Persistence of invariant tori on sub-manifolds in Hamiltonian systems

Shui-Nee Chow
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
Yong Li*
Department of Mathematics, Jilin University, Changchun 130023, PRC
Yingfei Yi†
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract

Generalizing the degenerate KAM theorem under the Rüssmann non-degeneracy and the isoenergetic KAM theorem, we employ a quasi-linear iterative scheme to study the persistence and frequency preservation of invariant tori on a smooth sub-manifold for a real analytic, nearly integrable Hamiltonian system. Under a nondegenerate condition of Rüssmann type on the sub-manifold, we shall show the following: a) the majority of the unperturbed tori on the sub-manifold will persist; b) the perturbed toral frequencies can be partially preserved according to the maximal degeneracy of the Hessian of the unperturbed system and be fully preserved if the Hessian is nondegenerate; c) the Hamiltonian admits normal forms near the perturbed tori of arbitrarily prescribed high order. Under a sub-isoenergetic nondegenerate condition on an energy surface, we shall show that the majority of unperturbed tori give rise to invariant tori of the perturbed system of the same energy which preserve the ratio of certain components of the respective frequencies.

Keywords. Hamiltonian system, KAM theory, persistence on sub-manifolds.

AMS(MOS). Mathematics Subject Classification. 58F05, 58F27, 58F30.

1 Introduction

We consider an analytic family of real analytic Hamiltonian systems of the following action-angle form

$$H = N(y) + \varepsilon P(x, y, \varepsilon), \tag{1.1}$$

where (x, y) lies in a complex neighborhood $\{(x, y) : |\mathrm{Im} x| < r, \mathrm{dist}(y, G) < \beta\}$ of $T^d \times G$, $G \subset R^d$ ($d > 1$) is a bounded closed region, and ε is a small parameter.

*Partially supported by NSFC grant 19971042, National 973 key Project: Nonlinearity in China and the outstanding youth project of Ministry of Education of China.

†Partially supported by NSF grant DMS9803581.

This work is partially done when the second and third authors were visiting the National University of Singapore.

With the symplectic form

$$\sum_{i=1}^{n} \mathrm{d}x_i \wedge \mathrm{d}y_i,$$

the associated unperturbed motion of (1.1) is simply described by the equation

$$\begin{cases} \dot{x} & = \omega(y), \\ \dot{y} & = 0, \end{cases}$$

where $\omega(y) = \dfrac{\partial N}{\partial y}(y)$. Thus, for $\varepsilon = 0$, the phase space $G \times T^d$ is foliated into invariant tori $T_y = \{y\} \times T^d$ with the frequency vectors $\omega(y)$, $y \in G$.

Under the Kolmogorov nondegenerate condition, i.e.,

K) *the Hessian* $A(y) \equiv \dfrac{\partial^2 N}{\partial y^2}(y)$ *is nonsingular for all* $y \in G$,

the classical KAM theorem (see Kolmogorov [17], Arnold [1], Moser [20]) says that the majority of the invariant d-tori will persist as ε sufficiently small. More precisely, there is a family of Cantor sets $G_\varepsilon \subset G$, with $|G \backslash G_\varepsilon| \to 0$, as $\varepsilon \to 0$, such that for each $y \in G_\varepsilon$, the torus T_y persists and gives rise to a slightly deformed, analytic, quasi-periodic, invariant torus T_y^ε of the perturbed system. Moreover, for each $y \in G_\varepsilon$ the perturbed torus T_y^ε preserves the frequency $\omega(y)$ of the corresponding unperturbed torus T_y.

Recently, a fair amount of attention was given to the persistence of a fixed Diophantine torus with the preservation of the toral frequency, see Benettin et. al. [4] for a KAM approach, Eliasson [12], Gallavotti [13], Chierchia and Falcolini [10] for a direct method using Lindstedt series, and Gallavotti, Gentile and Mastropietro [14] and Bricmont, Gawedzki, and Kupiainen [5] for using renormalization groups techniques. Important generalizations to the classical KAM theorem were also made for various degenerate cases (i.e., when the Hessian $A(y)$ becomes singular), see Bruno ([8]), Cheng and Sun ([9]), Rüssmann ([24]), Xu, You and Qiu ([27]), Sevryuk ([25]) and references therein. The persistence of KAM tori has been shown under various partially nondegenerate conditions. The weakest such condition was given by Rüssmann in [24] which says that the frequencies $\{\omega(y) : y \in G\}$ should not lie in any hyperplane of R^d. In the real analytic case, it is shown in [27] that the Rüssmann condition is equivalent to that

R) $\max_{y \in G} \mathrm{rank}\{\dfrac{\partial^\alpha \omega}{\partial y^\alpha} : |\alpha| \leq d-1\} = d.$

The matrix in the above is formed by d dimensional column vectors of all the partial derivatives of $\omega(y)$ of orders up to $d-1$.

In this paper, instead of the persistence of invariant tori in the whole domain of the action variable, we shall study the persistence problem on a given smooth sub-manifold M in the action space G, e.g., a curve or a surface, which is either closed or with boundary. Clearly, such persistence will depend on both the non-degeneracy of the unperturbed system and the differential structure of the sub-manifold. By using a quasi-linear iterative scheme introduced in [18], we shall show the following results for (1.1) as ε sufficiently small:

1) The majority of the unperturbed tori $\{T_y : y \in M\}$ will persist under a nondegenerate condition of Rüssmann type on M.

