho t}$. Similarly, for the convenience of later estimates,

$$\begin{aligned} \sum_{\lambda \in A} \left\|F_{A,\theta_\lambda}\right\|_{r-\rho,s} &\le \sum_k \Delta([\![k]\!]) \cdot |k| \, \Delta(|k|) \left|R_{A,k}\right| e^{(r-\rho)|k|} \\ &\le \Delta([A])\Gamma_1(\rho)\|R_A\|_{r,s}, \end{aligned}$$

where $\Gamma_1(\rho) = \sup_{t \ge 0} (1+t)\Delta(t)e^{-\rho t}$. Putting the spatial components together,

$$\begin{aligned} |||F|||_{m-\mu,r-\rho,s} &\le \sum_A \Delta([A])\Gamma_0(\rho)\|R_A\|_{r,s} e^{(m-\mu)[A]} \\ &\le \Gamma_0(\mu)\Gamma_0(\rho)|||R|||_{m,r,s} \end{aligned}$$

and

$$\sum_\lambda \left|\left|\left|F_{\theta_\lambda}\right|\right|\right|_{m-\mu,r-\rho,s} \le \Gamma_0(\mu)\Gamma_1(\rho)|||R|||_{m,r,s}$$

for $0 < \mu < m$.

In view of the estimate $\Gamma_0(\rho) \le \rho\Gamma_1(\rho)$ in Lemma 6 we may summarize these estimates by writing

$$\rho^{-1}|||F|||_{m-\mu,r-\rho,s}, \; \sum_\lambda \left|\left|\left|F_{\theta_\lambda}\right|\right|\right|_{m-\mu,r-\rho,s} \le \Gamma_\mu \Gamma_\rho |||R|||_{m,r,s}, \tag{15}$$

with $\Gamma_\mu = \Gamma_0(\mu)$ and $\Gamma_\rho = \Gamma_1(\rho)$.

The derivatives of F

On the domain $\mathcal{D}_{r-\rho,s}$ we obtain the estimate

$$
\begin{aligned}
|F_\theta|_w = \sum_\lambda \left|F_{\theta_\lambda}\right| e^{w[\lambda]} &\le \sum_\lambda \sum_{A\ni\lambda} \left\|F_{A,\theta_\lambda}\right\|_{r-\rho,s} e^{w[A]} \\
&\le \sum_\lambda \left\|\left|F_{\theta_\lambda}\right|\right\|_{w,r-\rho,s} .
\end{aligned}
$$

Similarly, on the domain $\mathcal{D}_{r-\rho,s/2}$ we obtain the estimate

$$
\begin{aligned}
|F_I|_\infty = \sup_\lambda \left|F_{I_\lambda}\right| &\le \sup_\lambda \frac{2}{s} \sum_{A\ni\lambda} |F_A|_{r-\rho,s}\, e^{w[\lambda]} \\
&\le \frac{2}{s} \sum_A \|F_A\|_{r-\rho,s} e^{w[A]} \\
&\le \frac{2}{s} |||F|||_{w,r-\rho,s} .
\end{aligned}
\tag{16}
$$

Requiring that

$$
m - \mu \ge w \tag{17}
$$

and recalling the estimates for F, F_θ we thus have

$$
\frac{1}{\rho}\, |F_I|_\infty\, ,\; \frac{2}{s} |F_\theta|_w \le 2\varGamma_\mu \varGamma_\rho \cdot s^{-1} |||R|||_{m,r,s} \le 4\varGamma_\mu \varGamma_\rho\, \frac{\varepsilon}{s}
$$

uniformly on the domain $\mathcal{D}_{r-\rho,s/2}$.

These estimates are expressed more conveniently by means of a *weighted phase space norm.* Let

$$
|(\theta, I)|_{\mathcal{P}} = \max\left(|\theta|_\infty\, , |I|_w\right), \qquad W = \operatorname{diag}\left(\rho^{-1} I_\varLambda, 2s^{-1} I_\varLambda\right).
$$

Then the above estimates are equivalent to

$$
|W X_F|_{\mathcal{P}} \le 4\varGamma_\mu \varGamma_\rho E, \qquad E = \frac{\varepsilon}{s}
$$

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

Transforming the coordinates

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

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

to the boundary of $\mathcal{D}_a$ is exactly one half. Hence, if $16\Gamma_\mu\Gamma_\rho E \le 1$, then $|WX_F|_{\mathcal{P}}$ is less than or equal one fourth on $\mathcal{D}_a$ and consequently

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

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

$$|W(\Phi - id)|_{\mathcal{P};\mathcal{D}_b} \le 4\Gamma_\mu\Gamma_\rho E \tag{18}$$

holds.

In fact, under the present smallness condition on E this statement holds as well for the larger domain $\mathcal{D}_{r-\kappa\rho,\kappa s/4} \subset \mathcal{D}_a$ instead of $\mathcal{D}_b$, where $\kappa = 3/2$. The $|W\cdot|_{\mathcal{P}}$-distance of its boundary to $\mathcal{D}_b$ is exactly one fourth. Applying the general Cauchy inequality of Appendix B to the last estimate it follows that in addition,

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

where $\|\cdot\|_{\mathcal{P}}$ denotes the operator norm induced by $|\cdot|_{\mathcal{P}}$. Finally, if we require

$$4\Gamma_\mu\Gamma_\rho E \le \alpha \le 1/2,$$

then

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

by the same arguments as before.

Transforming the frequencies

To put $N_+ = N + \hat{N}$ into normal form, the frequency parameters are transformed by setting $\omega_+ = \omega + v(\omega)$. Proceeding just as in (16) the estimate for $\hat{N}$ implies that $|v|_\infty \le 2E$ uniformly on $\mathcal{W}_h$. Refering to Lemma 11 it follows that for

$$E \le h/8 \tag{19}$$

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

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

satisfying

$$|\varphi - id|_\infty\,, \quad \frac{h}{4}\left\|\frac{\partial\varphi}{\partial\omega} - I\right\|_\infty \le 2E \tag{20}$$

uniformly on $\mathcal{W}_b$.

The Poisson bracket

The estimate of the new error term hinges on an estimate for the Poisson bracket

$$\{F, G\} = \langle F_\theta, G_I\rangle - \langle F_I, G_\theta\rangle$$

in terms of the norm $|||\cdot|||$.

Consider the term $\langle F_I, G_\theta\rangle$. Given that F and G have a spatial expansion over the same structure $\mathcal{S}$ we have

$$\begin{aligned}\langle F_I, G_\theta\rangle &= \sum_\lambda F_{I_\lambda} G_{\theta_\lambda} \\ &= \sum_\lambda \Big(\sum_{A\ni\lambda} F_{A,I_\lambda}\Big)\Big(\sum_{B\ni\lambda} G_{B,\theta_\lambda}\Big) \\ &= \sum_{A,B}\sum_{\lambda\in A\cap B} F_{A,I_\lambda} G_{B,\theta_\lambda} \\ &= \sum_{\substack{A,B\\ A\cap B\neq\emptyset}} \big\langle F_{A,I}, G_{B,\theta}\big\rangle.\end{aligned}$$

The term $\big\langle F_{A,I}, G_{B,\theta}\big\rangle$ "lives" on $A\cup B$ which belongs to $\mathcal{S}$ by the definition of a spatial structure. Hence, $\langle F_I, G_\theta\rangle$ has the same structure.

