Commun. Math. Phys. 192, 145 – 168 (1998)

Communications in
Mathematical Physics

A KAM Theorem for Hyperbolic-Type Degenerate Lower Dimensional Tori in Hamiltonian Systems ⋆

Jiangong You

Department of Mathematics, Nanjing University, Nanjing 210093, P.R. China

Received: 23 October 1996 / Accepted: 24 June 1997

Abstract: A KAM theorem for degenerate lower dimensional tori in nearly integrable Hamiltonian systems is given in this paper. For the non-degenerate cases, both hyperbolic and elliptic, the KAM theorem has been well established by many authors ([8, 9, 11, 13, 14, 17]).

1. Introduction and Result

Consider a real analytic Hamiltonian

$$H(x, y, u, v) = h_0(y, u, v) + P(x, y, u, v) \tag{1.1}$$

in a complex neighbourhood in $C^{2n} \times C^{2m}$ with the symplectic structure $\sum_{i=1}^{n} dx_i \wedge dy_i + \sum_{i=1}^{n} du_i \wedge dv_i$, of a $2n$ dimensional real domain $u = v = 0, (x, y) \in T^n \times D \subset T^n \times R^n$, where D is an open set of R^n. Denote by $z = (u_1, \cdots, u_m, v_1, \cdots, v_m) \in R^{2m}$ for simplicity.

If $\frac{\partial h_0}{\partial z}(y, 0) = 0$, the unperturbed Hamiltonian system defined by h_0 possesses a $2n$ dimensional invariant subspace $u = v = 0$ foliated by a family of invariant tori $y = y_0, u = v = 0$ and the flow on each torus is given by $x(t) = x_0 + \frac{\partial h_0(y_0,0)}{\partial y} t$.

If $det \dfrac{\partial^2 h_0(y, 0)}{\partial y^2} \neq 0$, i.e., if h_0 is non-degenerate, the frequencies $(\omega_1, \cdots, \omega_n) = \dfrac{\partial h_0(y, 0)}{\partial y}$ can be regarded as parameters and one can equivalently consider perturbations of a family of linear integrable Hamiltonians, parameterised by the frequencies $\omega = \frac{\partial h_0(y,0)}{\partial y} \in \mathcal{O} \subset R^n$ with positive Lebesgue measure,

---

⋆ The work was supported by AvH Foundation in Germany.

$$H = N + P = < \omega, y > + \frac{1}{2} < A(\omega)z, z > + P, \tag{1.2}$$

in $(x, y, z) \in T^n \times R^n \times R^{2m}$, where $A(\omega) = \frac{\partial^2 h_0^2}{\partial z^2}((\frac{\partial h_0}{\partial y})^{-1}(\omega), 0)$ is a $2m \times 2m$ matrix. This setting has been frequently used by many authors. In the following we will treat ω as independent parameters varying over a positive measure set $\mathcal{O}$.

The persistence of invariant tori for the perturbed Hamiltonians have been extensively studied in the case that A is non-degenerate, i.e., $det \frac{\partial^2 h_0}{\partial z^2}(y, 0) \neq 0$.

If all the eigenvalues of A are not on the imaginary axis, the torus is called hyperbolic. For this case, if $\omega = (\omega_1, \cdots, \omega_n) \in \mathcal{O}$ satisfies the Diophantine condition

$$| < k, \omega > | > \gamma |k|^{-\tau}, \tag{1.3}$$

for $\tau > n + 1, 0 \neq k \in Z^n$, Moser[14], Graff[8] and Zehnder[26] had proved that, there is a ω^* such that (1.2) at ω^* possesses an invariant torus with prescribed inner frequencies ω if perturbations are sufficiently small. Back to the original system (1.2), it follows that (1.1) has a Cantor family of invariant tori if the perturbation is smooth and small.

If all the eigenvalues of $A(\omega)$ belong to $iR^1 \setminus 0$, the torus is called elliptic. More precisely, for the following system

$$H = N + P = \sum_{i=1}^{n} \omega_i y_i + \frac{1}{2} \sum_{j=1}^{m} \Omega_j(\omega)(u_j^2 + v_j^2) + P, \tag{1.4}$$

Melnikov ([13]) in 1967 announced that, for a positive Lebesgue measure subset $\mathcal{O}_\gamma \subset \mathcal{O}$, $(1.4)_{\omega \in \mathcal{O}_\gamma}$ possesses a n dimensional invariant torus with frequencies satisfying the non-resonant conditions

$$| < k, \tilde{\omega} > + < l, \tilde{\Omega} > | > \gamma |k|^{-\tau}, \quad |l| \leq 2 \tag{1.5}$$

for $k \in Z^n, l \in Z^m, |k| + |l| \neq 0, \tau > n + 1$, where $(\tilde{\omega}, \tilde{\Omega}) = (\tilde{\omega}_1, \cdots, \tilde{\omega}_n, \tilde{\Omega}_1, \cdots, \tilde{\Omega}_m)$, provided the perturbation is sufficiently small. The complete proof was carried out fifteen years later by Eliasson, Kuksin and Pöschel ([9, 11, 17]). In this case, only the measure estimate is available. One can't tell if (1.4) has a torus with prescribed frequencies.

More recently, developing Craig and Wayne's method [7], Bourgain [3] proved the existence of quasi-periodic solutions for various Hamiltonian PDEs under the first Melnikov non-resonant condition, i.e., (1.5) holds for $|l| \leq 1$. The result was applied to some kinds of PDEs with periodic boundary condition. His proof is based on the Liapounov-Schmidt reduction introduced by [7], and some sophisticated estimates of controling the inverse of matrices with singular sites.

The remaining question is what happens when A contains zero eigenvalues? In this paper, we shall consider the simplest case of this degenerate problem: $m = 1$, the spectrum Ω of A vanishes and ω satisfies (1.3). In this case the invariant torus T^n of the unperturbed system is resonant, and it may be eliminated by some special perturbations, for example $P = \epsilon(u^3 + u)$, if no further assumption is presented. Thus if we do not want to impose further restriction on the perturbation except the smallness and smoothness, the higher order terms of the unperturbed integrable system have to be taken into account. The simplest degenerate case is the nilpotent case (See Takens [22]), i.e.,

$$H = < \omega, y > + \frac{1}{2} v^2 + f(u, v) + P(x, y, u, v, \omega), \tag{1.6}$$

in $T^n \times R^n \times R^2$. If $T^n \times \{0\}$ is an isolated torus of the unperturbed system $(1.6)_{P=0}$, it follows that $f(u,v) = o(|u|^2 + |v|^2) \not\equiv 0$. By normal form theory, in this case, we can assume that $f(u,v) = u^3 G(u)$, where $G(u) \not\equiv 0$ is a polynomial of u. The case with leading terms $G(u) = cu^{2d-4} \not\equiv 0 (d = 2, 3, \cdots)$ is not of interest in this paper, because in this case the torus can be eliminated by a simple perturbation, $P = \epsilon cu$. Thus we consider the case with leading terms $G(u) = cu^{2d-3} (d \geq 2)$. Concerning the unperturbed system, there are two different types of n-dimensional invariant tori:

1. $c < 0$, which we call a hyperbolic type degenerate torus. In this case the origin in the (u,v) plane is a saddle-type degenerate singular point. This is the main subject of this paper.

2. $c > 0$, which we call an elliptic type degenerate torus. The origin in the (u,v) plane is then a degenerate centre. This case is more complicated. The persistence result has not been available. We shall give more comments for this case to show where is the problem.

In this paper, we consider the persistence of an hyperbolic-type degenerate lower dimensional torus for a real analytical, perturbed integrable Hamiltonian system continuously depending on the parameter ω in an open set $\mathcal{O} \subset R^n$. Without loss of generality, we normalise c as -1 and consider the Hamiltonian of the following form:

$$H = < \omega, y > + \frac{1}{2} v^2 - u^{2d} + P(x,y,u,v), \quad d \geq 2, \tag{1.7}$$

in (x,y,u,v)-space $T^n \times R^n \times R^2$ with standard symplectic structure $dx \wedge dy + du \wedge dv$.

The goal of this paper is to prove, for any ω_0 satisfying (1.3), there is a ω^* such that (1.7) at ω^* possesses a n dimensional invariant torus carrying rotational flow of frequencies ω_0, provided that the perturbation P is analytic and small enough. The obtained torus might be still degenerate, but it is saddle-like in the (u,v) plane.

From now on, we fix a ω_0 satisfying (1.3) and consider the complex extension of the real Hamiltonian (1.7) on the complex neighbourhood

$$D(r,s,s_u) = \{(x,y,u,v) \mid |Imx| < r, \ |y| < s^{2d}; |v| < s^d, |u| < s_u\},$$

with parameter ω in

$$B(\omega_0, \delta) = \{\omega \in R^n : |\omega - \omega_0| \leq \delta\},$$

where Imx is the image part of x and $|\cdot|$ is the $sup-$norm for complex vectors, $s \geq s_u > 0$.

The norm of P on $D(r,s,s_u) \times B$ is defined as

$$||P||_{D \times B} = \sup_{D \times B} |P|.$$

Theorem 1.1. *Suppose that Hamiltonian (1.7) is analytic in $(x,y,u,v) \in D$ and continuous in $\omega \in B$. Moreover,*

$$\frac{1}{s^{2d}} ||P||_{D(r,s,s) \times B} \leq \epsilon_0 \leq \frac{1}{2} \delta,$$

where ϵ_0 is a small constant depending on n, d, r, γ, τ. Then there is a $\omega^* \in B$ with $|\omega^* - \omega_0| < 2\epsilon_0$ such that H at ω^* possesses a n dimensional invariant tori carrying rotational flow with internal frequencies ω_0.

As a consequence of Theorem 1.1, we have the following

Theorem 1.2. *Suppose that the analytic Hamiltonian*

$$H = h(y) + \frac{1}{2}v^2 - u^{2d} + \epsilon P(x, y, u, v), \tag{1.8}$$

is defined on $\{|Imx| < r\} \times \{y \in D \subset C^n\} \times \{|u|, |v| < s\}$*. Then there is a* $\epsilon^*(n, d, D, s, r, \gamma, \tau)$ *such that if* $\epsilon \leq \epsilon^*$*,* $\frac{\partial h}{\partial y}(y_0)$ *satisfies (1.3) and* $dist\{y_0, D\} \geq M\epsilon^*$ *with* $M = \sup_D |\frac{\partial h}{\partial y}|$*, Hamiltonian (1.8) has an invariant torus carrying rotational flow of frequencies* $\frac{\partial h}{\partial y}(y_0)$*.*

Remark. For general perturbations, the degenerate torus $T^n \times \{y = u = v = 0\}$ of the unperturbed system will break into $2d - 1$ non-degenerate tori. But we can not expect more than one if we do not present a further restriction on the perturbation besides smoothness and smallness. Moreover the normal behaviour of the obtained torus can not be predetermined; it depends on the perturbation.