2) The maximal number of the preserved frequency components of a perturbed torus is characterized by the maximal rank of the Hessian matrices $\{A(y) : y \in M\}$.

3) If $A(y)$ is nonsingular on M, i.e., if the Kolmogorov nondegenerate condition is satisfied on M, then all Diophantine tori of the unperturbed system on M persist with unchanged toral frequencies.

4) If the unperturbed system admits a sub-isoenergetic non-degeneracy on an energy surface, then the majority of the unperturbed tori on the energy surface will persist and give rise to perturbed tori of the same energy, whose frequency ratios of the respective 'nondegenerate' components are preserved.

These results generalize both the KAM theorem under the Rüssmann non-degeneracy and the isoenergetic KAM theorem ([2],[3],[6]). Similar to the isoenergetic case, one interesting phenomenon is that the Kolmogorov and the Rüssmann non-degeneracy can be independent conditions on a sub-manifold of G, i.e., one can have the Kolmogorov but not the Rüssmann non-degeneracy on a sub-manifold and vice versa. In contrast to the KAM theory on the entire region G, the validity of the Kolmogorov nondegenerate condition on a sub-manifold does not automatically guarantee the existence of a Diophantine torus on the manifold (hence the persistence of any torus), unless the Rüssmann nondegenerate condition is also satisfied on the manifold. It should be noted the Rüssmann nondegenerate conditions on the whole domain and on a sub-manifold are also independent conditions. Hence the persistence on a particular sub-manifold does not follow from the Rüssmann non-degeneracy on the entire domain (see the examples in Section 2 for detail).

The quasi-linear scheme we employed follows the standard KAM iterative procedure but involves solving a system of quasi-linear equations at each KAM step instead of linear ones. This has the advantage of eliminating any prescribed number of high order angular-dependent terms in just one iteration, resulting in a normal form in the vicinity of a perturbed torus of arbitrarily high order.

The paper is organized as follows. In Section 2, we state our results with respect to both (1.1) and a parameterized Hamiltonian system, along with some discussion and examples. The quasi-linear iterative scheme will be described in Section 3 for one KAM cycle. We complete the proof of our results in Section 4 by deriving an iteration lemma and giving measure estimates.

Acknowledgment. The authors would like to thank the referees and editors for their valuable suggestions and comments which leads to a significant improvement of the paper.

2 Main results

Below, unless specified otherwise, we shall use the same symbol $|\cdot|$ to denote an equivalent (finite dimensional) vector norm and its induced matrix norm, absolute value of functions, and measure of sets etc., and use $|\cdot|_D$ to denote the supremum norm of functions on a domain D. Also, for any two complex column vectors ξ, ζ of the same dimension, $\langle \xi, \zeta \rangle$

always stands for $\xi^\top\zeta$, i.e., the transpose of ξ times ζ. For the sake of briefness, we shall not specify smoothness orders for functions having obvious orders of smoothness indicated by their derivatives taken. Moreover, all Hamiltonians in the sequel are associated to the standard symplectic structure.

We first consider the following parameter-dependent, real analytic Hamiltonian system

$$H = e(\lambda) + \langle \omega(\lambda), y\rangle + h(y,\lambda) + P(x,y,\lambda), \tag{2.1}$$

where (x,y) lies in a complex neighborhood $D(r,s) = \{(x,y) : \ |\mathrm{Im}x| < r, |y| < s\}$ of $T^d \times \{0\} \subset T^d \times R^d$, λ is a parameter lying in a bounded closed region $\Lambda \subset R^{d_0}$, $h(y,\lambda) = O(|y|^2)$. In the above, all λ dependence are of class C^{l_0} for some $l_0 \geq d$.

Write

$$h = \frac{1}{2}\langle y, A(\lambda)y\rangle + \hat{h}(y,\lambda),$$

where $A(\lambda)$ is real symmetric for each $\lambda \in \Lambda$ and $\hat{h}(y,\lambda) = O(|y|^3)$. We assume the following conditions:

A1) $\mathrm{rank}\{\frac{\partial^\alpha \omega}{\partial \lambda^\alpha} : |\alpha| \leq d-1\} = d$ *for all* $\lambda \in \Lambda$.

A2) $\mathrm{rank}A(\lambda) \equiv n$ *on* Λ*, and, there is a smoothly varying, nonsingular,* $n \times n$ *principal minor* $\mathcal{A}(\lambda)$ *of* $A(\lambda)$.

Remark 2.1 1) *In the case that* ω *is real analytic and* Λ *is connected, the condition* **A1)** *can be replaced by* $\max_{\lambda\in\Lambda} \mathrm{rank}\{\frac{\partial^\alpha \omega}{\partial \lambda^\alpha} : |\alpha| \leq d-1\} = d$*, which becomes the Rüssmann condition* **R)** *when* $d_0 = d$*. Indeed, as pointed out in* [27] *for the case* $d_0 = d$*, the Rüssmann condition implies that* **A1)** *holds on an open subset* $\Lambda_0 \subset \Lambda$ *with* $|\Lambda \setminus \Lambda_0| = 0$*.*

2) *The condition* **A2)** *is more or less automatic in the sense that if* $n = \max_{\lambda\in\Lambda} \mathrm{rank}A(\lambda) = A(\lambda_0)$*, then due to the symmetry of* $A(\lambda)$ *there is an* $n \times n$ *principal minor of* $A(\lambda)$ *which is smoothly varying and nonsingular in a neighborhood of* λ_0*.*

Denote $i_1, i_2, \cdots, i_n$ as the row indices (in the natural order) of $\mathcal{A}(\lambda)$ in $A(\lambda)$, and set $d_* = \max\{d_0, d\}$.