One easily verifies that the norm $\|\cdot\|$ is multiplicative. Moreover, the "amplified" Cauchy inequalities

$$\sum_{\lambda\in C}\big\|H_{\theta_\lambda}\big\|_{r-\rho,s} \le \sup_k |k|\, e^{-\rho|k|}\|H\|_{r,s} \le \frac{1}{e\rho}\|H\|_{r,s}$$

and

$$\sup_{\lambda\in C}\big\|H_{I_\lambda}\big\|_{r,s-\sigma} \le \sup_{\lambda\in C}\frac{e^{w[\lambda]}}{\sigma}\|H\|_{r,s} \le \frac{e^{w[C]}}{\sigma}\|H\|_{r,s} \tag{21}$$

hold for C in $\mathcal{S}\cap\mathcal{S}$ in view of the definition of $[\lambda]$ no matter on which set the function H actually “lives”. It follows that

$$\begin{aligned}\left\|\left\langle F_{A,I}, G_{B,\theta}\right\rangle\right\|_{r-\rho,s-\sigma} &\leq \sum_{\lambda\in A\cap B} \left\|F_{A,I_\lambda}\right\|_- \left\|G_{B,\theta_\lambda}\right\|_- \\ &\leq \sup_{\lambda\in A\cap B} \left\|F_{A,I_\lambda}\right\|_- \sum_{\lambda\in A\cap B} \left\|G_{B,\theta_\lambda}\right\|_- \\ &\leq \frac{1}{e\rho\sigma} e^{w[A\cap B]} \|F_A\|_{r-\rho,s} \|G_B\|_{r,s},\end{aligned} \tag{22}$$

where $\|\cdot\|_- = \|\cdot\|_{r-\rho,s-\sigma}$. Now recall that

$$[A\cup B] + [A\cap B] \leq [A] + [B].$$

Thus, for $v \geq w$, we obtain

$$\begin{aligned}|||\langle F_I, G_\theta\rangle|||_{v,r-\rho,s-\sigma} &\leq \sum_{\substack{A,B\\ A\cap B\neq\emptyset}} \left\|\left\langle F_{A,I}, G_{B,\theta}\right\rangle\right\|_{r-\rho,s-\sigma} e^{v[A\cup B]} \\ &\leq \frac{1}{e\rho\sigma} \sum_{\substack{A,B\\ A\cap B\neq\emptyset}} e^{v[A\cup B]} e^{v[A\cap B]} \|F_A\|_{r-\rho,s} \|G_B\|_{r,s} \\ &\leq \frac{1}{e\rho\sigma} \sum_{A,B} e^{v[A]} \|F_A\|_{r-\rho,s} e^{v[B]} \|G_B\|_{r,s} \\ &= \frac{1}{e\rho\sigma} |||F|||_{v,r-\rho,s} |||G|||_{v,r,s}.\end{aligned}$$

The term $\langle F_\theta, G_I\rangle$ is handled similarly. However, in order to avoid an unnecessary shrinking of the θ-domain and take advantage of the estimate of F_θ in (15), one may vary the argument to obtain

$$|||\langle F_\theta, G_I\rangle|||_{v,r-\rho,s-\sigma} \leq \frac{1}{\sigma} |||G|||_{v,r,s} \sum_{\lambda} \left|\left|\left|F_{\theta_\lambda}\right|\right|\right|_{v,r-\rho,s}.$$

Hence, if

$$\rho^{-1} |||F|||_{v,r-\rho,s}, \quad \sum_{\lambda} \left|\left|\left|F_{\theta_\lambda}\right|\right|\right|_{v,r-\rho,s} \leq M$$

and $v \geq w$, then

$$|||\{F, G\}|||_{v,r-\rho,s-\sigma} \leq \frac{2M}{\sigma} |||G|||_{v,r,s} \tag{23}$$

for $0 < \sigma < s$.

Estimating the new error term

The new error term is

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

where $R_t = (1-t)\hat{N} + tR$. By Lemma 10 and estimate (15),

$$\left|\!\left|\!\left| G \circ X_F^t \right|\!\right|\!\right|_{m-\mu, r-2\rho, \alpha s/2} \le 2 |||G|||_{m-\mu, r-\rho, \alpha s}, \qquad 0 \le t \le 1,$$

provided that

$$4C_0 \Gamma_\mu \Gamma_\rho E \le \alpha \le \frac{1}{2} \tag{24}$$

where $C_0 = 8$ is the constant of Lemma 10. Hence, with this assumption,

$$|||P_+|||_{m-\mu, r-2\rho, \alpha s/2} \le \int_0^1 2|||\{R_t, F\}|||_{m-\mu, r-\rho, \alpha s} \, dt \\ + 2|||P - R|||_{m-\mu, r-\rho, \alpha s}.$$

Obviously, $|||R_t|||_{m,r,s} \le 2\varepsilon$ for $0 \le t \le 1$ by the estimates for $\hat{N}$ and R and therefore

$$|||\{R_t, F\}|||_{m-\mu, r-\rho, \alpha s} \le 8\Gamma_\mu \Gamma_\rho E \varepsilon$$

by the estimates for F, F_x and (23) with $\sigma = s/2$. Combined with (13) we altogether obtain

$$|||P_+|||_{m-\mu, r-2\rho, \alpha s/2} \le 16\Gamma_\mu \Gamma_\rho E \varepsilon + 2e^{-T} \varepsilon + 4\alpha^2 \varepsilon \tag{25}$$

for the new error term.

7 Iteration and Convergence

Heuristic considerations

Choosing $e^{-T} \sim \Gamma_* E$ and $\alpha^2 \sim \Gamma_* E$ with $\Gamma_* \sim \Gamma_\mu \Gamma_\rho$ all terms in the error estimate (25) are approximately of the same size so that $\varepsilon_+ \sim \Gamma_* E \varepsilon$. Dividing by $s_+ \sim \alpha s$,

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

That is,

$$E_+ \sim \Gamma_*^{\kappa-1} E^\kappa, \qquad \kappa = \frac{3}{2}.$$

This estimate is iterated with small divisor functions $\Gamma_0 \le \Gamma_1 \le \dots$ in place of Γ_* arising from nonincreasing sequences $\mu_0 \ge \mu_1 \ge \dots > 0$ and $\rho_0 \ge \rho_1 \ge \dots > 0$. After n steps,

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

where

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

With an appropriate choice of the μ_ν and ρ_ν, the infinite product of the $\Gamma_\nu^{\kappa_\nu}$ converges to a constant multiple of $\Psi_0(\mu)\Psi_1(\rho)$ which by hypotheses is finite. Thus, if E_0 is sufficiently small, then the E_n concerge to zero exponentially fast.

The actual choice of the Γ_n has to take into account an important constraint. By comparison with (28) condition (14) turns out to be tantamount to

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

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

The iterative construction

Let $a = 13$, $b = 4$, $c = 5$, $d = 8$ and $e = 22$. The choice of these integer constants will be motivated later in the course of the proof of the iterative lemma.

Given $0 < \mu \le m - w$ and $0 < \rho < r/2$ there exist sequences $\mu_0 \ge \mu_1 \ge \dots > 0$ and $\rho_0 \ge \rho_1 \ge \dots > 0$ such that

$$\Psi_0(\mu)\Psi_1(\rho) = \prod_{\nu=0}^{\infty} \Gamma_{\mu_\nu}^{\kappa_\nu} \Gamma_{\rho_\nu}^{\kappa_\nu}$$

with

$$\sum_{\nu=0}^{\infty} \mu_\nu = \mu, \qquad \sum_{\nu=0}^{\infty} \rho_\nu = \rho,$$

where $\Gamma_{\mu_\nu} = \Gamma_0(\mu_\nu)$ and $\Gamma_{\rho_\nu} = \Gamma_1(\rho_\nu)$. Fix such sequences, and for $n \ge 0$ set

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

where $\Theta_0 = 1$. Furthermore, set

$$m_n = m - \sum_{\nu=0}^{n-1} \mu_\nu, \qquad r_n = r - 2\sum_{\nu=0}^{n-1} \rho_\nu,$$
$$s_n = s \prod_{\nu=0}^{n-1} \frac{\alpha_\nu}{2}, \qquad h_n = 2^{n+c} E_n,$$

where $\alpha_n^2 = 4^{-b} \Gamma_n E_n$. Then $m_n \downarrow m - w$, $r_n \downarrow r - 2\rho$ and $s_n \downarrow 0$, $h_n \downarrow 0$. These sequences define the complex domains

$$\mathcal{D}_n = \mathcal{D}_{r_n, s_n}, \quad \mathcal{W}_n = \mathcal{W}_{h_n}.$$

Finally, we introduce an extended phase space norm,

$$|(\theta, I, \omega)|_{\bar{\mathcal{P}}} = \max\left(|\theta|_\infty, |I|_w, |\omega|_\infty\right),$$

and the corresponding weight matrices,

$$\bar{W}_n = \operatorname{diag}\left(\rho_n^{-1} I_\Lambda, 2s_n^{-1} I_\Lambda, h_n^{-1} I_\Lambda\right).$$

Then we can state the iterative lemma.