Remark. The analyticity is not necessary but it considerably simplifies the proof.

Example. The theorem applies to the Duffing equation

$$\frac{d^2x}{dt^2} - x^3 = \epsilon p(\omega_1 t, \cdots, \omega_n t, x), \tag{1.9}$$

for proving the existence of quasi-periodic solutions with the frequencies $(\omega_1, \cdots, \omega_n)$. Similar results have been obtained in the pioneering work of Moser ([15]) for the equation with a linear term $a(\epsilon)x$, where $a(\epsilon) \approx 1$ depending on the perturbation is an artificial parameter.

I'd like to mention here, in a different setting, C. Cheng ([5]) considered the perturbation of the resonant tori in Hamiltonian systems. His result is a partial generalisation of the Poincaré-Birkhoff fixed point theorem to the invariant torus case. As far as the author knows, that is the first result dealing with the resonant problem without adding any restriction on the perturbation.

2. Outline of the Proof

The theorem will be proved by KAM iteration which involves an infinite sequence of coordinate transformations. The procedure is more complicated when we consider the normally resonant case. In the following, we outline the proofs.

Let $B_n = \{\omega, |\omega - \omega_0| \leq \frac{\gamma}{K_n^{\tau+1}}\}$. As we will see, at each step of the KAM scheme, a family of Hamiltonian

$$H_n = N_n + P_n$$

defined in $D(r_n, s_n, s_{un})$ with parameter $\omega \in B_n$ is considered near a n dimensional torus $\{y = 0, u = v = 0\}$, where

$$N_n =< \omega_0, y_n > + \frac{1}{2}v_n^2 + f_n(u_n),$$

with $f_n(u_n) = \sum_{i=2}^{i_n^*} a_i^n u_n^i, 2 \leq i_n^* \leq 2d$, $a_i^n < 2$ are small constants coming from the perturbation. The Hamiltonians satisfy

$$\frac{1}{s_n^2}||P_n||_{D(r_n,s_n,s_{un})} \le \epsilon_n,$$

$$-3s_n^{2d} \le f_n(u_n) \le \epsilon_n s_n^{2d}, \text{for } u \in [-s_{1n}, s_{1n}], -3s_n^{2d} \le f_n(s_{1n}), f_n(s_{2n}) \le -s_n^{2d} \tag{2.1}$$

for some s_{1n}, s_{2n} satisfying $-s_{un} \le s_{2n} < 0 < s_{1n} \le s_{un}$.

We shall prove that there is a symplectic change of variables,

$$\Psi_n : (x_{n+1}, y_{n+1}, u_{n+1}, v_{n+1}) \to (x_n, y_n, u_n, v_n);$$

defined in a smaller domain D_{n+1} with parameter in B_{n+1}, such that $P_n \circ \Phi_n = N_{n+1} + P_{n+1}$ in $D_{n+1} \times B_{n+1}$ satisfies $|P_{n+1}|_{D_{n+1}\times B_{n+1}} \le |P_n|^{\kappa}_{D_n \times B_n}, \kappa > 1$. Moreover, there is a continuous map $\phi_n^{-1} : \omega_{n+1} \to \omega_n$ which maps B_{n+1} into B_n.

In what follows the Hamiltonian without subscription denotes the Hamiltonian in n^{th} step, while those with subscription + denotes the Hamiltonian of $n + 1^{\text{th}}$ step. We outline one step of the KAM iteration.

We first truncate the perturbations and keep the higher order terms in the next iteration step since it can be made smaller by shrinking the definition domain. More precisely, rewrite H as the following:

$$H = N + R + (P - R),$$

where R is a higher order truncation of P such that $||P - R||$ is less than $||P||^{\kappa}_D, \kappa > 1$ in a smaller domain (see the next section for details).

Then we find a symplectic coordinate transformation to kill as many terms in R as possible. The transformation is generated by a Hamiltonian function F defined in a smaller domain $D(r - \rho, \frac{1}{8}s, \frac{1}{8}s_u)$. Let X_F be the vector field with the Hamiltonian F. Denote by X_F^t the flow of X_F and $\Phi = X_F^{t=1}$ the time 1 map of the flow. By a Taylor series, we have (see [12, 17]),

$$\begin{aligned} H \circ \Phi &= (N + R) \circ \Phi + (P - R) \circ \Phi \\ &= (N + R) \circ X_F^{t=1} + (P - R) \circ X_F^{t=1} \\ &= N + R + \{N, F\} \\ &\quad + \int_0^1 (1 - t)\{\{N + R, F\}, F\} \circ X_F^t dt + \{R, F\} + (P - R) \circ \Phi. \\ &= N + R + \{N, F\} + \bar{P} \qquad (2.2) \\ &= N + R + \{N_2, F\} + \{N_h, F\} + \bar{P} \\ &= \bar{N} + \{N_h, F\} + \bar{P}, \qquad (2.3) \end{aligned}$$

where

$$N_2 =< \omega, y > + \frac{1}{2}v^2 + a_2u^2, \; N_h = \sum_{i=3}^{i^*} a_i u^i, i^* \le 2d,$$

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

The philosophy of the KAM method ([1, 2, 10, 14]) is to find a special F defined in a shrunk domain which makes the new perturbation $\bar{P}$ in (2.2) much smaller and $N + R + \{N, F\}$ a new normal form. In the non-degenerate case, i.e., (1.5) is satisfied, we need not put the higher order terms of u into the normal form, i.e., $N_h = 0$, and F is obtained by solving a linear partial differential equation

$$N + R + \{N, F\} = N_+, \tag{2.4}$$

with a N_+ similar to N, where

$$\{N, F\} = \frac{\partial F}{\partial x}\frac{\partial N}{\partial y} + \frac{\partial F}{\partial v}\frac{\partial N}{\partial u} - \frac{\partial F}{\partial u}\frac{\partial N}{\partial v}.$$

In the non-degenerate case, N_+ can be put into a normal form simultaneously.

In the degenerate case, N consists of the higher order terms of u, and Eq. (2.4) is no longer a linear partial differential equation. We can not solve it completely as in the non-degenerate case. Note that the purpose of solving (2.4) is to find a function F so that (2.3) becomes a new normal form with a smaller perturbation. For this sake, we establish the KAM step as follows:

Firstly, instead of solving (2.4), we solve

$$\{N_2, F\} + R - \int_0^{2\pi} R\, dx = 0, \tag{2.5}$$

and treat $\{N_h, F\}$ as a part of the new perturbation. The trouble is $\bar{N} = N + \int R\, dx$ is no longer a normal form (see (3.16) for the precise formulation). Especially, it contains the linear term in u, and thus $T^n \times \{0\}$ might be no longer an invariant torus of $\bar{N}$. However, due to the existence of higher order terms, there does exist a hyperbolic type torus in the effective domain corresponding to a hyperbolic equilibrium in the (u, v) plane.

Secondly, we find a linear coordinate transformation to move the origin of the (u, v) plane to the hyperbolic equilibrium, which makes $\bar{N}$ a normal form, say $\tilde{N}$(containing no linear terms). As a matter of fact, the linear transformation does not smoothly but only continuously depend on parameters even if the original Hamiltonians smoothly depend on parameters. This observation means the method doesn't apply to the elliptic degenerate case which requires measure estimates. For the hyperbolic case, however, the measure estimate is not necessary since we can fix the frequency vector at each KAM step.

We have done one cycle of iteration if we could prove the new perturbation is smaller. This can be done by shrinking the definition domain. More precisely, we have to shrink the domain so that $\{N_h, F\}$ and $P - R$ are smaller. Certainly, this is available if the domain is sufficiently small. But it has no meaning if we don't care about the position of the torus of $\bar{N}$. If all tori of $\bar{N}$ are outside of the shrunk domain, we would lose an object at the next iteration step. Thus we have to balance the above two conflicting requirements and find a domain such that it does contain a hyperbolic torus of $\bar{N}$; meanwhile, $\{N_h, F\}$ and $P - R$ are smaller in this domain.

Obviously, the frequencies of the invariant torus of $\tilde{N}$ differs from that of N a little bit. After a KAM step, the frequencies of $\tilde{N}$ are $\omega_+ = \phi(\omega) = \omega + P_{0100}(\omega)$ for the Hamiltonian H at ω. However, since $P_{0100}(\omega)$ is continuous and small, $\phi(B)$ still contains a smaller ball B_+ centred at ω_0 and the new perturbation is uniformly smaller in B_+, so that we can do the next iteration.

Besides, there is an additional technical difficulty. We have to discard some higher order terms in $\tilde{N}$ at each iteration step since the coefficients of those terms might be big enough to destroy the convergence of the normal form series.

Iterating the above step, we get a family of symplectic change of variables Ψ^n and a family of Hamiltonian H_n defined in nested domains $\mathcal{B}_n \subset B$ such that $(D\Psi^n)^* X_H \circ \Psi^n = X_{N_n} + X_{P_n}$. By passing to the limit, it follows that H at $\omega^* \in \cap \mathcal{B}_n$ is conjugated to an integrable Hamiltonian system N_∞ which has a n torus carrying rotational flow of frequencies ω_0. That means H at $\omega^* \in \cap \mathcal{B}_n$ has a invariant torus of frequencies ω_0. We refer to Sect. 4 for more details.

In the next section we describe one KAM step in detail. Throughout this paper, we denote by c the constants which depend only on $n, d, \gamma, \delta, \tau$.

3. KAM Step

In what follows, the Hamiltonian without subscription denotes the Hamiltonian in the ν^{th} step, while those with subscription + denote the Hamiltonian of the $\nu + 1^{\text{th}}$ step. We consider one step of KAM iteration in full details.