Our main result is as follows.

Theorem A. *Consider* (2.1) *and let* $m \geq 2$ *be a given integer.*

1) *Assume* **A1),A2)** *and let* $\tau > d(d-1)-1$ *be fixed. Then there exists a* $\mu = \mu(r,s,m,l_0,\tau) > 0$ *sufficiently small such that if*

$$|\partial_\lambda^l P|_{D(r,s)\times\Lambda} \leq \gamma^{2m+l_0+5} s^m \mu, \quad |l| \leq l_0, \tag{2.2}$$

then there exist Cantor sets $\Lambda_\gamma \subset \Lambda$ *with* $|\Lambda \setminus \Lambda_\gamma| = O(\gamma^{\frac{1}{d_*-1}})$ *and a* C^{l_0-1} *Whitney smooth family of* C^m *symplectic transformations*

$$\Psi_\lambda : D(\frac{r}{2}, \frac{s}{2}) \to D(r,s), \quad \lambda \in \Lambda_\gamma,$$

which is real analytic in x and C^m uniformly close to the identity such that

$$H \circ \Psi_\lambda(x,y) = e_*(\lambda) + \langle \omega_*(\lambda), y\rangle + h_*(y,\lambda) + P_*(x,y,\lambda), \tag{2.3}$$

where, for all $\lambda \in \Lambda_\gamma$ and $(x,y) \in D(\frac{r}{2},\frac{s}{2})$, $h_*(y,\lambda) = O(|y|^2)$, $P_*(x,y,\lambda) = O(|y|^{m+1})$,

$$\begin{aligned}
|\partial_\lambda^l (e_* - e)| &= O(\gamma^{m+l_0+4}\mu), \quad |l| \le l_0 - 1,\\
|\partial_\lambda^l (\omega_* - \omega)| &= O(\gamma^{m+l_0+4}\mu), \quad |l| \le l_0 - 1,\\
|\partial_\lambda^l \partial_y^j (h_* - h)| &= O(\gamma^{m+l_0+4}\mu^{\frac{1}{2}}), \quad |l| \le l_0 - 1,\ |j| \le m,
\end{aligned}$$

and moreover,

$$\begin{aligned}
|\langle k, \omega_*(\lambda)\rangle| &> \frac{\gamma}{|k|^\tau}, \quad \text{for all } k \in Z^d \setminus \{0\},\\
(\omega_*(\lambda))_{i_q} &\equiv (\omega(\lambda))_{i_q}, \quad \text{for all } 1 \le q \le n.
\end{aligned}$$

Thus, for each $\lambda \in \Lambda_\gamma$, the unperturbed torus $T_\lambda = T^d \times \{0\}$ associated to the toral frequency $\omega(\lambda)$ persists and gives rise to an analytic, Diophantine, invariant torus of the perturbed system with the toral frequency $\omega_*(\lambda)$ which preserves the frequency components $\omega_{i_1}(\lambda), \cdots, \omega_{i_n}(\lambda)$ of the unperturbed toral frequency $\omega(\lambda)$. Moreover, these perturbed tori form a C^{l_0-1} Whitney smooth family.

2) *Assume that $A(\lambda)$ is nonsingular on Λ and let $\tau > d-1$ be fixed. Then there exists a $\mu = \mu(r,s,m,l_0,\tau) > 0$ sufficiently small such that if (2.2) holds, then each Diophantine torus $T_\lambda = T^d \times \{0\}$, $\lambda \in \Lambda$, whose toral frequency $\omega(\lambda)$ having the Diophantine type (γ,τ), will persist, with the normal form (2.3), and gives rise to an analytic, Diophantine, invariant perturbed torus with the same toral frequency.*

In the above theorem, d_0 can be any positive integer. The case $d_0 > d$ will typically occur when the nondegenerate condition **A1)** fails with respect to the original parameters of a Hamiltonian system and extra deformation parameters need to be added so that a joint nondegenerate condition of type **A1)** can hold with respect to the extended parameters.

When $d_0 \le d$, the theorem has a direct application to nearly integrable Hamiltonian systems of form (1.1) with respect to the persistence of invariant tori on a sub-manifold of G.