Iterative Lemma. *Suppose that*

$$s^{-1} |||H - N|||_{m,r,s,h} \le \frac{\bar{\alpha}\varepsilon_*}{\Psi_0(\mu)\Psi_1(\rho)} \le \frac{h}{2^c},$$

where $\bar{\alpha} = 2$ *and* $\varepsilon_* = 2^{-e}$. *Then for each* $n \ge 0$ *there exists a normal form* N_n *and a real analytic transformation*

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

of the form described in Section 5, which is symplectic for each ω, *such that*

$$s_n^{-1} \left|\left|\left| H \circ \mathcal{F}^n - N_n \right|\right|\right|_{m_n, r_n, s_n, h_n} \le E_n. \tag{26}$$

Moreover,

$$\left|\bar{W}_0\left(\mathcal{F}^{n+1}-\mathcal{F}^n\right)\right|_{\bar{\mathcal{P}}} \leq 4 \max\left(2^{1-a-n}\Gamma_n E_n,\, E_n/h_n\right),$$

and

$$\left\| T\mathcal{F}^{n+1} - T\mathcal{F}^n \circ \mathcal{F}_n \right\|_w \leq 2^{5-a-n}\Gamma_n E_n$$

on $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$*, where* $T\mathcal{F}$ *is defined in (11), and* $\|\cdot\|_w$ *also denotes the operator norm induced by* $\|\cdot\|_w$*.*

In components the weighted operator norm of $W = T\mathcal{F}$ more explicitly reads

$$\|W\|_w = \sup_{\mu\in\Lambda} \sum_{\lambda\in\Lambda} \left|W_{\lambda\mu}\right| e^{w([\lambda]-[\mu])}.$$

Auxiliary inequalities

Before giving the proof of the lemma we collect some useful facts. The κ_ν satisfy the identities

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

This and the monotonicity of the Γ-function imply that

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

Together with the definition of E_n we obtain the estimate

$$\Gamma_n E_n \leq \left(\prod_{\nu=0}^{\infty} \Gamma_\nu^{\kappa_\nu} E_0\right)^{\kappa^n} = \left(2^{2+a}\Psi_0\Psi_1 E_0\right)^{\kappa^n}. \tag{27}$$

Moreover, $\Gamma_n^{\kappa-1}E_n^\kappa = E_{n+1}$ by a straightforward calculation.

Proof of the lemma

The lemma is proven by induction. Choosing $\mathcal{F}^0 = id$ and

$$E_0 = \frac{\bar{\alpha}\varepsilon_*}{\Psi_0(\mu)\Psi_1(\rho)},$$

there is nothing to prove for $n = 0$. Just observe that $h_0 \le h$ by the very definition of h_0 and E_0.

So let $n \ge 0$. To apply the KAM-step to $H_n = H \circ \mathcal{F}^n$ and N_n we need to verify its assumptions (14), (17), (19) and (24). Clearly, $m_n - \mu_n \ge w$ by construction, and $E_n \le h_n/8$ in view of the definition of h_n and $c \ge 3$, so the second and third requirements are met. Taking squares, the fourth requirement is equivalent to

$$4^{2-a-n} C_0^2 \Gamma_n^2 E_n^2 \le 4^{-b} \Gamma_n E_n \le 4^{-1}.$$

This holds for all $n \ge 0$, since $C_0 = 8$,

$$\Gamma_n E_n \le 2^{3+a-e}$$

by (27) and $a \ge b + 2$, $b \ge 0$, $e \ge a + 9$.

As to the first requirement, define T_n by $e^{-T_n} = 2^{-d}\Gamma_n E_n$ and subsequently $\langle \cdot \rangle$ as in (12). For arbitrary A in $\mathcal{S}$ with $\langle A \rangle > 0$ we then have

$$\begin{aligned} \frac{1}{\langle A \rangle \Delta(\langle A \rangle)\Delta([A])} &= \frac{e^{-\rho_n \langle A \rangle} e^{-\mu_n [A]}}{\langle A \rangle \Delta(\langle A \rangle) e^{-\rho_n \langle A \rangle} \cdot \Delta([A]) e^{-\mu_n [A]}} \\ &\ge \frac{e^{-T_n}}{\Gamma_{\mu_n}\Gamma_{\rho_n}} = \frac{2^{-d}\Gamma_n E_n}{2^{-n-a}\Gamma_n} = 2^{n+a-d} E_n \ge h_n, \end{aligned} \tag{28}$$

since $a \ge c + d$. This estimate holds even more when $\langle A \rangle = 0$. Hence, also requirement (14) is satisfied.

The KAM-construction now provides a normal form N_{n+1}, a coordinate transformation Φ_n and a parameter transformation φ_n. By the definition of r_n and s_n, Φ_n maps $\mathcal{D}_{n+1}$ into $\mathcal{D}_n$, while φ_n maps $\mathcal{W}_{n+1}$ into $\mathcal{W}_n$, since

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

in view of (27) and $e \ge a + 9$. Setting

$$\mathcal{F}^{n+1} = \mathcal{F}^n \circ \mathcal{F}_n, \qquad \mathcal{F}_n = \Phi_n \circ \varphi_n,$$

we obtain a transformation $\mathcal{F}^{n+1}$ from $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$ into $\mathcal{D}_0 \times \mathcal{W}_0$. For the new error term

$$P_{n+1} = H \circ \mathcal{F}^{n+1} - N_{n+1} = H_n \circ \mathfrak{F}_n - N_{n+1}$$

we obtain

$$\begin{aligned}
|||P_{n+1}|||_{n+1} &\le 16\Gamma_{\mu_n}\Gamma_{\rho_n}E_n\varepsilon_n + 2e^{-T_n}\varepsilon_n + 4\alpha_n^2\varepsilon_n \\
&\le \left(2^{4-a} + 2^{1-d} + 2^{2-2b}\right)\Gamma_n E_n \varepsilon_n .
\end{aligned}$$

Dividing by $s_{n+1} = \alpha_n s_n/2$ this yields

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

since $a \ge b+7$, $b \ge 4$ and $d \ge b+4$.

To prove the first of the estimates, write

$$\begin{aligned}
\left|\bar{W}_0\left(\mathcal{F}^{n+1} - \mathcal{F}^n\right)\right|_{n+1} &= \left|\bar{W}_0\left(\mathcal{F}^n \circ \mathfrak{F}_n - \mathcal{F}^n\right)\right|_{n+1} \\
&\le \left\|\bar{W}_0 \bar{D}\mathcal{F}^n \bar{W}_n^{-1}\right\|_n \left|\bar{W}_n\left(\mathfrak{F}_n - id\right)\right|_{n+1},
\end{aligned}$$

where $|\cdot|_n = |\cdot|_{\bar{\mathcal{P}},\mathcal{D}_n\times\mathcal{W}_n}$, and $\bar{D}$ denotes differentiation with respect to (θ, I, ω). By the estimates for Φ and φ and the definition of Γ_n,

$$\begin{aligned}
\left|\bar{W}_n\left(\mathfrak{F}_n - id\right)\right|_{n+1} &\le \max\left(\left|W_n\left(\Phi_n - id\right)\right|_{\mathcal{P}}, h_n^{-1}\left|\varphi_n - id\right|_\infty\right) \\
&\le \max\left(2^{2-a-n}\Gamma_n E_n, 2E_n/h_n\right).
\end{aligned}$$

It remains to show that the first factor is bounded by 2. By the inductive construction, $\mathcal{F}^n = \mathfrak{F}_0 \circ \cdots \circ \mathfrak{F}_{n-1}$, and

$$\begin{aligned}
&\left\|\bar{W}_\nu \bar{D}\mathfrak{F}_\nu \bar{W}_\nu^{-1}\right\|_{\nu+1} \\
&\quad\le \max\left(\left\|W_\nu D\Phi_\nu W_\nu^{-1}\right\|_{\mathcal{P}} + h_\nu |W_\nu \partial_\omega \Phi_\nu|_{\mathcal{P},\infty}, |\partial_\omega \varphi_\nu|_\infty\right) \\
&\quad\le \max\left(1 + 2^{5-a-\nu}\Gamma_\nu E_\nu, 1 + 8E_\nu/h_\nu\right) \\
&\quad\le 1 + 2^{3-c-\nu}
\end{aligned}$$

by (18) and (20). Since the weights of $\bar{W}_\nu^{-1}$ do *not* decrease as ν decreases, and since $c \geq 5$, we obtain

$$\left| \bar{W}_0 \bar{D}\mathcal{F}^n \bar{W}_n^{-1} \right|_n \leq \prod_{\nu=0}^{n-1} \left| \bar{W}_\nu \bar{D}\mathcal{F}_\nu \bar{W}_{\nu+1}^{-1} \right|_{\nu+1}$$
$$\leq \prod_{\nu=0}^{\infty} \left(1 + 2^{3-c-\nu}\right) \leq 2.$$