Throughout this section, we consider a family of Hamiltonians

$$H = N + P, \tag{3.1}$$

defined in $D(r, s, s_u), \omega \in B(\omega_0, \frac{1}{2}\gamma K^{-\tau-1})$ with $\frac{1}{2}s^d \le s_u \le s$, where

$$N = N_2 + N_h = < \omega, y > + \frac{1}{2}v^2 + a_2 u^2 + \sum_{i=3}^{i^*} a_i u^i, \quad 3 \le i^* \le 2d.$$

Since ω_0 satisfies (1.3), we have that

$$| < k, \omega > | > \frac{1}{2}\gamma |k|^{-\tau}, \tag{3.2}$$

holds for $0 \ne |k| \le K$ and all $\omega \in B$, where K is the minimum integer satisfying $2K^n e^{-K\rho} \le \epsilon$ for a given ρ.

Remark. In this section, for given s, ϵ, we denote by

$$\alpha = \epsilon^{\frac{8d}{16d^2+2d-1}}, \quad s_+ = \alpha s, \epsilon_+ = \epsilon^\kappa, \kappa = 1 + \frac{1}{4d+1}. \tag{3.3}$$

Since a constant factor c independent of iteration step is irrelevant, in the following "$< c$" is abbreviated to be "$< \cdot$".

We assume that H satisfies

$$\frac{1}{s^{2d}} ||P||_{D\times B} = \epsilon, \tag{3.4}$$

and

$$\frac{\epsilon s^{2d}}{s_u^{i^*}} < \cdot\, \alpha^{\frac{1}{4d}}, \text{ for } i = 2, \cdots, i^*. \tag{3.5}$$

Moreover, there are s_1, s_2 with $-s_u \le s_2 < 0 < s_1 \le s_u$, such that

$$-3s^{2d} \le f(u) \le \epsilon s^{2d}, \text{ for } u \in [s_2, s_1], \ -3s^{2d} \le f(s_1), f(s_2) \le -s^{2d}, \tag{3.6}$$

$$a_2 \le 0, \tag{3.7}$$

$$\max_{|u|\le s_u} |N_h(u)| <\cdot s^{2d}\alpha^{-\frac{1}{d}}, \tag{3.8}$$

$$|a_2 s_u| <\cdot s^d, \tag{3.9}$$

where $f(u) = a_2u^2 + N_h(u)$.

It follows from (3.6) that

$$s_u \ge \frac{1}{2}s^d, \tag{3.10}$$

if $s_u \le s \le \frac{1}{4}$.

The purpose of this section is to find a change of variables defined in a smaller domain $D_+ \times B_+(\omega_0)$, such that the transformed Hamiltonian $H_+ = N_+ + P_+$ satisfies all the above iteration assumptions with smaller $s_+, s_{u+}, \alpha_+, c\epsilon_+$(see (3.43), (3.51), (3.37), (3.39), (3.31), (3.40), (3.41) below).

In order to help the reader understand the complicated iteration assumptions, we add the following remarks:

1. Assumption (3.5) is used to prove the convergence of the normal form series.
2. Assumption (3.6) guarantees that the critical points of f in $[-s_u, s_u]$ will not be completely moved out by the perturbation.
3. Assumption (3.7) means that the origin is a saddle-like singular point of the unperturbed system in the uv plane, which might be degenerate: i.e., $a_2 = 0$.
4. Assumption (3.8) is crucial for controlling the size of $\{N_h, F\}$, which is treated as a part of the new perturbation.
5. Assumption (3.9) is used to control the size of F, and thus the size of the generated symplectic transformation.

Truncating Perturbations. Expanding P into a Fourier-Taylor series,

$$P(x, y, u, v) = \sum P_{klpq}e^{i(k,x)}y^l u^p v^q,$$

where the sum is taken over

$$k = (k_1, \cdots, , k_n) \in Z^n, l = (l_1, \cdots, l_n) \in Z_+^n, p, q \in Z_+.$$

Let

$$R = \sum_{|k|\le K, 2|l|+q+2\frac{p}{i^*}\le 2} P_{klpq}e^{i(k,x)}y^l u^p v^q, \tag{3.11}$$

be a truncation of P, where K is the minimum integer satisfying $2K^n e^{-K\rho} \le \epsilon$, $|l| = \sum_{i=1}^n l_i$.

Note that if $2|l| + q + 2\frac{p}{i^*} > 2$, then $2|l| + q + 2\frac{p}{i^*} \ge 2 + \frac{1}{i^*} \ge 2 + \frac{1}{2d}$, since all the qualities are integer.

The following estimates come from the Cauchy estimates.

Lemma 3.1.

$$||R||_{D(r-\rho,\frac{1}{2}s,\frac{1}{2}s_u)\times B} \le c(\rho)||P||_{D\times B} = c(\rho)s^{2d}\epsilon. \tag{3.12}$$

Moreover, in a smaller domain, $D_1 = D(r - \rho, \alpha s, s_{u+})$, we have

$$||P - R||_{D_1} <\cdot (K^n e^{-K\rho} + \alpha^{2d+\frac{d}{2i^*}})||P||_D <\cdot s^{2d}\epsilon^2, \tag{3.13}$$

if $\alpha = \epsilon^{\frac{8d}{16d^2+2d-1}}$, $\frac{s_{u+}}{s_u} < \alpha^{\frac{2d}{i^*}-\frac{1}{4d}}$.

Proof. Equation (3.12) directly follows from the Cauchy inequality. Now we prove (3.13). Note that

$$P - R \;=\; \sum_{|k|\geq K} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q \;+\; \sum_{|k|\leq K, 2|l|+q+2\frac{p}{i^*}>2} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q.$$

Equation (3.13) follows from the following two estimates:

$$\begin{aligned} |\sum_{|k|\geq K} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q|_{D_1} &\leq \sum_{|k|\geq K} |P|\mathrm{e}^{|k|r}\mathrm{e}^{|k|(r-\rho)} \\ &\leq s^{2d}\epsilon \sum_{\kappa\geq K} \kappa^n \mathrm{e}^{-\kappa\rho} \leq \quad s^{2d}\epsilon^2 \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} &\|\sum_{|k|\leq K, 2|l|+q+2\frac{p}{i^*}>2} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q\|_{D_1} \\ \leq\; & \sum_{2<2|l'|+q'+2\frac{p'}{i^*}\leq 3} \|\int \frac{\partial^{|l'|+q'+p'}}{\partial y^{l'}\partial u^{p'}\partial v^{q'}}(\sum_{|k|\leq K, 2|l|+q+2\frac{p}{i^*}>2} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q)\|_{D_1} \\ <\cdot\; & \alpha^{(2|l|+q)d}(\frac{s_{u+}}{s_u})^p | \sum_{|k|\leq K, 2|l|+q+2\frac{p}{i^*}>2} P_{klpq}\mathrm{e}^{\mathrm{i}(k,x)}y^l u^p v^q| \\ <\cdot\; & s^{2d}\epsilon\alpha^{2d+\frac{2d-1}{8d}} <\cdot s^{2d}\epsilon^2, \end{aligned} \tag{3.15}$$

in D_1, where $\int$ stands for $\underbrace{\int_0^y \cdots \int_0^{y_1}}_{l'} \underbrace{\int_0^u \cdots \int_0^u}_{p'} \underbrace{\int_0^v \cdots \int_0^v}_{q'}$. □

Solving the linear equation. Let

$$\begin{aligned} \bar{N} &= N + \int_0^{2\pi} R\,dx \\ &= N + P_{0000} + \sum_{1\leq j\leq i^*} P_{00j0}u^j + \sum_{1\leq j\leq \frac{i^*}{2}} P_{00j1}u^j v + P_{0002}v^2 + \langle P_{0100}, y\rangle. \end{aligned} \tag{3.16}$$

In the following, the sum "$\sum$" is always taken over

$$\{(k,l,p,q)|\;\; 0 \neq |k| \leq K, 2|l| + q + 2\frac{p}{i^*} \leq 2\}$$

if there is no further claim. We denote $\Delta = \sqrt{-1}(k,\omega)$ for simplicity. Let

$$F = \sum f_{klpq}e^{i(k,x)}y^l u^p v^q, \tag{3.17}$$

be the solution of the following partial differential equation:

$$\{N_2, F\} + R = \sum \Delta f_{klpq} + \frac{\partial F}{\partial v}\frac{\partial N_2}{\partial u} - \frac{\partial F}{\partial u}\frac{\partial N_2}{\partial v} + R = \bar{N},$$

where

$$\frac{\partial F}{\partial v}\frac{\partial N_2}{\partial u} - \frac{\partial F}{\partial u}\frac{\partial N_2}{\partial v} = v \;\; 2a_2 u \sum q f_{klpq} e^{i(k,x)} y^l u^p v^{q-1} \\ - \sum p f_{klpq} e^{i(k,x)} y^l u^{p-1} v^q.$$

That is

$$f_{kl00} = -\Delta^{-1} P_{kl00},$$

$$f_{kl20} = (\Delta^3 + 8a_2\Delta)^{-1} det \begin{pmatrix} -P_{kl20} & 2a_2 & 0 \\ -P_{kl11} & \Delta & 4a_2 \\ -P_{kl02} & -1 & \Delta \end{pmatrix},$$

$$f_{kl11} = (\Delta^3 + 8a_2\Delta)^{-1} det \begin{pmatrix} \Delta & -P_{kl20} & 0 \\ -2 & -P_{kl11} & 4a_2 \\ 0 & -P_{kl02} & \Delta \end{pmatrix},$$

$$f_{kl02} = (\Delta^3 + 8a_2\Delta)^{-1} det \begin{pmatrix} \Delta & 2a_2 & -P_{kl20} \\ -2 & \Delta & -P_{kl11} \\ 0 & -1 & -P_{kl02} \end{pmatrix}.$$

For $0 < p \neq 2$,

$$f_{klp0} = -(\Delta^2 + 2pa_2)^{-1}(\Delta P_{klp0} - 2a_2 P_{kl,p-1,1}),$$
$$f_{kl,p-1,1} = -(\Delta^2 + 2pa_2)^{-1}(\Delta P_{kl,p-1,1} + pP_{klp0}).$$

Remark. Since do not know the size of a_2, which comes from the perturbation, we have to keep all the terms with $k = 0$. It makes $\bar{N}$ a bit complicated. Later, we have to employee further transformations to put $\bar{N}$ into the normal form.