Consider (1.1) and let M be a d_0 $(\le d)$ dimensional, C^{l_0} $(l_0 \ge d)$ sub-manifold of G which is either closed or with boundary. Denote

$$\omega(y) = \frac{\partial N}{\partial y}(y), \quad A(y) = \frac{\partial^2 N}{\partial y^2}(y), \quad y \in G.$$

We assume the following conditions:

A1)$'$ *For any coordinate chart (ϕ, U) of M, $\text{rank}\{\frac{\partial^\alpha(\omega\circ\phi^{-1})}{\partial\lambda^\alpha} : |\alpha| \le d-1\} = d$ for all $\lambda \in \phi(U) \subset R^{d_0}$.*

A2)$'$ rank$A(y) \equiv n$ *on* M*, and, there is a smoothly varying, nonsingular,* $n \times n$ *principal minor* $\mathcal{A}(y)$ *of* $A(y)$ *on* M*.*

Corollary. *Consider* (1.1)*. Let* $m \geq 2$ *be given and* r, β *be as in* (1.1)*.*

1) *Assume* **A1)**$'$,**A2)**$'$ *and let* $\tau > d(d-1)-1$ *be fixed. Then there is an* $\varepsilon_0 = \varepsilon_0(r, \beta, l_0, m, M, \tau) > 0$ *and a family of Cantor sets* $M_\varepsilon \subset M$, $0 < \varepsilon \leq \varepsilon_0$*, with* $|M \setminus M_\varepsilon| = O(\varepsilon^{\frac{1}{2(d_*-1)(2m+l_0+5)}})$*, where* $d_* = \max\{d_0, d\}$*, such that for each* $y \in M_\varepsilon$*, the unperturbed torus* T_y *persists and gives rise to an analytic, Diophantine, invariant torus of the perturbed system whose toral frequency* $\omega_\varepsilon(y)$ *satisfies*

$$\begin{aligned} |\langle k, \omega_\varepsilon(y)\rangle| &> \frac{\gamma}{|k|^\tau}, \quad \text{for all } k \in Z^d \setminus \{0\}, \\ (\omega_\varepsilon(y))_{i_q} &= (\omega(y))_{i_q}, \quad \text{for all } 1 \leq q \leq n, \end{aligned}$$

where $0 < \gamma \leq \varepsilon^{\frac{1}{2(2m+l_0+5)}}$, $i_1, \cdots, i_n$ *are the row indices* (*in the natural order*) *of* $\mathcal{A}(y)$ *located in* $A(y)$*. Moreover, these perturbed tori form a Whitney smooth family.*

2) *Assume that* $A(y)$ *is non-singular on* M *and let* $\tau > d-1$ *be fixed. Then each Diophantine torus* T_y, $y \in M$*, whose toral frequency* $\omega(y)$ *having the Diophantine type* (γ, τ) *for some* $0 < \gamma \leq \varepsilon^{\frac{1}{2(2m+l_0+5)}}$*, will persist and gives rise to an analytic, Diophantine, invariant perturbed torus with the same toral frequency.*

3) *Let* $y_0 \in M_\varepsilon$ *in* 1) *or* $\omega(y_0)$ *be Diophantine in* 2)*. Then* (1.1) *admits the following normal form on* $D_{y_0}(\frac{r}{2}, \frac{\beta}{2}) = \{(x,y) : |\text{Im}x| < r, |y - y_0| < \beta\}$*:*

$$H_{y_0}(x,y) = e_*(y_0) + \langle \omega_*(y_0), y - y_0\rangle + h_*(y, y_0) + P_*(x, y, y_0), \tag{2.4}$$

where $\omega_*(y_0)$ *is the toral frequency of the perturbed torus associated to* y_0 (*hence equals* $\omega(y_0)$ *in the case* 2))*,* $h_*(y,y_0) = O(|y-y_0|^2)$*,* $P_*(x,y,y_0) = O(|y-y_0|^{m+1})$*, which satisfy similar properties as described in part* 1) *of Theorem A with* $\mu = \varepsilon^{\frac{2}{2m+l_0+5}}, \gamma = \varepsilon^{\frac{1}{2(2m+l_0+5)}}$*.*

Remark 2.2 1) *Under the Kolmogorov or isoenergetic non-degeneracy, the arbitrarily high order normal forms of type* (2.4) *around a perturbed Diophantine torus plays the role of the classical Birkhoff normal forms and the existence of such has implications on the measure of the set of invariant tori around the torus* (*see* [11],[21],[22] *and references therein*)*. In particular, a more or less straightforward application of* Theorem 4 *and its proof in* [11] *to the normal form* (2.4) *gives rise to an exponential measure estimate of the set of invariant tori around a perturbed Diophantine torus of* (1.1)*. More precisely, if* y_0 *is as in part* 3) *of the Corollary,* $\mathcal{B}_\rho$ *is a ball in* R^d *centered at* y_0 *with sufficiently small radius* ρ*, and* $T_\rho \subset T^d \times \mathcal{B}_\rho$ *is the set of points lying in invariant d-tori of* (2.4)*, then* $|(T^d \times \mathcal{B}_\rho) \setminus T_\rho| = O(e^{-\frac{m}{8}})$*. Furthermore, corresponding to the original Hamiltonian* (1.1)*, one can choose* $m = [(c\gamma/\rho)^{\frac{1}{\tau+1}}]$ *for some constant* c *to conclude that the Lebesgue* *measure of the set of points in* $T^d \times \mathcal{B}_\rho$ *which do not lie in any invariant d-torus of* (1.1) *is of the order of* $O(exp\{-(c\gamma/16\rho)^{\frac{1}{\tau+1}}\})$ (*see* [11], page 293).