To prove the second estimate, observe that

$$T\mathcal{F}^{n+1} = T\mathcal{F}^n \circ \mathcal{F}_n \cdot T\mathcal{F}_n,$$

since θ and ω are transformed independently of the I-coordinate. It follows that

$$\left| T\mathcal{F}^{n+1} - T\mathcal{F}^n \circ \mathcal{F}_n \right|_w \leq \left| T\mathcal{F}^n \circ \mathcal{F}_n \right|_w \left| T\mathcal{F}_n - I \right|_w$$

uniformly on $\mathcal{D}_{n+1} \times \mathcal{W}_{n+1}$. By (18) and the definition of Γ_n,

$$\left| T\mathcal{F}_n - I \right|_w \leq \left\| W_n \left(D\Phi_n - I\right) W_n^{-1} \right\|_{\mathcal{P}} \leq 2^{4-a-n} \Gamma_n E_n,$$

and by a standard telescoping argument as above, $\left| T\mathcal{F}^n \circ \mathcal{F}_n \right|_w \leq 2$. This completes the proof of the iterative lemma.

Convergence

By the estimates of the iterative lemma the $\mathcal{F}^n$ and subsequently the $T\mathcal{F}^n$ converge uniformly on

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

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

$$\left| \bar{W}_0 \left(\mathcal{F}_* - id\right) \right|_{\bar{\mathcal{P}}}, \ \left| T\mathcal{F}_* - I \right|_w \leq \frac{1}{2}$$

on $\mathcal{D}_* \times \mathcal{O}_*$ by the usual telescoping argument.

But by construction, the $\mathcal{F}^n$ are affine linear in each fiber over $\mathbb{T}^\Lambda \times \mathcal{O}_*$. Therefore they indeed converge uniformly on *any* domain $\mathcal{D}_{r-2\rho,\sigma} \times \mathcal{O}_*$ with $\sigma > 0$ to a map $\mathcal{F}_*$ that is real analytic and symplectic for each ω. In particular,

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

by piecing together the above estimates.

Going to the limit in (26) and using Cauchy's inequality we finally obtain

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

This completes the proof of Theorem A.

Estimates

The scheme so far provides only a very crude estimate of $\mathcal{F}_*$ since the actual size of the perturbation is not taken into account in the estimates of the iterative lemma. But nothing changes when all inequalities are scaled down by the factor $\varepsilon / E \leq 1$, where

$$\varepsilon = s^{-1} |||H - N|||_{m,r,s,h} \leq E = \frac{\bar{\alpha} \varepsilon_*}{\Psi_\mu \Psi_\rho} .$$

It follows that

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

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

8 The Measure Estimate

In this section the proof of Theorem B is given.

The set of all illegitimate frequencies in the entire frequency space $\mathbb{R}^\Lambda$ for a given parameter value α is

$$\mathbb{R}^\Lambda - \mathbb{R}^\Lambda_\alpha = \bigcup_{0 \neq k \in \mathbb{Z}^\Lambda} \mathcal{R}_k(\alpha),$$

where

$$\mathcal{R}_k(\alpha) = \left\{ \omega \in \mathbb{R}^\Lambda : |\langle k, \omega \rangle| < \frac{\alpha}{\Delta([\![k]\!]) \Delta(|k|)} \right\}$$

are the individual open *resonance zones*. The first step is to design a probability measure μ that gives a useful estimate of the size of these zones.

Lemma 1. *Under the hypotheses of Theorem B there exists a probability* *measure* μ *on* $\mathbb{R}^\Lambda$ *with support at any prescribed point such that*

$$\alpha^{-1}\mu(\mathcal{R}_k) \le \frac{1 + [\![k]\!]}{\Delta([\![k]\!])\Delta(|k|)}$$

for all $0 \ne k \in \mathbb{Z}^\Lambda$.

Proof. We construct a measure with support at the origin. As we will see the pertaining estimates are not affected by translations of this measure, so its support may be shifted to any prescribed point.

Let

$$d\sigma(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

be the standard gaussian measure on the real line with mean zero and variance one, and set

$$d\mu(\omega) = \prod_{\lambda\in\Lambda} a_\lambda^* d\sigma(\omega_\lambda) = \prod_{\lambda\in\Lambda} \frac{a_\lambda}{\sqrt{2\pi}}\, e^{-a_\lambda^2\omega_\lambda^2/2}\, d\omega_\lambda$$

with scale factors $a_\lambda = 1 + [\lambda]$. The measure of the resonance zones then is

$$\mu(\mathcal{R}_k) = \int_{\mathcal{R}_k} \prod_{\lambda\in\Lambda} a_\lambda^* d\sigma(\omega_\lambda) = \int_{L_a\mathcal{R}_k} \prod_{\lambda\in\Lambda} d\sigma(\omega_\lambda),$$

where L_a is the unbounded linear operator mapping ω_λ into $a_\lambda\omega_\lambda$, and

$$L_a\mathcal{R}_k = \left\{ \omega : \left|\left\langle L_a^{-1}k, \omega\right\rangle\right| < \frac{\alpha}{\Delta([\![k]\!])\Delta(|k|)} \right\}.$$

As in the finite dimensional case, this yields

$$\alpha^{-1}\mu(\mathcal{R}_k) \le \frac{1}{\|L_a^{-1}k\|} \cdot \frac{1}{\Delta([\![k]\!])\Delta(|k|)},$$

where $\|\cdot\|$ denotes the euclidean length. The leading factor is bounded by $\|k\|^{-1}$ $\max_{\lambda\in K} a_\lambda$ with $K = \operatorname{supp} k$. Since

$$\max_{\lambda\in K}[\lambda] = \max_{\lambda\in K} \min_{\lambda\in A\in\mathcal{S}\cap\mathcal{S}} [A] \le \min_{K\subseteq A\in\mathcal{S}} [A] = [\![k]\!],$$

we indeed obtain a slightly better result than stated. Obviously, this estimate is not worsened by any translation of the measure μ.

To study the support of μ, let $Q_\varepsilon\colon |\omega|_\infty < \varepsilon$ be a basis of neighbourhoods of the origin. Clearly,

$$\begin{aligned}
\mu(Q_\varepsilon) &= \prod_{\lambda\in\Lambda} \int_{-\varepsilon}^{\varepsilon} a_\lambda^* d\sigma(\omega_\lambda) \\
&= \prod_{\lambda\in\Lambda} \frac{1}{\sqrt{2\pi}} \int_{-a_\lambda\varepsilon}^{a_\lambda\varepsilon} e^{-x^2/2} dx \\
&= \prod_{\lambda\in\Lambda} \left(1 - \sqrt{2/\pi} \int_{a_\lambda\varepsilon}^{\infty} e^{-x^2/2} dx\right) \\
&\geq \prod_{\lambda\in\Lambda} \left(1 - e^{-a_\lambda^2\varepsilon^2/2}\right)
\end{aligned}$$

in view of the estimate

$$\sqrt{2/\pi} \int_h^\infty e^{-x^2/2} dx \leq e^{-h^2/2}$$

for $h \geq 0$. Proceeding exactly as in the proof of the next lemma and letting $\delta = \varepsilon^2/2$ we now have the very crude estimate

$$\begin{aligned}
\sum_{\lambda\in\Lambda} e^{-a_\lambda^2\delta} &\leq \sum_{A\in\mathcal{S}} |A|\, e^{-[A]^2\delta} \\
&\leq \sum_{n=1}^{\infty} n \left(N_n(0) + \int_0^\infty e^{-t^2\delta} dN_n(t)\right) < \infty
\end{aligned}$$

for every $\delta > 0$. This suffices to show that $\mu(Q_\varepsilon) > 0$ for all $\varepsilon > 0$, whence the measure μ has support at the origin. ∎