We first give estimates for F. Let $D_i = D(r-8\rho+i\rho, \frac{1}{i}s, \frac{1}{i}s_u) \subset D_1 = D, 0 < i \leq 4$, and

$$\Gamma(\rho) = \sup_{t\geq 0} t^{3\tau} e^{-\rho t}.$$

Lemma 3.2.

$$\frac{1}{s^{2d}} ||F||_{D_2\times B} < \cdot \Gamma\epsilon. \tag{3.18}$$

Proof. Rewrite F into the following:

$$F = F_1 + F_2 + F_3 + F_4 + \cdots + F_9,$$

where

$$F_1 = \sum f^1_{klpq} e^{i(k,x)} y^l u^p v^q, \tag{3.19}$$

with

$$f^1_{kl00} = -\Delta^{-1} P_{kl00},$$
$$f^1_{klpq} = -(\Delta^2 + 2pa_2)^{-1} \Delta P_{klpq}, \text{for } \; 0 < p + q \neq 2, \tag{3.20}$$

$$f^1_{kl11} = -(\Delta^2 + 8a_2)^{-1} \Delta P_{kl11},$$
$$f^1_{klpq} = -(\Delta^2 + 8a_2)^{-1} \Delta^{-1}(\Delta^2 - 4a_2) P_{klpq}, (p, q) = (2, 0), (0.2);$$

$$F_2 = -\sum(\Delta^2 + 8a_2)^{-1} 8a_2^2 P_{kl02} e^{i(k,x)} y^l u^2;$$
$$F_3 = \sum(\Delta^2 + 8a_2)^{-1} 2a_2 P_{kl11} e^{i(k,x)} y^l u^2;$$
$$F_4 = -\sum(\Delta^2 + 8a_2)^{-1} 4a_2 P_{kl02} e^{i(k,x)} y^l uv;$$
$$F_5 = -\sum(\Delta^2 + 8a_2)^{-1} 2P_{kl20} e^{i(k,x)} y^l uv;$$
$$F_6 = -\sum(\Delta^2 + 8a_2)^{-1} \Delta^{-1} 2P_{kl20} e^{i(k,x)} y^l v^2;$$
$$F_7 = -\sum(\Delta^2 + 8a_2)^{-1} P_{kl11} e^{i(k,x)} y^l v^2;$$
$$F_8 = \sum(\Delta^2 + 2pa_2)^{-1}(2a_2) P_{kl,p-1,1} e^{i(k,x)} y^l u^p, \ p \neq 2;$$
$$F_9 = -\sum p(\Delta^2 + 2pa_2)^{-1} P_{klp0} e^{i(k,x)} y^l u^{p-1} v, \ p \neq 2.$$

To prove Lemma 3.2, it suffices to prove that

$$\frac{1}{s^{2d}} ||F_i||_{D_2 \times B} < \cdot \Gamma\epsilon, \ i = 1, \cdots, 9.$$

In the following, we only give the estimates for F_1, F_8, F_9, the estimates for the others are similar to one of F_1, F_8, F_9.

By Cauchy estimates, for fixed k,

$$|\sum P_{klpq}(\omega) y^l u^p v^q|_{D \times B} \leq s^{2d} \epsilon e^{-|k|r}.$$

In view of (3.1) - (3.9), we know that $\frac{1}{2}s^d < s_u \leq s, a_2 \leq 0$ and $|a_2|s_u \leq 2s^d$. It follows that

$$\begin{aligned}
|| \ & F_1(x,y,u,v)||_{D_2 \times B} \\
\leq \ & \sum_{k \neq 0} (\sum |f^1_{klpq}(\omega) y^l u^p v^q|_{D_2 \times B}) e^{|k|(r-\rho)} \\
\leq \ & \sum_{k \neq 0} |\Delta^3| e^{-|k|\rho} s^{2d} \epsilon = \Gamma s^{2d} \epsilon,
\end{aligned}$$

$$\begin{aligned}
|| \ & F_8(x,y,u,v)||_{D_2 \times B} \\
= \ & || \sum (\Delta^2 + 2pa_2)^{-1}(2a_2) P_{kl,p-1,1} e^{i(k,x)} y^l u^p ||_{D_2 \times B} \\
= \ & || \frac{d}{dv} \sum (\Delta^2 + 2pa_2)^{-1}(2a_2 u) P_{kl,p-1,1} e^{i(k,x)} y^l u^{p-1} v ||_{D_2 \times B} \\
\leq \ & |2a_2 u| \cdot 2|s^{-d}| \cdot || \sum (\Delta^2 + 2pa_2)^{-1} P_{kl,p-1,1} e^{i(k,x)} y^l u^{p-1} v ||_{D \times B} \\
< \cdot \ & \Gamma s^{2d} \epsilon,
\end{aligned}$$

$$\begin{aligned}
|| \ & F_9(x,y,u,v)||_{D_2 \times B} \\
= \ & || \sum (\Delta^2 + 2pa_2)^{-1} p P_{klp0} e^{i(k,x)} y^l u^{p-1} v ||_{D_2 \times B}
\end{aligned}$$

$$= ||\frac{d}{du}(\sum(\Delta^2 + 2pa_2)^{-1}P_{klp0}e^{i(k,x)}y^l u^p) \cdot v||_{D_2\times B}$$
$$= 2s_u^{-1}s^d||\sum(\Delta^2 + 2pa_2)^{-1}P_{klp0}e^{i(k,x)}y^l u^p)||_{D\times B}$$
$$<\cdot\ s_u^{-1}s^d\Gamma s^{2d}\epsilon <\cdot \Gamma s^{2d}\epsilon. \qquad \square$$

Again by the Cauchy inequality, it follows that

$$\rho||F_x||, s^{2d}||F_y||, s_u||F_u|, s^d||F_v|| <\cdot \Gamma s^{2d}\epsilon, \tag{3.21}$$

uniformly on $D_2 \times B$.

Let

$$||D^i F||_{D\times B} = \max\{\rho^{-i_1}|\frac{\partial^i F}{\partial x^{i_1}\partial y^{i_2}\partial u^{i_3}\partial v^{i_4}}|_{D\times B}, |i_1| + |i_2| + i_3 + i_4 = i\}.$$

Note that F is a polynomial in y of order 1, in v of order 2 and in u of order i^*. From the Cauchy inequality and (3.5), it also follows that

$$||D^i F||_{D_4\times B} <\cdot \Gamma\frac{s^{2d}\epsilon}{s_u^{i^*}} <\cdot \Gamma\alpha^{\frac{1}{4d}}, \tag{3.22}$$

for $i = 2, \cdots, i^*$.

If ϵ is sufficiently small so that $\cdot\Gamma\epsilon < (\frac{1}{2}\alpha)^{4d}$ in (3.21), the symplectic coordinate transformation $\Phi = X_F^t|_{t=1}$ maps $D_{\frac{1}{2}\alpha}$ into D_α. Moreover, Φ transforms H into $\bar{H} = \bar{N} + \{N_h, F\} + \bar{P}$(see (2.3)).

The new normal form. From (3.13), we know that $\bar{P}$ is smaller only when the new definition domain of u is smaller, say $\frac{s_{u+}}{s_u} < \alpha^{\frac{2d}{i^*}-\frac{1}{4d}}$. In what follows we fix this domain and find a symplectic change of variables $\Phi_2 \circ \Phi_1$ which transforms $\bar{N}$ of (3.16) into the normal form up to a smaller term; i.e., $N_+ = \bar{N} \circ \Phi_2 \circ \Phi_1 + O(\epsilon_+)$ satisfying (3.5)–(3.9) with s_+, s_{u+}, ϵ_+, where s_{u+} is determined later.

Firstly, we use

$$\Phi_1 : u = u_1, v = v_1 - (1 + 2P_{0002})^{-1}\sum_{1\le p\le \frac{i^*}{2}} P_{00p1}u_1^j, \tag{3.23}$$

to kill the mixing terms $\sum_{p=1}^{\frac{i^*}{2}} P_{00p1}u^p v$ in $\bar{N}$(see (3.16)). It follows that

$$\bar{N}\circ\Phi_1 = N + P_{0000} + \sum_{i^*\ge p\ge 1} P_{00j0}u_1^p + (1 + P_{0002})^{-1}(\sum_{\frac{i^*}{2}\ge p\ge 1} P_{00p1}u_1^p)^2$$
$$+ P_{0002}v_1^2 + P_{0100}y. \tag{3.24}$$

Denote by

$$f_1(u_1) = f(u_1) + \sum_{i^*\ge p\ge 1} P_{00p0}u_1^p + (1 + P_{0002})^{-1}(\sum_{\frac{i*}{2}\ge p\ge 1} P_{00p1}u_1^p)^2. \tag{3.25}$$

Now we fix the definition domain of the u_1 variable of the next iteration step, which is a bit complicated. Let

$$L = \{u_1 \in [s_2, s_1], f_1(u_1) \ge -2s_+^{2d}\},$$

where s_1, s_2 are defined in (3.6).

Lemma 3.3.

$$|L| \leq C_0 \alpha^{\frac{2d}{i^*}} s_1, \tag{3.26}$$

where $|L|$ denotes the Lebesgue measure of L and $C_0 = 4(2 + 3 + \cdots + 2d + 4^d)$.

Proof. In view of (3.3), (3.6), together with the fact $\epsilon < \frac{1}{2}\alpha^{2d}$ if $\epsilon \leq 2^{-8d-1}$, we have

$$-3s^{2d} - s^{2d}\epsilon \leq f_1(u_1) \leq 2s^{2d}\epsilon, \text{for } s_2 \leq u \leq s_1, \quad -4s^{2d} \leq f_1(s_1), f_1(s_2) \leq -\frac{1}{2}s^{2d}. \tag{3.27}$$

It follows that

$$L = \{u_1 \in [s_2, s_1], |f_1(u_1)| \leq 2s_+^{2d}\},$$

and

$$\max_{s_2 \leq u_1 \leq s_1} |f(u_1)| \geq \frac{1}{2}s^{2d}.$$

Now Lemma 5.2 in the Appendix leads to the estimate in Lemma 3.3. □

Note that $f_1(0) = 0 \geq -2s_+^{2d}$ implies $0 \in L$. Denote by $\overline{L} = [\overline{s}_2, \overline{s}_1]$ the connected component of L which contains zero. Since $f_1(0) = 0 > -2s_+^{2d} = f_1(\overline{s}_1) = f_1(\overline{s}_2)$, f_1 will reach its maximum at an interior point of $\overline{L}$, say u_m. Let $\tilde{s}_1 = \overline{s}_1 - u_m$, $\tilde{s}_2 = \overline{s}_2 - u_m$. We have

$$f_1(\tilde{s}_1 + u_m) = f_1(\tilde{s}_2 + u_m) = -2s_+^{2d}. \tag{3.28}$$

Without loss of generality, we assume that $|\tilde{s}_2| \leq \tilde{s}_1$.