Given the above, it would be interesting to know whether the normal form (2.4) *can also lead to a similar measure estimate under the Rüssmann non-degeneracy, or more generally the condition* $\mathbf{A1)}'$. *In the later, relative measure estimates of a similar nature on a sub-manifold should be considered.*

2) *Both Theorem A and the Corollary trivially hold when* $d = 1$. *In this case, one can simply take* $\tau > 0, d_* = 2$ (*see* [19] *for more discussions*).

To illustrate the significance and application of the Corollary, we now consider (1.1) with $d = 2$ and assume that $N(y)$ has the form

$$N(y) = h_1(y_1) + h_2(y_2).$$

Particular examples of $N(y)$ to be considered are

$$\begin{aligned} N_1(y) &= y_1 + \frac{1}{2}y_2^2, \\ N_2(y) &= \frac{1}{2}y_1^2 + \frac{1}{3}y_2^3, \\ N_3(y) &= \frac{1}{2}y_1^2 + \frac{1}{2}y_2^2. \end{aligned}$$

It is clear that the Hessian matrices associated to N_1, N_2, N_3 read

$$A_1(y) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_2(y) = \begin{pmatrix} 1 & 0 \\ 0 & 2y_2 \end{pmatrix}, \quad A_3(y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

respectively.

Below, we discuss the application of the Corollary to the Hamiltonians above on three type of analytic plan curves: line segment, parabola, and circle. Since $d_0 = 1$ in these cases, the non-degenerate condition $\mathbf{A1)}'$ will give rise to certain twist conditions on these curves.

Example 1 (*Line segment*). Consider the line segment:

$$M_1 : y_1(\lambda) = a_1\lambda, \quad y_2(\lambda) = a_2\lambda, \quad \lambda \in [1, 2],$$

where $(a_1, a_2)^\top$ is a non-zero vector. Then it is easy to see that $\mathbf{A1)}'$ is equivalent to the twist condition

$$a_2 \frac{\partial h_1}{\partial y_1}\frac{\partial^2 h_2}{\partial y_2^2} - a_1 \frac{\partial h_2}{\partial y_2}\frac{\partial^2 h_1}{\partial y_1^2} \neq 0 \tag{2.5}$$

on M_1.

For N_1, (2.5) becomes $a_2 \neq 0$. Thus, if a_2 is non-zero and a_1 is arbitrarily given, then it follows from part 1) of the Corollary and the expression of A_1 that the majority of 2-tori on M_1 will persist with unchanged second components of toral frequencies. Since A_1 is singular, part 2) of the Corollary is not applicable.

For N_2, (2.5) becomes $a_1 a_2 \neq 0$. Thus, if both a_1 and a_2 are non-zero, then part 1) of the Corollary and the fact that $\text{rank}A_2 \equiv 2$ imply that all Diophantine 2-tori on M_2 will persist with unchanged toral frequencies. In fact, since A_2 is non-singular on M_1 if $a_2 \neq 0$, the same conclusion also follows from part 2) of the Corollary ($a_1 \neq 0$ is still required since Diophantine tori are considered). We note that only due to the application of part 1) of the Corollary one knows that the set of Diophantine 2-tori on M_2 is a nearly full measure set (hence non-empty).

For N_3, (2.5) will never be satisfied with any choice of a_1, a_2. Hence part 1) of the Corollary is not applicable. But since A_3 is always non-singular on M_1, one can apply part 2) of the Corollary to conclude that all Diophantine 2-tori on M_2 will persist with unchanged toral frequencies. However, we note in this special situation that a toral frequency of N_3 on M_1 is Diophantine if only if $(a_1, a_2)^\top$ is. Hence the above conclusion holds only if $(a_1, a_2)^\top$ is Diophantine, in which case the persistence of all 2-tori on M_1 follows.

Example 2 (*Parabola*). Consider the following parabola:

$$M_2 : y_1(\lambda) = a_1\lambda, \quad y_2(\lambda) = a_2\lambda^2, \quad \lambda \in [1, 2],$$

where $(a_1, a_2)^\top$ is a nonzero vector. Then it is easy to see that $\mathbf{A1)}'$ is equivalent to the twist condition

$$2a_2\lambda \frac{\partial h_1}{\partial y_1} \frac{\partial^2 h_2}{\partial y_2^2} - a_1 \frac{\partial h_2}{\partial y_2} \frac{\partial^2 h_1}{\partial y_1^2} \neq 0 \tag{2.6}$$

on M_2.

For N_1, (2.6) becomes $a_2 \neq 0$. Hence the same conclusion for N_1 in Example 1 is valid.

For N_2, (2.6) becomes $a_1 a_2 \neq 0$. Thus, the same conclusion for N_2 in Example 1 is valid.

For N_3, (2.6) also becomes $a_1 a_2 \neq 0$. Still, A_3 is always non-singular on M_2. Hence, if both a_1 and a_2 are non-zero, then one can apply either parts 1) 2) of the Corollary to conclude the persistence and the preservation of toral frequencies of all Diophantine 2-tori on M_2. But, again, it is because of the application of part 1) of the Corollary that one can actually conclude the existence of Diophantine 2-tori on M_2 in the case $a_1 a_2 \neq 0$.