From now on it is convenient to choose different approximation functions for $|k|$ and $[\![k]\!]$ in the definition of the resonance zones. Replacing the latter by $(1+t)\Delta(t)$, which is again an approximation function, we obtain

$$\begin{aligned}
\alpha^{-1}\mu\left(\mathbb{R}^\Lambda - \mathbb{R}^\Lambda_\alpha\right) &\leq \sum_{k\in\mathbb{Z}^\Lambda} \alpha^{-1}\mu(\mathcal{R}_k) \\
&\leq \sum_{k\in\mathbb{Z}^\Lambda} \frac{1}{\Delta([\![k]\!])\Delta(|k|)} \\
&\leq \sum_{A\in\mathcal{S}} \left(\frac{1}{\Delta([A])} \sum_{k\in\mathbb{Z}^A} \frac{1}{\Delta(|k|)}\right)
\end{aligned}$$

$$\leq \sum_{n=1}^{\infty} \Bigg(\sum_{A \in \mathcal{S},\, |A|=n} \frac{1}{\Delta([A])} \Bigg) \Bigg(\sum_{k \in \mathbb{Z}^n} \frac{1}{\Delta(|k|)} \Bigg).$$

Thus the sum is broken up with respect to the cardinality and the weight of the spatial components of $\mathcal{S}$. Each of these factors is now studied separately.

Lemma 2. *Under the hypotheses of Theorem B, there exists an approximation function Δ for every given approximation function Θ such that*

$$\sum_{A \in \mathcal{S},\, |A|=n} \frac{1}{\Delta([A])} \leq \frac{2N_0}{\Theta(t_n)}, \qquad n \geq 1.$$

Proof. In view of the definition of the distribution function N_n and the monotonicity of approximation functions the sum in question may be written as a Stieltjes integral:

$$\begin{aligned} \sum_{A \in \mathcal{S},\, |A|=n} \frac{1}{\Delta([A])} &= \inf_{t_\nu} \; N_n(0) + \sum_{\nu=0}^{\infty} \frac{N_n(t_{\nu+1}) - N_n(t_\nu)}{\Delta(t_\nu)} \\ &= N_n(0) + \int_0^\infty \frac{dN_n(t)}{\Delta(t)}, \end{aligned}$$

where the infimum is taken over all partitions $0 = t_0 < t_1 < t_2 < \ldots$ of the positive real axis. Integrating by parts,

$$\begin{aligned} \int_0^\infty \frac{dN_n(t)}{\Delta(t)} &= \frac{N_n(t)}{\Delta(t)}\bigg|_0^\infty + \int_0^\infty N_n(t)\, \frac{d \log \Delta(t)}{\Delta(t)} \\ &= -N_n(0) + \int_{t_n}^\infty N_n(t)\, \frac{d \log \Delta(t)}{\Delta(t)} \end{aligned}$$

for every sufficiently "strong" approximation function Δ. In particular, for

$$\Delta(t) = \Theta(t)\Phi^2(t)$$

we have $d \log \Delta(t) \leq 2\, d \log \Theta(t)\Phi(t)$, so that together with $N_n(t) \leq N_0 \Phi(t)$,

$$\int_{t_n}^\infty N_n(t)\, \frac{d \log \Delta(t)}{\Delta(t)} \leq 2N_0 \int_{t_n}^\infty \frac{d \log \Theta(t)\Phi(t)}{\Theta(t)\Phi(t)} = \frac{2N_0}{\Theta(t_n)\Phi(t_n)},$$

which is even better than our claim. ∎

Lemma 3.

$$\sum_{k\in\mathbb{Z}^n} \frac{1}{\Delta(|k|)} \le 2^n \int_0^\infty \binom{n+t}{n} \frac{d\log\Delta(t)}{\Delta(t)},$$

provided that $t^n/\Delta(t) \to 0$ *as* $t \to \infty$.

Proof. Let $V_n(t) = \operatorname{card}\{\, k \in \mathbb{N}^n : |k| \le t \,\}$. Proceeding exactly as in the last proof we have

$$\sum_{k\in\mathbb{Z}^n} \frac{1}{\Delta(|k|)} \le 2^n \sum_{k\in\mathbb{N}^n} \frac{1}{\Delta(|k|)}$$

and

$$\begin{aligned}\sum_{k\in\mathbb{N}^n} \frac{1}{\Delta(|k|)} &= \inf_{t_\nu}\ 1 + \sum_{\nu=0}^{\infty} \frac{V_n(t_{\nu+1}) - V_n(t_\nu)}{\Delta(t_\nu)} \\ &= 1 + \int_0^\infty \frac{dV_n(t)}{\Delta(t)} = \int_0^\infty V_n(t)\, \frac{d\log\Delta(t)}{\Delta(t)}\end{aligned}$$

by partial integration. The upper boundary term vanishes by our assumption on the function Δ.

We claim that

$$V_n(t) = \binom{[n+t]}{n} \le \binom{n+t}{n}$$

for all $n \ge 1$ and $t \ge 0$, where $[n+t]$ denotes the largest integer not bigger than $n+t$. This will prove the lemma.

For the proof let $t = l$ be an integer. For $n = 1$,

$$V_1(l) = 1 + l = \binom{1+l}{1},$$

so the equality is correct in this case. Proceeding by induction,

$$V_{n+1}(l) = \sum_{k=0}^{l} V_n(k) = \sum_{k=0}^{l} \binom{n+k}{n} = \binom{n+l+1}{n+1}$$

by a well-known identity for binomial coefficients. Thus the identity holds for all $n \ge 1$ and integer values of t. The general statement follows from the fact that V_n is constant on every interval $l \le t < l+1$. ∎

The estimate of Lemma 3 is of a rather general nature. This is now specialized for a certain class of approximation functions in a way that suits our needs.

Lemma 4. *There are approximation functions* $\varDelta$ *such that*

$$\sum_{k\in\mathbb{Z}^n} \frac{1}{\varDelta(|k|)} \leq K^{n\log\log n}$$

for all sufficiently large n *with some constant* K.

Of course, this also gives a bound for all small n, since the left hand side is monotonically increasing with n.

Proof. For $t \leq n$,

$$\binom{n+t}{n} \leq \binom{2n}{n} = \frac{(2n)!}{(n!)^2} \leq 4^n$$

for all $n \geq 1$ by well known estimates for the factorial function. Hence,

$$\int_0^n \binom{n+t}{n} \frac{d\log\varDelta(t)}{\varDelta(t)} \leq 4^n \int_0^\infty \frac{d\log\varDelta(t)}{\varDelta(t)} = 4^n$$

for *every* approximation function $\varDelta$. So this part of the integral is alright under all circumstances.

Now consider $t \geq n$, where

$$\binom{n+t}{n} = \frac{1}{n!}(t+1)\ldots(t+n) \leq \frac{2^n}{n!}t^n.$$

Let φ be given by $\varphi(s) = \log^2 s$, and define $\varDelta$ by stipulating that $t \mapsto s = \log\varDelta(t)$ is the inverse function of $s \mapsto t = s\varphi(s)$, at least for large t and s respectively. One easily verifies that this gives rise to an approximation function. Since

$$s\varphi(s)\Big|_{n/\varphi(n)} = \frac{n}{\varphi(n)}\,\varphi\left(\frac{n}{\varphi(n)}\right) \leq n$$

by the monotonicity of φ, the change of variables formula yields

$$\begin{aligned} \int_n^\infty \binom{n+t}{n} \frac{d\log\varDelta(t)}{\varDelta(t)} &\leq \frac{2^n}{n!} \int_n^\infty t^n \frac{d\log\varDelta(t)}{\varDelta(t)} \\ &\leq \frac{2^n}{n!} \int_{s_n}^\infty s^n\varphi^n(s)e^{-s}\,ds \end{aligned} \tag{29}$$

with $s_n = n/\varphi(n)$. Now, for all large n and $s \geq s_n$,

$$\varphi(s) = \log^2 s \leq s^{h_n}, \qquad h_n = \frac{\log \varphi(s_n)}{\log s_n} \leq \frac{4 \log\log n}{\log n}.$$