Corollary 1.

$$\tilde{s}_1 \leq 4C_0 \alpha^{\frac{2d}{i^*}} s_1. \tag{3.29}$$

Now we move the origin to the maximum by the following linear symplectic change of variables:

$$\Phi_2 : v_1 = (1 + P_{0002})^{-\frac{1}{2}} v_+, u_1 = (1 + P_{0002})^{\frac{1}{2}} u_+ + u_m. \tag{3.30}$$

Let $\tilde{N} = \overline{N} \circ \Phi_1 \circ \Phi_2$. It is easy to see that

$$\tilde{N} =< \omega_+, y > + \frac{1}{2}v_+^2 + \tilde{f}(u_+) + f_1(u_m),$$

with

$$\tilde{f}(u_+) = \sum_{i=2}^{i^*} a_{+i} u_+^i, \ \ a_{2+} \leq 0. \tag{3.31}$$

Remark. The above symplectic change of variables depends only continuously on the parameter ω. As a consequence, the resulting Hamiltonian depends only continuously on ω.

The following estimate will be used to prove the convergence of the coefficients of f and thus the convergence of normal form series.

Lemma 3.4.

$$|a_{i+} - a_i| < \cdot \alpha^{\frac{1}{4d}}, \ i = 1, \cdots, i^*. \tag{3.32}$$

Proof. Equation (3.32) can be directly verified from (3.5), (3.25), (3.26), (3.30) by the formula

$$a_{i+} = \frac{1}{i!}\frac{d^i \tilde{f}}{du_+^i}|_{u_+=0} = \frac{1}{i!}\frac{d^i f_1}{du^i}|_{u=u_m}(1 + P_{0002})^{\frac{i}{2}}. \qquad \square$$

Now $\tilde{N}$ is already in normal form. But if s_{1+} is too small, we can not get the result of Lemma 3.4 in the next KAM step which is crucial for proving the convergence of the normal form series. Since we don't know the exact size of s_{1+}, we have to discard some higher order terms in $\tilde{f}$ according to s_{1+}. For the same reason we also need to shift s_{1+} up a little bit when necessary. Certainly, this modification will make the new perturbation a bit larger, but it does not influence the convergence of the iteration.

Let

$$[s_{2+}, s_{1+}] = [(1 + P_{0002})^{-\frac{1}{2}}\tilde{s}_2, (1 + P_{0002})^{-\frac{1}{2}}\tilde{s}_1]. \tag{3.33}$$

More precisely, let i_+^* be the integer such that

$$s_{1+} \in \left((s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}})^{\frac{1}{i_+^*}}, (s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}})^{\frac{1}{i_+^*+1}}\right], \tag{3.34}$$

if $s_{1+}^{i^*} \leq s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}}$, otherwise let $i_+^* = i^*$.

In view of (3.28) and (3.30), it follows that that $\tilde{f}(s_{1+}) = \tilde{f}(s_{2+}) = f_1(\bar{s}_1) = f_1(\bar{s}_2) = 2s_+^{2d}$. Together with (3.32), it follows that $\bar{s}_1, |\bar{s}_2| \geq s_+^d$. Note that (3.34) implies

$$\begin{aligned} s_{1+} &> \frac{3}{4}\tilde{s}_1 \geq \frac{3}{4}(\bar{s}_1 - u_m) \geq \frac{3}{4}[(\bar{s}_1 - u_m) - (\bar{s}_2 - u_m)] \\ &> \frac{3}{4}(\bar{s}_1 - \bar{s}_2) \geq \frac{3}{4}s_+^d > (s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}})^{\frac{1}{2}}. \end{aligned} \tag{3.35}$$

It follows that $i_+^* \geq 2$.

Define

$$s_{u+} = \begin{cases} 10s_{1+}\alpha_+^{-\frac{1}{2di_+^*}}, & \text{if } s_{1+}^{i_+^*} \in (s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}}, s_+^{2d}\epsilon_+\alpha_+^{-\frac{1}{4d}}), \\ 10s_{1+}, & \text{otherwise,} \end{cases} ; \tag{3.36}$$

here we assume that $|s_{2+}| \leq s_{1+}$, otherwise we replace s_{1+} in (3.36) by $|s_{2+}|$.

In view of (3.34), (3.36), we have

$$\frac{s_+^{2d}\epsilon_+}{s_{u+}^{i_+^*}} \leq \alpha_+^{\frac{1}{4d}}. \tag{3.37}$$

By definition (3.36),

$$s_{u+}^{i_+^*+1} \leq \max\{(10^{i_+^*+1}(s_+^{2d}\epsilon_+\alpha_+^{-\frac{3}{4d}})^{\frac{i_+^*+1}{i_+^*}}, s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}}\} < \cdot\, s_+^{2d}\epsilon_+\alpha_+^{\frac{1}{4d}}.$$

It follows that

$$\begin{aligned} \max_{|u_+|\leq s_{u+}} |\sum_{i=i_+^*+1}^{i^*} a_{+i}u_+^i| &\leq \max_{|u_+|\leq s_u} \sum_{i=i_+^*+1}^{i^*} |a_{+i}u_+^i| \\ &\leq 2|\sum_{i=i_+^*+1}^{i^*} s_{u+}^i| \leq 8s_{u+}^{i_+^*+1} < \cdot\, s_+^{2d}\epsilon_+. \end{aligned} \tag{3.38}$$

Thus we can put the higher order terms $\sum_{i=i_+^*+1}^{i^*} a_{+i}u_+^i$ in $\tilde{N}$ into the new perturbation and keep only the terms which are bigger than ϵ_+.

More precisely, we set

$$f_+(u_+) = \sum_{i=2}^{i_+^*} a_{i+}u_+^i,$$

for the next iteration step.

As a matter of fact, we mention that

$$s_{u+} < \cdot\, \alpha^{\frac{2d}{i^*} - \frac{1}{4d}} s_u,$$

by (3.33) and (3.36), which will lead to the estimate of (3.13) in D_1.

Now we prove

Lemma 3.5.

$$-3s_+^{2d} \leq f_+(u_+) \leq s_+^{2d}\epsilon_+,\ \textit{for } u \in [s_{2+}, s_{1+}], \textit{and } -3s_+^{2d} \leq f_+(s_{1+}), f(s_{2+}) \leq -s_+^{2d}, \tag{3.39}$$

Proof. In fact, since $-2s_+^{2d} \leq f_1 \leq s^{2d}\epsilon$ in $[\bar{s}_2, \bar{s}_1]$, by the definition of $\bar{L}$, we have $-\frac{5}{2}s_+^{2d} \leq -2s_+^{2d} - f_1(u_m) \leq \tilde{f} \leq 0$ in $[\tilde{s}_1, \tilde{s}_2]$. Combining with (3.33) and (3.38), it follows that $-3s_+^{2d} \leq f_+ \leq s_+^{2d}\epsilon_+$ in $[s_{1+}, s_{2+}]$. Another half of (3.39) is proved similarly. □

Lemma 3.6.

$$\max_{|u_+|\leq s_{u^+}} |N_{h+}| = \max_{|u_+|\leq s_{u^+}} |\sum_{i=3}^{i_+^*} a_{+i}u_+^i| \leq c_0 s_+^{2d}\alpha_+^{-\frac{1}{2d}}, \tag{3.40}$$

where $c_0 = 10C^{2d}$, the constant C is defined in Lemma 5.3.

Proof. In view of (3.39), Lemma 5.3 and the definition of s_{u+}, we have

$$\begin{aligned}\max_{|u_+|\leq s_{u^+}} |N_{h+}| &\leq \sum_{i=2}^{i_+^*} |a_{+i}s_{u+}^i| \leq \sum_{i=2}^{i_+^*} |a_{+i}| 10^i s_{1+}^i \alpha_+^{-\frac{i}{2di_+^*}} \\ &\leq 10^{i_+^*}\alpha_+^{-\frac{1}{2d}} \sum_{i=2}^{i_+^*} |a_{+i}| s_{1+}^i \leq 10^{i_+^*} C\alpha_+^{-\frac{1}{2d}} \max_{0\leq u\leq s_{1+}} |\sum_{i=2}^{i_+^*} a_{+i}u^i| \\ &\leq c_0\alpha_+^{-\frac{1}{2d}} s_+^{2d}. \qquad \square\end{aligned}$$

Lemma 3.7.

$$|a_{2+}s_{u+}| < \cdot\, s_+^d. \tag{3.41}$$

Proof. If $i_+^* \geq 3$, then $s_{u+} \geq s_+^{\frac{2d+1}{3}}$ by (3.34), (3.36). Combining with (3.39), it follows that

$$|a_{2+}s_{u+}^2| \leq |a_{2+}s_{1+}^2|\alpha_+^{-\frac{1}{d}} \leq \max_{|u|\leq s_{1+}} \sum_{i=2}^{i_+^*} |a_{i+}u^i|\alpha_+^{-\frac{1}{d}}$$
$$< \cdot \max_{|u|\leq s_{1+}} |\sum_{i=2}^{i_+^*} a_{i+}u^i|\alpha_+^{-\frac{1}{d}} < \cdot s_+^{2d}\alpha_+^{-\frac{1}{d}}. \tag{3.42}$$

Thus

$$|a_{2+}s_{u+}| < \cdot s_+^{2d}\alpha_+^{-\frac{1}{d}}s_u^{-1} < \cdot s_+^{\frac{4d-1}{3}}\alpha_+^{-\frac{1}{d}} \leq s_+^d.$$

In case that $i_+^* = 2$, we have $s_{u+} = 10s_{1+}$ in (3.36) since $s_{1+} \geq s_+^d$. It follows that

$$|a_{2+}s_{u+}| \leq 10|a_{2+}|s_{1+}^2 s_{1+}^{-1} \leq 10s_+^{-d}|a_{2+}s_{1+}^2| \leq 10Cs_+^{-d}\max_{|u|\leq s_{1+}}|f_+(u)| < \cdot s_+^d. \qquad \square$$

So far, we have found a symplectic change of variables such that the normal form part N_+ of the transformed Hamiltonian $H_+ = N_+ + P_+$ satisfies all the iteration assumptions with $s_+, r_+ = r - \rho, \epsilon_+$ uniformly for $\omega \in B$.