Example 3 (*Circle*). Consider the unit circle:

$$M_3 : y_1(\lambda) = \cos 2\pi\lambda, \quad y_2(\lambda) = \sin 2\pi\lambda, \quad \lambda \in [0, 1].$$

Then it is easy to see that $\mathbf{A1)}'$ is equivalent to the twist condition

$$y_1 \frac{\partial h_1}{\partial y_1} \frac{\partial^2 h_2}{\partial y_2^2} + y_2 \frac{\partial h_2}{\partial y_2} \frac{\partial^2 h_1}{\partial y_1^2} \neq 0 \tag{2.7}$$

on M_3.

For N_1, (2.7) becomes $\cos 2\pi\lambda \neq 0$, i.e., $\mathbf{A1)}'$ is satisfied on M_3 except two points $(0, 1)^\top, (0, -1)^\top$. In view of Remark 2.1 1), one can still apply part 1) of the Corollary to conclude the persistence of the majority of invariant 2-tori on M_3 and the preservation of their second frequency components.

For N_2, (2.7) becomes $\sin 2\pi\lambda \neq 0$, i.e., $\mathbf{A1)}'$ is satisfied on M_3 except two points $(1, 0)^\top, (-1, 0)^\top$. Since the lower right minor of A_2 also vanishes at these points, the application of part 1) of the Corollary (based on Remark 2.1 1)) will only guarantee the persistence of the majority of invariant 2-tori on M_3 and the preservation of their first frequency components. In order to apply part 2) of the Corollary to obtain the persistence of all Diophantine 2-tori with unchanged toral frequencies, one needs to restrict to a closed portion of M_3 which does not contain these two points.

For N_3, (2.7) always holds and A_3 is always non-singular on M_3. One can use both parts of the Corollary to conclude the persistence and the frequency preservation of all Diophantine 2-tori on M_3.

We now consider the case that M is a fixed energy surface $\{N(y) = E\}$ in G. Then the usual isoenergetic non-degeneracy on M implies the Rüssmann non-degeneracy $\mathbf{A1)}'$ on M. Indeed, if (ϕ, U) is any coordinate chart on M, then one clearly has $(\frac{\partial N}{\partial y})^\top \frac{\partial \phi^{-1}}{\partial \lambda} \equiv 0$ on $\phi(U)$. It follows from the isoenergetic non-degeneracy that

$$\begin{aligned} d &= \operatorname{rank} \begin{pmatrix} \frac{\partial^2 N}{\partial y^2} & \frac{\partial N}{\partial y} \\ (\frac{\partial N}{\partial y})^\top & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial \phi^{-1}}{\partial \lambda} & 0 \\ 0 & 1 \end{pmatrix} \\ &= \operatorname{rank} \begin{pmatrix} \frac{\partial^2 N}{\partial y^2} \frac{\partial \phi^{-1}}{\partial \lambda} & \frac{\partial N}{\partial y} \\ (\frac{\partial N}{\partial y})^\top \frac{\partial \phi^{-1}}{\partial \lambda} & 0 \end{pmatrix} \\ &= \operatorname{rank} \begin{pmatrix} \frac{\partial^2 N}{\partial y^2} \frac{\partial \phi^{-1}}{\partial \lambda} & \frac{\partial N}{\partial y} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

on $\phi(U)$, hence

$$\operatorname{rank} \left(\frac{\partial^2 N}{\partial y^2} \frac{\partial \phi^{-1}}{\partial \lambda}, \frac{\partial N}{\partial y} \right) = \operatorname{rank} \left(\frac{\partial N}{\partial y}, \frac{\partial^2 N}{\partial y^2} \frac{\partial \phi^{-1}}{\partial \lambda} \right) = d$$

on $\phi(U)$. Thus, our Corollary asserts that the perturbed system admits invariant tori conjugating to certain unperturbed ones on the energy surface M, and moreover, the perturbed toral frequencies are preserved if the Kolmogorov non-degeneracy also holds on M. However, it should be noted that such perturbed tori do not lie on the same energy level E in general, simply because the perturbed tori on the same energy level are generally equivalent (not necessary conjugated) to the unperturbed ones and only the preservation of frequency ratios can be expected (see [6]).

To generalize the standard isoenergetic KAM theorem, it turns out that an additional sub-isoenergetic nondegenerate condition is needed besides the Rüssmann non-degeneracy on an energy surface. More precisely, let M be a sufficiently smooth, relatively open, bounded subset of $\{N(y) = E\}$. We assume $\mathbf{A1)}'$ on M and also the following *sub-isoenergetic non-degeneracy*:

$\mathbf{A1)}''$ *There is a smoothly varying $n \times n$ principal minor $\mathcal{A}(y)$ of $A(y)$ on M such that*

$$\det \begin{pmatrix} \mathcal{A}(y) & \omega^*(y) \\ \omega^*(y)^\top & 0 \end{pmatrix} \neq 0$$

on M, where $\omega^*(y) = \frac{\partial N}{\partial y_*}(y)$, $y_* = (y_{i_1}, \cdots, y_{i_n})^\top$, and $i_1, \cdots, i_n$ denote the row indices of $\mathcal{A}(y)$ in $A(y)$.

Theorem B. *Consider* (1.1). *Let* $m \geq 2$ *be given,* r, β *be as in* (1.1), *and* M *be a sufficiently smooth, relatively open, bounded subset of* $\{N(y) = E\}$.