Thus, for all large n,

$$\begin{aligned}\int_n^\infty \binom{n+t}{n} \frac{d \log \Delta(t)}{\Delta(t)} &\leq \frac{2^n}{n!} \int_0^\infty s^{n+nh_n} e^{-s}\, ds \\ &\leq \frac{2^n}{n^n} (n + nh_n)^{n+nh_n+1} \\ &= 2^n A_n^n\, n^{nh_n+1}\end{aligned}$$

with $A_n = (1+h_n)^{1+h_n+\frac{1}{n}}$. The final estimate follows, since $A_n \to 1$ as $n \to \infty$ and $nh_n \log n = 4n \log\log n$. ∎

There is nothing special about our choice of the function φ in the last proof. Indeed, any nonnegative function φ satisfying

$$\varphi(s) \nearrow \infty \quad \text{as } s \to \infty, \qquad \int^\infty \frac{ds}{s\varphi(s)} < \infty$$

gives rise to an approximation function Δ by stipulating that $t \mapsto s = \log \Delta(t)$ is the inverse function of $s \mapsto t = s\varphi(s)$, and vice versa. This provides an alternate way of characterizing approximation functions. The argument of the proof applies to every such function which in addition satisfies the monotonicity condition

$$\frac{\log \varphi(s)}{\log s} \searrow 0 \qquad \text{as } s \to \infty,$$

giving a bound $K^{n \log \varphi(n)}$ for the sum in question.

Nothing is gained, however, from such greater generality. For, in any event we have $\varphi(s) > \log s$ asymptotically, so that in (29) already the integral around $s = n$ yields a bound of the order $K^{n \log\log n}$ for all sufficiently large n.

Summarizing all our estimates so far we arrive at

$$\begin{aligned}\alpha^{-1} \mu\Big(\mathbb{R}^\Lambda - \mathbb{R}^\Lambda_\alpha\Big) &\leq \sum_{n=1}^\infty \bigg(\sum_{A \in \mathcal{S},\, |A| = n} \frac{1}{\Delta([A])} \bigg) \bigg(\sum_{k \in \mathbb{Z}^n} \frac{1}{\Delta(|k|)} \bigg) \\ &\leq C + C \sum_{n=n_0}^\infty \frac{K^{n \log\log n}}{\Theta(t_n)}\end{aligned}$$

with some constant C and n_0 so large that $t_n \geq n \log^\sigma n$ for $n \geq n_0$ by hypotheses. Here we are still free to choose a suitable approximation function Θ, and choosing

$$\Theta(t) = \exp\left(\frac{t}{\log t \, \log^\sigma \log t}\right), \qquad t > e,$$

the infinite sum does converge. Thus there are approximation functions such that

$$\alpha^{-1}\mu\Big(\mathbb{R}^\Lambda - \mathbb{R}^\Lambda_\alpha\Big) < \infty.$$

This proves Theorem B.

A Approximation Functions

In [29] Rüssmann introduced the notion of an approximation function in order to characterize a large class of small divisors to which the KAM procedure is applicable. A similar characterization was already used by Brjuno [6] in his extension of Siegel's famous result on the linearization of complex mappings in the plane. Incidentally, those results do *not* rely on an iteration technique but on an ingenious application of the majorant method. See also [25] for a brief exposition of this method.

A nondecreasing function $\Delta\colon [0,\infty) \to [1,\infty)$ is called an *approximation function*, if

$$\frac{\log \Delta(t)}{t} \searrow 0, \qquad 0 \leq t \to \infty \tag{30}$$

and

$$\int^\infty \frac{\log \Delta(t)}{t^2}\, dt < \infty. \tag{31}$$

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

Obviously, any positive power of an approximation function is again an approximation function. So is the product of two such functions.

Given a characterization of small divisors in terms of an approximation function Δ their effect in a perturbation problem is described by two functions Γ_k and Ψ_k defined on the positive real axis in terms of Δ. For $k \geq 0$ and $1 < \kappa \leq 2$,

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

and

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

where the infimum is taken over all sequences $\rho_0 \geq \rho_1 \geq \cdots > 0$ such that $\rho_0 + \rho_1 + \cdots \leq 0$. The parameter κ is different for different kinds of small divisor problems. In our case, $\kappa = 3/2$.

Evidently, if Δ is an approximation function, then so is $(1+t)^k \Delta$ for any $k \geq 0$. We may therefore restrict our attention to the case $k = 0$, writing Γ and Ψ for Γ_0 and Ψ_0 respectively.

The supremum in the definition of Γ is attained and finite in view of condition (30). The infinite product in the definition of Ψ is lower semi-continuous when considered as a function on the set of sequences over which the infimum is taken endowed with the topology of pointwise convergence. Consequently, the infimum is also attained. For every $\rho > 0$, there exists a sequence $\rho_0^* \geq \rho_1^* \geq \cdots > 0$ whose sum is not bigger than ρ such that

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

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

Still, Ψ may be infinite for some $\rho > 0$. The following lemma which is essentially due to Rüssmann [29] rules that out.

Lemma 5. *The function Ψ is finite for all $\rho > 0$. Specifically, if*

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

then $\Psi(\rho) \leq e^{(\kappa-1)\rho T}$.

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

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

for $\nu \geq 0$. By the monotonicity hypothesis (30) we have $\rho_0 \geq \rho_1 \geq \cdots > 0$ and

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

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

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

by the definition of ρ_ν and hence

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

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

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

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

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

Lemma 6. *If Δ is sufficiently increasing, then $\Gamma_0(\rho) \leq \rho^k \Gamma_k(\rho)$ for $k \geq 0$.*

Proof. Again, let $\delta = \log \Delta$. If $\rho \leq \dfrac{1}{1+t}$, then

$$\frac{d}{dt}(\delta(t) - \rho t) \geq \frac{d}{dt}\delta(t) - \frac{1}{1+t} \geq 0.$$

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

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

as we wanted to show. ∎

Typical approximation functions are

(1)	$(1 + t/n)^n,$	$n \geq 1,$
(2)	$\exp(t^\alpha/\alpha),$	$0 < \alpha < 1,$
(3)	$\exp\left(\dfrac{t}{1 + \log^\gamma(1+t)}\right),$	$\gamma > 1,$

where n need not be an integer. They are also sufficiently increasing.

Lemma 7. *For example (1),*

$$\Psi(\rho) \le \left(\frac{1}{\delta\hat{\rho}}\right)^n,$$

with $\hat{\rho} = \min(\rho, 1)$ *and* $\delta = \dfrac{\kappa - 1}{4}$. *For example (2),*

$$\Psi(\rho) \le \exp\left(\frac{1}{\alpha \log 2}\left(\frac{1}{\delta\rho}\right)^{\alpha/(1-\alpha)}\right)$$

with $\delta = (1-\alpha)(\kappa-1)\log 2$. *And for example (3),*

$$\Psi(\rho) \le \exp\left((\kappa-1)\rho\, \exp\left(\left(\frac{1}{\delta\rho}\right)^{1/(\gamma-1)}\right)\right)$$

with $\delta = (\gamma - 1)\log\kappa$.

Proof. For the first example, one easily finds $\Gamma(\rho) \le \hat{\rho}^{-n}$ by distinguishing the cases $\rho \le 1$ and $\rho > 1$. Choosing the sequence $\rho_\nu = \kappa_\nu \hat{\rho}$ and recalling that

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

we then obtain

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

Since

$$\kappa^{\kappa/(\kappa-1)} = \left(1 + \frac{1}{\mu}\right)^{\mu+1}\Bigg|_{\mu=1/(\kappa-1)} \le \left(1+\frac{1}{\mu}\right)^{\mu+1}\Bigg|_{\mu=1} = 4,$$

the estimate follows as claimed.