Before proving the smallness of P_+, we first note that after a cycle of iteration the frequencies of the normal form is shifted to $\omega_+ = \phi(\omega) = \omega + P_{0100}(\omega)$. Since $K^n e^{-K\rho} \leq \epsilon, |P_{0100}| \leq \epsilon$, the range of B under ϕ still contains the ball $B(\omega_0, \frac{1}{2}\gamma K_+^{-\tau-1})$ in ω_+ space.

Treating ω_+ as new parameters, we have

$$| < k, \omega_+ > | > \frac{1}{2}\gamma K_+^{-\tau} \tag{3.43}$$

for $\omega_+ \in B(\omega_0, \frac{1}{2}\gamma K_+^{-\tau-1})$ and $0 \neq k \leq K_+$, where K is the minimum integer satisfying $2K_+^n e^{-K_+\rho} \leq \epsilon_+$, if ϵ is sufficiently small.

Now we arrive at a new normal form

$$N_+ = e_+ + (\omega_+, y_+) + \frac{1}{2}v_+^2 + f_+(u_+),$$

with $e_+ = e + P_{0000} + f_1(u_m)$, satisfying all the inductive assumptions (3.1) - (3.9) with s_+, s_{u+}, ϵ_+.

Estimates for the new perturbation. To finish one cycle of iteration, the only thing which remains is to estimate the new error term.

Firstly, we give some estimates for X_F^t, Φ_1, Φ_2. The following (3.44), which will be used to prove our coordinate transformations is well defined. Equation (3.45) is useful for proving the convergence of the composition of coordinate transformations.

Let $D_{\frac{i}{8}\alpha} = D(r - 9\rho + i\rho, \frac{i}{8}s_+, \frac{i}{8}s_{u+})$. Especially, we denote $D_{\frac{1}{8}\alpha}$ by D_+.

Lemma 3.8.

$$\Phi_2 : D_+ \to D_{\frac{1}{4}\alpha},$$
$$\Phi_1 : D_{\frac{1}{4}\alpha} \to D_{\frac{1}{2}\alpha},$$
$$X_F^t : D_{\frac{1}{2}\alpha} \to D_\alpha, \quad 0 \leq t \leq 1, \tag{3.44}$$

if s, ϵ is sufficiently small. Moreover,

$$||D\Psi - Id||_{D_+} < E, ||D^2\Psi||_{D_+} < c, \tag{3.45}$$

where $\Psi = \Phi \circ \Phi_1 \circ \Phi_2$.

Proof. Note that Φ_2 is a linear transformation defined by (3.30). It follows that, in D_+,

$$\begin{aligned}|v_1 - v_+| &= |(1+P_{0002})^{-\frac{1}{2}}v - v| = |\frac{P_{0002}}{(1+P_{0002})^{\frac{1}{2}}}v| \le \frac{1}{4}Es_+^d \le \frac{1}{8}s_+^d,\\ |u - u_+| &= |(1+P_{0002})^{\frac{1}{2}}u_+ + u_m - u_+| = |u_m + \frac{P_{0002}}{1+(1+P_{0002})^{\frac{1}{2}}}u_+|\\ &\le \frac{1}{10}s_{u^+} + \frac{1}{4}\epsilon s_{u^+} \le \frac{1}{8}s_{u^+},\end{aligned}$$

which implies $\Phi_2 : D_+ \to D_{\frac{1}{4}\alpha}$. Moreover,

$$||D^i\Phi_2||_{D_+} < 2,\ i \ge 2. \tag{3.46}$$

Since $|\sum_{1\le p\le \frac{i^*}{2}} P_{00p1}u^p| \le \frac{\epsilon}{s^d} = s^d\epsilon \le \frac{1}{4}s_+^d$ if $\epsilon \le \frac{1}{16}$, It follows from (3.23), $\Phi_1 : D_{\frac{1}{4}\alpha} \to D_{\frac{1}{2}\alpha}$. Moreover, by (3.37),

$$||D^i\Phi_1 - Id||_{D_{\frac{1}{4}\alpha}} \le |\frac{d^i}{du^i}\sum_{1\le p\le \frac{i^*}{2}} P_{00p1}u^p|_{D_{\frac{1}{4}\alpha}} \le C\frac{s^d\epsilon}{s_u^{\frac{i^*}{2}}} \le C\epsilon^{\frac{1}{3}},$$

for $i \ge 2$. Thus

$$||D^i\Phi_1|| < 2,\ i \ge 2. \tag{3.47}$$

To get the estimates for X_F^t, we start from the integral equation,

$$X_F^t = id + \int_0^t J(\nabla F)\circ X_F^s\, ds.$$

$X_F^t : D_{\frac{1}{2}\alpha} \to D_\alpha,\quad 0 \le t \le 1$, follows directly from (3.21). Since

$$DX_F^t = Id + \int_0^t JD^2F\circ X_F^s \cdot DX_F^s\, ds.$$

It follows that

$$||DX_F^t - Id|| \le 2\max|D^2F| \le 2\epsilon. \tag{3.48}$$

The estimates of higher order derivatives $D^iX_F^t (i \ge 2)$ follows from (3.22).

Equation (3.45) can be verified, in view of (3.46), (3.47), (3.48), from the following formula:

$$D\Psi = (D\Phi)\circ\Phi_1\circ\Phi_2\cdot(D\Phi_1)\circ\Phi_2\cdot D\Phi_2,$$

$$\begin{aligned}D^2\Psi &= (D^2\Phi)\circ\Phi_1\circ\Phi_2\cdot((D\Phi_1)\circ\Phi_2\cdot D\Phi_2)^2\\ &+ (D\Phi)\circ\Phi_1\circ\Phi_2\cdot(D^2\Phi_1)\circ\Phi_2\cdot(D\Phi_2)^2\\ &+ (D\Phi)\circ\Phi_1\circ\Phi_2\cdot(D\Phi_1)\circ\Phi_2\cdot(D\Phi_2)^2. \qquad \square\end{aligned}$$

By the definitions of $\Phi, \Phi_1, \Phi_2, \Psi$ and Lemma 3.8, we know that

$$H \circ \Psi = H \circ \Phi \circ \Phi_1 \circ \Phi_2 = N_+ + P_+$$
$$= N_+ + (\bar{P} + \{F, N_h\}) \circ \Phi_1 \circ \Phi_2 + \sum_{i^{*+}+1}^{i^*} a_{i+} u^i,$$

is well defined in $D_+ \times B_+$. Moreover, we have the following estimates:

$$\begin{aligned}
|| \quad P_+||_{D_+ \times B_+} &= ||(\bar{P} + \{F, N_h\}) \circ \Phi_1 \circ \Phi_2 + \sum_{i^{*+}}^{i^*} a_{i+} u^i ||_{D_{\frac{1}{8}\alpha} \times B_+} \\
= \quad & ||(\int_0^1 \{R_t, F\} \circ X_F^t dt + \{R, F\} + (P - R) \circ \Phi - \frac{\partial F}{\partial v} \frac{\partial N_h}{\partial u}) \circ \Phi_1 \circ \Phi_2 \\
& + \sum_{i^{*+}}^{i^*} a_{i+} u_+^i ||_{D_{\frac{1}{8}\alpha} \times B_+} \\
< \cdot \quad & ||(\int_0^1 \{R_t, F\} \circ X_F^t dt + \{R, F\} + (P - R) \circ \Phi)||_{D_{\frac{1}{2}\alpha} \times B} \\
& + || \frac{\partial F}{\partial v} \frac{\partial N_h}{\partial u} ||_{D_{\frac{1}{4}\alpha} \times B} + s_+^{2d} \epsilon_+ \\
< \cdot \quad & ||\{R_t, F\}||_{D_\alpha \times B} + ||\{R, F\}||_{D_\alpha \times B} + ||(P - R)||_{D_\alpha \times B} \\
& + || \frac{\partial F}{\partial v} \frac{\partial N_h}{\partial u} ||_{D_{\frac{1}{2}\alpha} \times B} + s_+^{2d} \epsilon_+ \\
< \cdot \quad & \rho^{-1} (\Gamma \epsilon^2 s^{2d} + s^{2d} \epsilon^2 + || \frac{\partial F}{\partial v} \frac{\partial N_h}{\partial u} ||_{D_{\frac{1}{2}\alpha} \times B} + s_+^{2d} \epsilon_+) \\
< \cdot \quad & \rho^{-1} (\Gamma \epsilon^2 s^{2d} + s^{2d} \epsilon^2 + \Gamma \epsilon s^d || \frac{\partial N_h}{\partial u} ||_{D_{\frac{1}{2}\alpha} \times B} + s_+^{2d} \epsilon_+),
\end{aligned} \tag{3.49}$$

where $R_t = (1 - t)\{N + R, F\}$. Since $s_u \geq s^{\frac{2d+1}{3}}$ (otherwise $N_h = 0$), we have

$$\begin{aligned}
|| \quad & \frac{\partial N_h}{\partial u} ||_{D_{\frac{1}{2}\alpha} \times B} \leq 2 s_u^{-1} \max_{s \in [-s_{u+}, s_{u+}]} ||N_h(u)|| \\
< \cdot \quad & s^{2d} \alpha^{-\frac{1}{d}} s_u^{-1} < \cdot s^{2d} \alpha^{-\frac{1}{d}} s^{-\frac{2d+1}{3}} \\
< \cdot \quad & s^d s^{\frac{d-1}{3}} \alpha^{-\frac{1}{d}} \leq s^d \epsilon,
\end{aligned} \tag{3.50}$$

if s is sufficiently small.

Combining with (3.49) and (3.50), we have

$$||P_+||_{D_+ \times B_+} < \cdot \rho^{-1} \Gamma s^{2d} \epsilon^2 + s_+^{2d} \epsilon_+ < \cdot \Gamma s_+^{2d} \epsilon_+. \tag{3.51}$$

It means there is a constant c depending on d, n, m, γ, δ, but not on the iteration steps such that

$$\frac{1}{s_+^{2d}} ||P_+||_{D_{\frac{1}{8}\alpha} \times B_+} \leq c \rho^{-1} \Gamma \epsilon_+.$$

One circle of the KAM step is finished.

Remark. Equations (3.43), (3.51), (3.37), (3.39), (3.31),(3.40),(3.41) play the same roles as (3.2), (3.4), (3.5), (3.6),(3.7),(3.8), (3.9) in the next KAM step.