1) *Assume* $\mathbf{A1)}'$ *on* M *and let* $\tau > d(d-1) - 1$ *be fixed. Then there is an* $\varepsilon_0 = \varepsilon_0(r, \beta, l_0, m, M, \tau) > 0$ *and a family of Cantor sets* $M_\varepsilon \subset M$, $0 < \varepsilon \leq \varepsilon_0$, *with* $|M \setminus M_\varepsilon| = O(\varepsilon^{\frac{1}{2(d_*-1)(2m+l_0+5)}})$, *where* $d_* = \max\{d_0, d\}$, *such that for each* $y \in M_\varepsilon$, *the unperturbed torus* T_y *persists and gives rise to an analytic, Diophantine, invariant torus* $T_{\varepsilon,y}$ *of the perturbed system on the energy surface* $\{H(x,y) = E\}$, *whose toral frequency* $\omega_\varepsilon(y)$ *satisfies*

$$|\langle k, \omega_\varepsilon(y)\rangle| \quad > \quad \frac{\gamma}{|k|^\tau} \quad \text{for all } k \in Z^d \setminus \{0\}.$$

Moreover, these perturbed tori form a local Whitney smooth family.

2) *If* $\mathbf{A1)}''$ *also holds on* M, *then each perturbed torus* $T_{\varepsilon,y}$ *preserves the ratio of the* $i_1, i_2, \cdots, i_n$ *components of its toral frequency* $\omega_\varepsilon(y)$, *i.e.,*

$$[\omega_{\varepsilon,i_1}(y) : \cdots : \omega_{\varepsilon,i_n}(y)] = [\omega_{i_1}(y) : \cdots : \omega_{i_n}(y)],$$

where $\omega_{\varepsilon,i_j}(y)$ *and* $\omega_{i_j}(y)$ *are the* i_j*-th components of* $\omega_\varepsilon(y)$ *and* $\omega(y) = \frac{\partial N}{\partial y}(y)$ *respectively, for* $j = 1, 2, \cdots, n$.

3) *For* $y_0 \in M_\varepsilon$, (1.1) *admits the same normal form as in part* 3) *of the Corollary.*

In the case that $\mathcal{A}(y) \equiv \frac{\partial^2 N(y)}{\partial y^2}(y)$, the condition $\mathbf{A1)}''$ coincides with the isoenergetic non-degeneracy which also implies the Rüssmann condition $\mathbf{A1)}'$ on the energy surface. Hence, part 2) of the Theorem B generalizes the standard isoenergetic KAM theorem.

3 KAM step

In this section, we describe the quasi-linear iterative scheme for the Hamiltonian (2.1) in one KAM cycle, say, from a νth KAM step to the $(\nu+1)$th-step. For simplicity, we set $l_0 = d$.

Consider (2.1) and define

$$\begin{aligned} &r_0 = r,\ \gamma_0 = 4\gamma,\ \beta_0 = s,\ \Lambda_0 = \Lambda,\ H_0 = H,\ e_0 = e, \\ &A_0 = A,\ \mathcal{A}_0 = \mathcal{A},\ h_0 = h,\ \hat{h}_0 = \hat{h},\ P_0 = P, \\ &N_0 = e_0(\lambda) + \langle \omega_0(\lambda), y\rangle + h_0. \end{aligned}$$

Without loss of generality, we assume that $0 < r_0, \beta_0, \gamma_0 \leq 1$ and $\mathcal{A}_0$ is the ordered $n \times n$ principal minor of A_0.

By monotonicity, we define μ_0, s_0 implicitly through the following equations

$$\begin{aligned} \mu &= \frac{4^{d+5}\mu_0}{(M^*+1)^m K_1^{m\tau}}, \\ s_0 &= \frac{\beta_0\gamma_0}{16(M^*+1)K_1^\tau}, \end{aligned} \tag{3.1}$$

where

$$M^* = \max_{|l|\le d, |j|\le m+5, |y|\le \beta_0, \lambda\in\Lambda_0} |\partial_\lambda^l \partial_y^j h_0(y,\lambda)|, \quad K_1 = ([\log \frac{1}{\mu_0}]+1)^{3\eta},$$

η is a fixed positive integer such that $(1+\sigma)^\eta > 2$ for $\sigma = \frac{1}{2(m+1)}$. It is clear that μ_0 is small if and only if μ is, and,

$$\mu_0 = o(\mu^{1-\epsilon}) \tag{3.2}$$

for any $0<\epsilon<1$. By making μ small, we assume without loss of generality that

$$16(M^*+1)K_1^\tau > 1.$$

Hence $s_0 < \min\{\beta_0, \gamma_0\}$.