Considering the second example, a straightforward calculation shows that $\Gamma(\rho) = \exp(\rho^{-\tilde{\alpha}}/\tilde{\alpha})$ with $\tilde{\alpha} = \alpha/1 - \alpha$. Choosing the geometric sequence $\rho_\nu = \tilde{\kappa}_\nu \rho$, where $\tilde{\kappa}_\nu$ is defined analogously to κ_ν using $\tilde{\kappa} = \kappa^{1-\alpha}$ we obtain

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

with

$$\sum_{\nu=0}^{\infty} \frac{\kappa_\nu}{\tilde{\kappa}_\nu^{\tilde{\alpha}}} = \frac{\kappa - 1}{(\tilde{\kappa} - 1)^{\tilde{\alpha}}(\kappa\tilde{\kappa}^{-\tilde{\alpha}} - 1)} = \frac{\kappa - 1}{(\kappa^{1-\alpha} - 1)^{1/(1-\alpha)}}.$$

Since

$$\kappa^{1-\alpha} - 1 \geq (1 - \alpha)\log\kappa \geq (1 - \alpha)(\kappa - 1)\log 2$$

for $0 < \alpha < 1$ and $1 < \kappa \leq 2$, the last term is bounded from above by one over $((1 - \alpha)\log 2)^{\tilde{\alpha}+1}(\kappa - 1)^{\tilde{\alpha}}$, which gives the desired result.

As to the third example, we follow Rüssmann and apply Lemma 5. We have

$$\begin{aligned}\int_T^\infty \frac{\log \Delta(t)}{t^2}\,dt &= \int_T^\infty \frac{dt}{t(1 + \log^\gamma(1 + t))} \\ &\leq \int_T^\infty \frac{dt}{t\log^\gamma t} = \frac{1}{(\gamma - 1)\log^{\gamma-1} T}.\end{aligned}$$

Choosing T so that

$$\log^{\gamma-1} T = \frac{1}{(\gamma - 1)\log\kappa} \cdot \frac{1}{\rho},$$

the hypothesis of this lemma is satisfied, and the estimate follows. ∎

The result for the first example may be refined for $\rho > 1$. Let $\rho = l + \bar{\rho}$ with an integer $l \geq 0$ and $0 < \bar{\rho} \leq 1$. Choosing $\rho_\nu = 1$ for $0 \leq \nu < l$ and otherwise optimal with respect to $\bar{\rho}$, you get

$$\Psi(\rho) \leq \prod_{\nu=l}^{\infty} \Gamma(\rho_\nu)^{\kappa_\nu} = \prod_{\nu=0}^{\infty} \Gamma(\rho_{l+\nu})^{\kappa_\nu/\kappa^l} = \Psi(\bar{\rho})^{1/\kappa^l}.$$

Hence, for the first example one has more generally

$$\Psi(\rho) \leq \left(\frac{\kappa - 1}{4\bar{\rho}}\right)^{n/\kappa^l} \qquad \text{for} \quad \rho = l + \bar{\rho}.$$

This case is of interest in finite dimensional problems where the perturbation consists of entire functions such as in [39].

B The Cauchy Inequality

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

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

is its induced operator norm.

Lemma 8 (Generalized Cauchy Inequality). *Let F be an analytic map from the open ball of radius r around v in A into B such that $|F|_B \leq M$ on this ball. Then*

$$\|d_vF\|_{B,A} \leq \frac{M}{r}.$$

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

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

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

The statement of the lemma is particularly transparent, when F is a complex valued function. Then d_vF is an element in the dual space A^* to A, and the induced operator norm is the norm $|\cdot|_{A^*}$ dual to $|\cdot|_A$. So, for instance, if F is bounded in absolute value by M on the balls

$$|v|_\infty, \quad |v|_2, \quad |v|_1 \quad < \quad r,$$

then

$$|d_0F|_1, \quad |d_0F|_2, \quad |d_0F|_\infty \quad < \quad \frac{M}{r}$$

respectively in both finite and infinite dimensional settings.

C Poisson Bracket and Transformation

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

First we need an estimate of the norm of the Poisson bracket of two functions that is more general than the one stated in the KAM-step.

Lemma 9. *Suppose that for some* $v \geq w$*,*

$$\rho_0^{-1} |||F|||_{v,r_0-\rho_0,s_0}, \ \sum_\lambda \left|\left|\left| F_{\theta_\lambda} \right|\right|\right|_{v,r_0-\rho_0,s_0} \leq M.$$

Then

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

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

Proof. Proceeding just as in the KAM-step we have

$$\left\| \left\langle F_{A,I}, G_{B,\theta} \right\rangle \right\|_{r-\rho,s-\sigma} \leq \frac{1}{s_0 - s + \sigma} \cdot \frac{1}{e\rho} \, e^{w[A \cap B]} \|F_A\|_{r_0-\rho_0,s_0} \|G\|_{r,s}$$

for $s - \sigma < s_0$, $r - \rho \leq r_0 - \rho_0$ and, of course, $0 < \rho < r$ and $0 < \sigma < s$. Consequently,

$$|||\langle F_I, G_\theta \rangle|||_{v,r-\rho,s-\sigma} \leq \frac{1}{s_0 - s + \sigma} \cdot \frac{\rho_0}{e\rho} \, M \, |||G|||_{v,r,s}.$$

Similarly,

$$\left\| \left\langle F_{A,\theta}, G_{B,I} \right\rangle \right\|_{r-\rho,s-\sigma} \leq \frac{1}{\sigma} \, e^{w[A \cap B]} \|G_B\|_{r,s} \sum_{\lambda \in A} \left\| F_{A,\theta_\lambda} \right\|_{r_0-\rho_0,s_0}$$

and consequently

$$\begin{aligned} \|\langle F_\theta, G_I \rangle\|_{v,r-\rho,s-\sigma} &\leq \frac{1}{\sigma} |||G|||_{v,r,s} \sum_A \sum_{\lambda \in A} e^{v[A]} \left\| F_{A,\theta_\lambda} \right\|_{r_0-\rho_0,s_0} \\ &\leq \frac{1}{\sigma} |||G|||_{v,r,s} \sum_\lambda \left|\left|\left| F_{\theta_\lambda} \right|\right|\right|_{v,r_0-\rho_0,s_0} \\ &\leq \frac{1}{\sigma} M |||G|||_{v,r,s}. \end{aligned}$$

The result follows. ∎

Lemma 10. *Suppose that for some* $v \geq w$,

$$\rho_0^{-1} |||F|||_{v,r_0-\rho_0,s_0}, \sum_{\lambda} \left|\left|\left| F_{\theta_\lambda} \right|\right|\right|_{v,r_0-\rho_0,s_0} \leq M < \frac{s}{8}.$$

Then

$$|||G \circ \Phi|||_{v,r-\rho,s/2} \leq \frac{1}{1-8M/s} |||G|||_{v,r,s}$$

for $0 < \rho_0 \leq \rho < r \leq \rho_0 - s_0$ *and* $0 < s \leq s_0/2$, *where* Φ *denotes the time-1-map of the hamiltonian vectorfield* X_F.

The hypotheses of the lemma imply that

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

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

Proof. Consider the Lie series expansion

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

where

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

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

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

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

$$\left|\left|\left| \mathrm{ad}_F^h G \right|\right|\right|_{v,r-h\rho,s-h\sigma} \leq \left(\frac{1}{\sigma} + \frac{\rho_0}{es\rho} \right)^h M^h |||G|||_{v,r,s}.$$

Replacing ρ, σ by $\rho/h, s/2h$ respectively and using the assumption $\rho_0 \leq \rho < r$ this yields

$$\left|\left|\left|\mathrm{ad}_F^h G\right|\right|\right|_{v,r-\rho,s/2} \leq \left(\frac{8Mh}{es}\right)^h |||G|||_{v,r,s}.$$

By Stirling's formula, $\dfrac{h^h}{h!} \leq e^h$ for $h \geq 1$. Hence,

$$\begin{aligned} |||G \circ \Phi|||_{v,r-\rho,s/2} &\leq \sum_{h\geq 0} \frac{1}{h!} \left|\left|\left|\mathrm{ad}_F^h G\right|\right|\right|_{v,r-\rho,s/2} \\ &\leq \sum_{h\geq 0} \left(\frac{8M}{s}\right)^h |||G|||_{v,r,s} \\ &= \frac{1}{1-8M/s} |||G|||_{v,r,s}, \end{aligned}$$

provided that $8M < s$. ∎

D An Inverse Function Theorem

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

Lemma 11. *Suppose f is real analytic from $\mathcal{W}_h$ into $\mathbb{C}^A$. If*

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

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

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

on this domain.