4. Iteration and Convergence

For any given r, let $\rho_i = \frac{1}{2^{i+4}}r$,

$$\Psi(r) = \prod_{i=1}^{\infty} \Gamma(\rho_i)^{\frac{1}{\kappa^{i+1}}},$$

where $\kappa = 1 + \frac{1}{4d+1}$, is a well defined finite function of r (see [17, 21]). For any given s_0, ϵ_0, r_0, we define some sequences inductively in the following:

$$\begin{aligned}
r_\nu &= r_{\nu-1} - 8\rho_{\nu-1} = r_0 - 8\sum_{i=1}^{\nu} \rho_i,\\
\epsilon_\nu &= c\rho_{\nu-1}^{-1}\Gamma(\rho_{\nu-1})\epsilon_{\nu-1}^{\kappa},\\
\alpha_\nu &= \epsilon_\nu^{\frac{8d}{16d^2+2d-1}},\\
s_\nu &= \frac{1}{8}\alpha_{\nu-1}s_{\nu-1} = 8^{-\nu}(\prod_{i=0}^{\nu-1}\epsilon_i)s_0,\\
K_\nu &= \min\{K : 2K^n e^{-K\rho_\nu} \le \epsilon_\nu\},\\
B_\nu &= B(\omega_0, \frac{1}{2}\gamma K_\nu^{-\tau-1}),\\
D_\nu &= D(r_\nu, s_\nu, s_{u\nu}),
\end{aligned}$$

where $s_\nu^d < s_{u\nu} \le s_\nu$ depends on each KAM step and can not be defined uniformly. However, because all the estimates are made in s_ν, the convergence can be proved even if $s_{u\nu}$ can not be defined explicitly,

Let $H_\nu = N_\nu + P_\nu(\omega_\nu, x_\nu, y_\nu, u_\nu, v_\nu)$. Summarizing conclusions (3.43), (3.37), (3.39), (3.31),(3.40),(3.41), (3.51), we have the following iteration lemma:

Lemma 4.1. *If s_0, ϵ_0 are sufficiently small, then the following holds for $\nu \ge 0$.*

Let $H_\nu = N_{2\nu} + N_{h\nu} + P_\nu$ satisfies (3.1)- (3.9) in $D_\nu \times B_\nu$ with $\epsilon = \epsilon_\nu, s = s_\nu, s_u = s_{u\nu}$. Then there is a frequency map $\phi_{\nu+1} : B_{\nu+1} \to B_\nu$ and a symplectic change of variables

$$\Psi_\nu : D_{\nu+1} \to D_\nu, \tag{4.1}$$

depending on $\omega \in \phi_{\nu+1}(B_{\nu+1})$, such that $H_{\nu+1} = H_\nu \circ \Psi_\nu$, defined on $D_{\nu+1} \times B_{\nu+1}$, has the form

$$H_{\nu+1} = N_{2,\nu+1} + N_{h,\nu+1} + P_{\nu+1}, \tag{4.2}$$

with

$$N_{2,\nu+1} = e_{\nu+1} + (\omega_{\nu+1}, y_{\nu+1}) + \frac{1}{2}v_{\nu+1}^2 + a_{2,\nu+1}u_{\nu+1}^2,\ N_{h,\nu+1}(u_{\nu+1}) = \sum_{i=3}^{i_{\nu+1}^*} a_{i,\nu+1}u_{\nu+1}^i. \tag{4.3}$$

Moreover, $H_{\nu+1}$ satisfies the estimates

$$\frac{1}{s_{\nu+1}^{2d}}||P_{\nu+1}||_{D_{\nu+1}} \le \epsilon_{\nu+1}, \tag{4.4}$$

$$|a_{i,\nu+1} - a_{i,\nu}| < c\alpha_\nu^{\frac{1}{4d}}, c\epsilon_+^{\frac{1}{d}} \text{ for } i = 2, \cdots, i_{\nu+1}^*. \tag{4.5}$$

Moreover, there are $s_{1,\nu+1}, s_{2,\nu+1}$ *with* $-s_{u,\nu+1} \le s_{2,\nu+1} \le 0 \le s_{1,\nu+1} \le s_{u,\nu+1}$, *such that*

$$-3s_{\nu+1}^{2d} \le f_{\nu+1}(u_{\nu+1}) \le s_{\nu+1}^{2d}\epsilon_{\nu+1}, \text{ for } u \in [s_{2,\nu+1}, s_{1,\nu+1}],$$
$$-3s_{\nu+1}^{2d} \le f_{\nu+1}(s_{1,\nu+1}), f_{\nu+1}(s_{2,\nu+1}) \le -s_{\nu+1}^{2d}, \tag{4.6}$$

$$a_{2,\nu+1} \le 0, \tag{4.7}$$

$$\max_{u \le s_{u,\nu+1}} |N_h(u_{\nu+1})| \le c_0 s_{\nu+1}^{2d} \alpha_{\nu+1}^{-\frac{1}{d}}, \tag{4.8}$$

$$|a_{2,\nu+1}u_{\nu+1}| \le c_0 s_\nu^d, \tag{4.9}$$

where $f_{\nu+1}(u_{\nu+1}) = \sum_{i=2}^{i^*_{\nu+1}} a_{i,\nu+1}u_{\nu+1}^i$.

Now we are in the position to prove the main theorems. We only give the proof of Theorem 1.1. Theorem 1.2 is an immediate consequence of Theorem 1.1.

Proof of Theorem 1.1. Since the assumptions of Theorem 1.1 are satisfied, the iteration lemma applies for $\nu = 0$ if we set

$$s_0 = s, s_{u0} = s, N_0 = N, P_0 = P, B_0 = B,$$

and ϵ_0 is sufficiently small.

Inductively, we obtain the following sequences:

$$\Psi^\nu = \Psi_0 \circ \cdots \circ \Psi_\nu : D_\nu \times B_\nu \to D_0 \times B_0, \nu \ge 1,$$
$$\phi^\nu = \phi_0^{-1} \circ \cdots \circ \phi_\nu^{-1} : B_\nu \to B_0, \nu \ge 0,$$
$$H \circ \Psi^\nu = H_\nu = N_{2\nu} + N_{h\nu} + P_\nu \quad \text{on} \quad D_\nu \times \mathcal{B}_\nu,$$

where $\mathcal{B}_\nu = \phi^\nu B_\nu$ are nested domains.

By same argument as in [17], in view of Lemma 3.8, $\Psi^\nu, D\Psi^\nu$ converges uniformly on $D_\infty \times \mathcal{B}_\infty = D(r - 8\rho, 0, 0) \times \cap_{\nu=0}^\infty \mathcal{B}_\nu$.

In view of (4.5), $N_{2\nu} + N_{h\nu}$ converges to, say $N_{2\infty} + N_{h\infty}$, with

$$N_{2\infty} + N_{h\infty} = e_\infty + (\omega_0, y) + \frac{1}{2}v^2 + \sum_{i=2}^{i_\infty} a_{i,\infty}u^i,$$

where i_∞ varies from 2 to $2d$ depending on the perturbation.

Since

$$\epsilon_\nu = c\Gamma(\rho_{\nu-1})\epsilon_{\nu-1}^\kappa = c^\nu \prod_{i=1}^{\nu-1} \Gamma(\rho_i)^{\nu-i}\epsilon_0^{\kappa^\nu} = (c^{\frac{\nu}{\kappa^\nu}} \prod_{i=1}^{\nu-1} \Gamma(\rho_i)^{\kappa^{-i-1}}\epsilon_0)^{\kappa^\nu} \le (C\Psi(r)\epsilon_0)^{\kappa^\nu},$$

where $C = \sup_\nu c^{\frac{\nu}{\kappa^\nu}}$. It follows that $\epsilon_\nu \to 0$ if ϵ_0 is sufficiently small.

The convergence of $\Psi^\nu, D\Psi^\nu, X_{H_\nu}$ implies that we can take the limit for

$$X_{H\circ\Psi^\nu} = (D\Psi^\nu)^* X_H \circ \Psi^\nu = X_{N_{2\nu}+N_{h\nu}} + X_{P_\nu}, \tag{4.10}$$

and arrive at $(D\Psi^\infty)^* X_{H\circ\Psi^\infty} = X_{N_{2\infty}+N_{h\infty}}$ on $D_\infty = D(r - 8\rho, 0, 0)$ uniformly for $\omega^* \in \mathcal{B}_\infty = \cap_{\nu=0}^\infty \mathcal{B}_\nu$, where

$$\Psi_\infty : T^n \to R^n \times T^n,$$

depending on $\omega \in \mathcal{B}_\infty$ and X_{N_∞} is an integrable system on $T^n \times \{0\}$ carrying the rotation flow of frequencies ω_0.

It follows from (4.10), for any $\omega^* \in \mathcal{B}_\infty$,

$$(D\Psi^\infty)^* X_{H_{\omega^*}} \circ \Psi^\infty(\{\omega_0\} \times T^n)) = X_{N_\infty}(\{\omega_0\} \times T^n)),$$

or equivalently,

$$\phi^t_{H_{\omega^*}}(\Psi^\infty(\{\omega_0\} \times T^n)) = \Psi^\infty(\{\omega_0\} \times T^n),$$

where $\phi^t_{H_{\omega^*}}$ is the flow of $X_{H_{\omega^*}}$.

That means $\Psi^\infty(\{\omega_0\} \times T^n)$ is an embedding invariant torus of the original perturbed Hamiltonian system at $\omega^* \in \mathcal{B}_\infty$.

5. Appendix

The following Lemma 5.1 has been proved in [25]. For the sake of completeness, we repeat the proof here.

Lemma 5.1. *Suppose that $g(u)$ is a m*th *differentiable function on the closure $\overline{I}$ of I, where $I \subset R^1$ is an interval. Let $I_h = \{u| \ \ |g(u)| < h, \ \}, \quad h > 0$. If for some constant $d > 0$, $|g^{(m)}(u)| \geq d$ for $\forall u \in I$, then $|I_h| \leq ch^{\frac{1}{m}}$, where $|I_h|$ denotes the Lebesgue measure of I_h and $c = 2(2 + 3 + \cdots + m + d^{-1})$.*

Proof. Let $I_h^{m-1} = \{u| \ \ |g^{(m-1)}(u)| < h, \ \}$. Since for $\forall u \in I$,

$$|(g^{(m-1)}(u))'| = |g^{(m)}| \geq d > 0,$$

so I_h^{m-1} has at most one connected component and it follows that $|I_h^{m-1}| \leq \frac{2h}{d}$.

Let $I_h^{m-2} = \{u| \ |g^{(m-2)}(u)| < h^2\}$. From the above, $I - I_h^{m-1} = \{u| \ |g^{(m-1)}(u)| \geq h, u \in I\}$ has at most two connected components. Denote these components by $I_{(1)}^{m-1}$ and $I_{(2)}^{m-1}$. Thus,

$$|(g^{(m-2)}(u))'| = |g^{(m-1)}(u)| \geq h, \ \ u \in I_{(1)}^{m-1} \cup I_{(2)}^{m-1}.$$

In the same way as the above, since $I_h^{m-2} \cap I_{(1)}^{m-1}$ and $I_h^{m-2} \cap I_{(2)}^{m-1}$ have at most one connected component in $I_{(1)}^{m-1}$ and $I_{(2)}^{m-1}$ respectively, we have

$$|I_h^{m-2} \cap I_{(1)}^{m-1}| \leq 2h, \ \ |I_h^{m-2} \cap I_{(2)}^{m-1}| \leq 2h.$$

Thus,

$$\begin{aligned} |I_h^{m-2}| &\leq |I_h^{m-2} \cap (I - I_h^{m-1})| + |I_h^{m-2} \cap I_h^{m-1}| \\ &\leq |I_h^{m-2} \cap I_{(1)}^{m-1}| + |I_h^{m-2} \cap I_{(2)}^{m-1}| + |I_h^{m-2} \cap I_h^{m-1}| \\ &\leq 4h + 2d^{-1}h = 2(2 + d^{-1})h. \end{aligned}$$

Let $I_h^1 = \{u| \ \ |g'(u)| < h^{m-1}, \ \}$. In the same way as the above , it follows that

$$|I_h^1| \leq 2(2 + 3 + \cdots + m - 1 + d^{-1})h,$$

and $I - I_h^1$ has at most m connected components. Denote these components by $I_{(1)}^1, I_{(2)}^1, \cdots I_{(m)}^1$ and let $I_h^0 = \{u| \ \ |g(u)| < h^m\}$. Then $|I_h^0 \cap I_{(1)}^1| \le 2h, \cdots |I_h^0 \cap I_{(m)}^1| \le 2h$. So

$$\begin{aligned} |I_h^0| &\le |I_h^0 \cap (I - I_h^1)| + |I_h^0 \cap I_h^1| \\ &\le [2m + 2(2 + 3 + \cdots + m - 1 + d^{-1})]h \\ &\le 2(2 + 3 + \cdots + m + d^{-1})h \le ch. \end{aligned}$$

Noticing that $I_h = I^0_{h^{\frac{1}{m}}}$, we have that $|I_h| \le ch^{\frac{1}{m}}$. □

Lemma 5.2. *Suppose that* $f_1(u) = \sum_{i=2}^m a_i u^i$ *is a polynomial of order* m*, which satisfies*

$$\max_{u \in [s_2, s_1]} |f_1(u)| \ge \frac{1}{2} s^{2d}.$$

Let

$$L = \{u \in [s_2, s_1], |f(u)| \le 2s_+^{2d} = 2(\alpha s)^{2d}\}.$$

Then

$$|L| \le C(\frac{s_{1+}}{s})^{\frac{2d}{m}} = C_0 s_1 \alpha^{\frac{2d}{m}},$$

where $|L|$ *denotes the Lebesgue measure of* L, $C_0 \le 4(2 + 3 + \cdots + m + 2^m)$.

Proof. Without loss of generality, we assume $|s_2| \le s_1$. Consider an auxiliary polynomial

$$P(U) = \sum_{i=2}^m \frac{a_i s_1^i}{s^{2d}} U^i$$

with $U = \frac{u}{s_1}$. Then $P(U)$ is well defined in $[\frac{s_2}{s_1}, 1] \subset [-1, 1]$ with $\max_{U \in [\frac{s_2}{s_1}, 1]} |P(U)| \ge \frac{1}{2}$. It follows that there must be a $m_0 \le m$ such that the coefficients of $P(U)$,

$$a_i s_1^{m_0} s^{-2d} \ge (m!)^{-1} 2^{-m_0}. \tag{5.1}$$

Let m_0 be the largest number so that (5.1) is satisfied. Then

$$\begin{aligned} |\frac{d^{m_0}}{dU} P(U)| &\ge |(m_0)! a_{m_0} s_1^{m_0} s^{-2d}| - \frac{1}{(m_0)!} \sum_{i=m_0+1}^m |i! a_i s_1^{m_0} s^{-2d} U^i| \\ &\ge 2^{-m_0} - \frac{1}{m_0!} \sum_{i=m_0+1}^m 2^{-i} \ge 2^{-m_0-1}, \end{aligned}$$

for $U \in [\frac{s_2}{s_1}, 1]$.

Since

$$L = \{s_1 U. |U \in [\frac{s_2}{s_1}, 1], P(U) \ge \frac{s_+^{2d}}{s^{2d}}\},$$

we have, by Lemma 5.1,

$$|L| = s_1 |\{U | U \in [\frac{s_2}{s_1}, 1], P(U) \ge \frac{s_+^{2d}}{s^{2d}}\}| \le C_0 s_1 \alpha^{\frac{2d}{m_0}} \le C_0 s_1 \alpha^{\frac{2d}{m}}.$$

□

Lemma 5.3. *Suppose that X is the space of polynomial of order m defined in a finite interval I. Then two norms*

$$|f(u)| = \max_{u\in I} |\sum_{i=2}^{m} a_i u^i|, ||f(u)|| = \max_{u\in I} \sum_{i=2}^{m} |a_i u^i|,$$

in X are equivalent, i.e., there is a constant C depending only on m, such that

$$|f(u)| \leq ||f(u)|| \leq C|f(u)|.$$

Acknowledgement. The work was done when the author visited the Mathematisches Institut der Universität zu Köln and Mathematisches Institut der Technischen Universiät München. He acknowledges the hospitality of both institutes. He would like to thank H.W. Broer, H.Hanssmann, T. Küpper, J. Pöschel, J. Scheurle and F. Takens for their interest and helpful comments. He also thanks the referee of this paper for his patient and careful reading of the manuscript and pointing out a number of errors in the previous version of the paper.

References

1. Arnold, V.I.: Proof of A. N. Kolmogorov's theorem on the preservation of quasi periodic motions under small perturbations of the Hamiltonian. Usp. Math. USSR **18**, 13–40 (1963)
2. Arnold, V.I.: Small denominators and problems of stability of motions in classical and celestial mechanics. Russ. Math. Surv. **18:6**, 85–191 (1963)
3. Bourgain, J.: Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE. Int. Math. Res. Notices. 475–497 (1994)
4. Bruno, A.D.: Analytic form of differential equations. Trudy MMO 25(1971), Trans. Moscow Math. Soc. **25** 131–288 (1971)
5. Cheng, C.Q.: Birkhoff-Kolmogorov-Arnold-Moser tori in convex Hamiltonian systems. Commun. Math. Phys. **177**, 529–559 (1996)
6. Cheng, C.-Q. and Sun, Y.-S.: Existence of KAM tori in degenerate Hamiltonian systems. J. Differ. Eqs. **114**, 288–335 (1994)
7. Craig, W. and Wayne, C.E.: Newton's method and periodic solutions of nonlinear wave equations. Commun. Pure Appl. Math. **46**, 1409–1498 (1993)
8. Graff, S.M.: On the continuation of stable invariant tori for Hamiltonian systems. J. Differ. Eqs. **15**, 1–69 (1974)
9. Eliasson, L.H.: Perturbations of stable invariant tori for Hamiltonian systems. Ann. Sc. Norm. Sup. Psia. **15**, 115–147 (1988)
10. Kolmogorov, A.N.: On quasi-periodic motions under small perturbations of the Hamiltonian. Dokl. Akad. Nauk. USSR. **98**, 1–20 (1954)
11. Kuksin, S.B.: Nearly integrable infinite dimensional Hamiltonian systems. *Lect. Notes in Math*. **1556** Berlin: Springer, 1993
12. Kuksin, S.B., Pöschel, J.: Invariant cantor manifolds of quasiperiodic oscillations for a nonlinear Schrödinger equation. Ann. of Math. **142**, 149–179 (1995)
13. Melnikov, V.K.: On some cases of the conservation of conditionally periodic motions under a small change of the Hamiltonian function. Sov. Math. Dokl. **6**, 1592–1596 (1965)
14. Moser, J.: Convergent series expansions for quasiperiodic motions. Math. Ann. **169(1)**, 136–176 (1967)
15. Moser, J.: Combination tones for Duffings equation. Comm.Pure Appl. Math. **18**, 1965
16. Moser, J. and Pöschel, J.: An extension of a result by Dinaburg and Sinai on quasi-periodic potentials. Comment. Math. Helv. **59**, 39–85 (1984)
17. Pöschel, J.: On elliptic lower dimensional tori in Hamiltonian systems. Math. Z. **202**, 559–608 (1989)
18. Pöschel, J.: A KAM-Theorem for some Nonlinear Partial Differential Equations. Ann. Scuola. Norm. sup. Pisa cl. **23**, 119–148 (1996)
19. Pöschel, J.: Integrability of Hamiltonian systems on Cantor Sets. Comm. Pure Appl. Math. **25**, 653–695 (1982)
20. Rüssmann, H.: On twist Hamiltonians. *Talk on the Colloque Int.: Mécanique céleste et systemes hamiltonians, Marseille*, 1990

21. Rüssman, H.: On the one dimensional Schrödinger equation with a quasiperiodic potential. Annals of the New York Acad. Sci. *375*, 90–107 (1980)
22. Takens, F.: Singularities of vector fields. Publ. Math. I.H.E.S. **43**, 479–528 (1974)
23. Wayne, C.E.: Periodic and quasi-periodic solutions for nonlinear wave equations via KAM theory. Commun. Math. Phys. **127**, 479–528 (1990)
24. Xia, Z.: The existence of the invariant tori in the volume preserving diffeomorphism. Erg. Th. & Dyn. Sys. **12**, 621–631 (1992)
25. Xu, J., You, J., Qiu, Q.: Invariant tori for nearly integrable Hamiltonian systems with degeneracy. To appear in Math. Z (1996)
26. Zehnder, E.: Generalized implicit function theorems with applications to some small divisor problem, I and II. Commun. Pure Appl. Math. **28**, 91–140 (1975), 49–111 (1976)

Communicated by A. Jaffe