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

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

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

$$\theta = \max \|\partial f - I\|_\infty < \delta/k \le 1$$

by Cauchy's inequality and so

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

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

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

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

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

E More Measure Estimates

This appendix provides the measure estimates for the example concerning finite chains of oscillators. Let

$$d\mu(\omega) = \prod_{i=1}^{N} \frac{a_i}{\sqrt{2\pi}} \, e^{-a_i^2 \omega_i^2/2} \, d\omega_i, \qquad a_i = \log i, \quad i \ge 2.$$

The weights are chosen so that for infinite N we have a gaussian probability measure $\tilde{\mu}$ on $\mathbb{R}^{\mathbb{N}}$ with support at the origin. Conversely, μ is the "projection" of $\tilde{\mu}$ onto $\mathbb{R}^N$ obtained by "integrating out" the extra dimensions.

Proceeding as in the proof of Lemma 1 we obtain

$$\alpha^{-1}\mu(\mathcal{R}_k) \le \frac{\log N}{\|k\|} \cdot \frac{1}{\Delta([\![k]\!])\Delta(|k|)}$$

for the k-th resonance zone $\mathcal{R}_k$, and consequently

$$\alpha^{-1}\mu(\mathbb{R}^N - \mathbb{R}^N_\alpha) \le \sum_{A\in\mathcal{I}} \sum_{\substack{k\neq 0\\ \operatorname{supp} k\subseteq A}} \alpha^{-1}\mu(\mathcal{R}_k)$$
$$\le N\log N \sum_{n=2}^{N} \frac{1}{\Delta([n])} \sum_{0\neq k\in\mathbb{Z}^n} \frac{1}{\|k\|\Delta(|k|)},$$

where $[n]$ stands for the weight of any interval of length n.

Now choose

$$\Delta(t) = \frac{D(t)}{1+t}, \qquad D(t) = (1+t/\tau)^\tau, \qquad \tau > N,$$

and recall that the function D is monotonically increasing in τ for all $t \ge 0$. By Lemma 3,

$$\sum_{k\in\mathbb{Z}^n} \frac{1}{(1+|k|)\Delta(|k|)} \le 2^n \int_0^\infty \binom{n+t}{n} \frac{d\log D(t)}{D(t)}$$
$$= 2^n \int_0^\infty \binom{n+t}{n} \frac{dt}{(1+t/\tau)^{\tau+1}}.$$

The binomial equals $(1+t/1)(1+t/2)\cdots(1+t/n)$, while the denominator is bounded from below by

$$(1+t/\tau+1)^{\tau+1} \ge (1+t/n+\sigma)^{n+\sigma}$$

for $2 \le n \le N$ with $\sigma = \min(2, 1+\tau-N) > 1$. Moreover,

$$\left(1+\frac{t}{n+\sigma}\right)^{-n} \prod_{k=1}^{n}\left(1+\frac{t}{k}\right) \le \prod_{k=1}^{n} \frac{n+\sigma}{k} \le \frac{e^\sigma n^n}{n!} \le e^{n+\sigma}$$

by Stirling's formula. Thus,

$$\sum_{k\in\mathbb{Z}^n} \frac{1}{(1+|k|)\Delta(|k|)} \le 2^n e^{n+\sigma} \int_0^\infty \left(1+\frac{t}{n+\sigma}\right)^{-\sigma} dt$$
$$= 2^n e^{n+\sigma} \frac{n+\sigma}{\sigma-1}.$$

Together with $\|k\| \ge |k| / \sqrt{|\operatorname{supp} k|}$ for $k \neq 0$ we obtain

$$\alpha^{-1}\mu(\mathbb{R}^N - \mathbb{R}^N_\alpha) \le B_N \sum_{n=2}^{N} \frac{2^n e^n n^2}{\Delta([n])}$$
$$\le B_N \sum_{n=2}^{N} \left(\frac{2e}{1+[n]/n}\right)^n n^2(1+[n])$$

with $B_N = CN\log N/\min(1, \tau - N)$. Choosing $[A] = f\,|A| - f$ with $f > 2e-1$ we obtain a uniform bound of the last sum for all N and so

$$\alpha^{-1}\mu\left(\mathbb{R}^N - \mathbb{R}^N_\alpha\right) = O(N\log N)$$

uniformly in N.

References

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

[2] V.I. ARNOLD, *Small denominators and problems of stability of motions in classical and celestial mechanics.*, Uspehi. Math. Nauk. **18** (1963), 96–196; Russian Math. Surveys **18** (1963), 85–193.

[3] V.I. ARNOLD, *Mathematical Methods of Classical Mechanics*, Springer, New York, Heidelberg, Berlin, 1978.

[4] V.I. ARNOLD (ED), *Dynamical Systems III*, Encyclopaedia of Mathematical Sciences Volume 3, Springer, Berlin, Heidelberg, New York, 1988.

[5] J.-B. BOST, *Tores invariants des systèmes dynamiques hamiltoniens*, Séminaire Bourbaki, Astérisque **133–134** (1986), 113–157.

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

[7] L. CHIERCHIA, G. GALLAVOTTI, *Smooth prime integrals for quasi-integrable Hamiltonian systems*, Il Nuovo Cimento **67 B** (1982), 277–295.

[8] L.H. ELIASSON, *Perturbations of stable invariant tori for hamiltonian systems*, Ann. Sc. Norm. Sup. Pisa **15** (1988), 115–147.

[9] L.H. ELIASSON, *Compensation of signs in a small divisor problem*, Typoscript, 1988.

[10] J. FRÖHLICH, T. SPENCER, C.E. WAYNE, *Localization in disordered, nonlinear dynamical systems*, Jour. Stat. Phys. **42** (1986), 247–274.

[11] A.B. KATOK, *Ergodic properties of degenerate integrable hamiltonian systems*, Math. USSR Izv. **7** (1973), 185–214.

[12] A.N. KOLMOGOROV, *On the conservation of conditionally periodic motions for a*

small change in Hamilton's function, in Russian, Dokl. Akad. Nauk SSSR **98** (1954), 527–530.

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

[14] S.B. Kuksin, *Perturbation of quasiperiodic solutions of infinite dimensional hamiltonian systems*, Math. USSR Izv. **32** (1989), 39–62.

[15] L. Markus, K.R. Meyer, *Generic hamiltonian systems are neither integrable nor ergodic*, Memoirs of the American Mathematical Society **144** (1974), 1–52.

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

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

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

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

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

[21] Jürgen Moser, *Is the solar system stable?*, The Mathematical Intelligencer **1** (1978), 65–71.

[22] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Celeste II*, Gauthier-Villars, 1892.

[23] J. Pöschel, *Über invariante Tori in differenzierbaren Hamiltonschen Systemen*, Bonn. Math. Schr. **120** (1980), 1–103.

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

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

[26] J. Pöschel, *A general infinite dimensional KAM-theorem*, IXth International Congress on Mathematical Physics, B. Simon, A. Truman, I.M. Davies (Editors), Adam Hilger, Bristol, New York, 1989.

[27] J. Pöschel, *On elliptic lower dimensional tori in hamiltonian systems*, Math. Z. (1989).

[28] R.C. Robinson, *Generic properties of conservative systems*, Amer. J. of Math. **92** (1970), 562–603.

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

[30] H. Rüssmann, *Non-degeneracy in the perturbation theory of integrable dynamical systems*, Number Theory and Dynamical Systems, M.M. Dodson, J. Vickers (Editors), Cambridge University Press, Cambridge, 1989.

[31] D. Salamon, *The Kolmogorov-Arnold-Moser theorem*, Preprint, Forschungsinstitut für Mathematik, ETH Zürich, 1986.

[32] D. Salamon, E. Zehnder, *KAM theory in configuration space*, Comment. Math. Helv. **64** (1989), 84–132.

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

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

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

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

[37] C.E. Wayne, *The KAM theory of systems with short range interactions, I and II*, Commun. Math. Phys. **96** (1984), 311–329 and 331-344.

[38] C.E. Wayne, *On the elimination of non-resonance harmonics*, Commun. Math. Phys. **103** (1986), 351–386.

[39] C.E. Wayne, *Bounds on the trajectories of a system of weakly coupled rotators*, Commun. Math. Phys. **104** (1986), 21–36.

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

[41] J.C. Yoccoz, *Linéarisation des germes de difféomorphismes holomorphes de* $(\mathbb{C},0)$, C. R. Acad. Sci. Paris **306** (1988), 55–58.

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