Since

$$\frac{\mu}{\mu_0} = 4^{d+m+5}\left(\frac{s_0}{\beta_0\gamma}\right)^m, \tag{3.3}$$

(2.2) becomes

$$|\partial_\lambda^l P_0|_{D(r_0,s_0)} \le \gamma_0^{d+m+5} s_0^m \mu_0, \quad |l|\le d. \tag{3.4}$$

Now, suppose that after a νth-step, we have arrived at the following real analytic Hamiltonian:

$$\begin{aligned} H &= N+P, \\ N &= e(\lambda) + \langle \omega(\lambda), y\rangle + h(y,\lambda) \end{aligned} \tag{3.5}$$

which is defined on a phase domain $D(r,s)$ and depends smoothly on $\lambda\in\Lambda$, where $\Lambda\subset\Lambda_0$,

$$h = \frac{1}{2}\langle y, A(\lambda)y\rangle + \hat{h},$$

$\hat{h} = \hat{h}(y,\lambda) = O(|y|^3)$. In addition, suppose that the $n\times n$ ordered principal minor $\mathcal{A}$ of A is non-singular on Λ, and, $P = P(x,y,\lambda)$ satisfies

$$|\partial_\lambda^l P|_{D(r,s)} \le \gamma^{d+m+5} s^m \mu, \quad |l|\le d \tag{3.6}$$

for some $0<\mu\le\mu_0$, $0<\gamma\le\gamma_0$. By considering both averaging and translation, we shall find a symplectic transformation Φ_+, which, on a small phase domain $D(r_+,s_+)$ and a smaller parameter domain Λ_+, transforms the Hamiltonian (3.5) into the Hamiltonian of the next KAM cycle (the $(\nu+1)$th-step), i.e.,

$$H_+ = H\circ\Phi_+ = N_+ + P_+,$$

where N_+, P_+ enjoy similar properties as N, P respectively on $D(r_+,s_+)\times\Lambda_+$.

For simplicity, we shall omit index for all quantities of the present KAM step (the νth-step) and index all quantities (Hamiltonian, normal form, perturbation, transformation, and domains, etc) in the next KAM step (the $(\nu+1)$-th step) by " $+$ ". All constants $c_1 - c_7$ below are positive and independent of the iteration process, and, we shall also use c to denote any intermediate positive constant which is independent of the iteration process. To simplify the notations, we shall suspend the λ dependence in most terms of this section.

Define

$$
\begin{aligned}
r_+ &= \frac{r}{2}+\frac{r_0}{4},\\
s_+ &= \frac{1}{8}\alpha s, \quad \alpha=\mu^{2\sigma}=\mu^{\frac{1}{m+1}},\\
\beta_+ &= \frac{\beta}{2}+\frac{\beta_0}{4},\\
\gamma_+ &= \frac{\gamma}{2}+\frac{\gamma_0}{4},\\
K_+ &= ([\log\frac{1}{\mu}]+1)^{3\eta},\\
D_{\frac{i}{8}\alpha} &= D(r_+ + \frac{i-1}{8}(r-r_+), \frac{i}{8}\alpha s), \quad i=1,2,\cdots,8,\\
D(\xi) &= \{y\in C^d : |y|<\xi\}, \quad \xi>0,\\
\hat{D}(\xi) &= D(r_+ + \frac{7}{8}(r-r_+),\xi), \quad \xi>0,\\
D_+ &= D_{\frac{1}{8}\alpha}=D(r_+,s_+),\\
\tilde{D}_+ &= D(r_+ + \frac{3}{4}(r-r_+),\beta_+),\\
\Lambda_+ &= \{\lambda\in\Lambda : |\langle k,\omega(\lambda)\rangle| > \frac{\gamma}{|k|^\tau}, \text{ for all } 0<|k|\le K_+\},\\
\Gamma(r-r_+) &= \sum_{0<|k|\le K_+} |k|^{(d+m+6)\tau+d+m+6}e^{-|k|\frac{r-r_+}{8}}.
\end{aligned}
$$

3.1 Truncation

Consider the Taylor-Fourier series of P:

$$P=\sum_{k\in Z^d, \jmath\in Z_+^d} p_{k\jmath}y^\jmath e^{\sqrt{-1}\langle k,x\rangle}$$

and let R be the truncation of P of the form

$$R=\sum_{|k|\le K_+, |\jmath|\le m} p_{k\jmath}y^\jmath e^{\sqrt{-1}\langle k,x\rangle}. \tag{3.7}$$

Lemma 3.1 *Assume that*

H1) $\int_{K_+}^{\infty} t^{d+m}e^{-t\frac{r-r_+}{16}}\mathrm{d}t\le\mu.$

Then there is a constant c_1 such that for all $j\in Z_+^d$, $|l|\le d$, $\lambda\in\Lambda$,

$$
\begin{aligned}
|\partial_\lambda^l(P-R)|_{D_{\frac{7}{8}\alpha}} &\le c_1\gamma^{d+m+5}s^m\mu^2,\\
|\partial_\lambda^l R|_{D_{\frac{7}{8}\alpha}} &\le c_1\gamma^{d+m+5}s^m\mu.
\end{aligned}
$$

Proof: Without loss of generality, we let $\mu_0 \leq \frac{1}{8}$. Hence $\alpha \leq \frac{1}{2}$.
Let

$$\begin{aligned} I &= \sum_{|k|>K_+, \jmath \in Z_+^d} p_{k\jmath} y^\jmath e^{\sqrt{-1}\langle k,x\rangle}, \\ II &= \sum_{|k|\leq K_+, |\jmath|\geq m+1} p_{k\jmath} y^\jmath e^{\sqrt{-1}\langle k,x\rangle}. \end{aligned}$$

Then

$$P - R = I + II.$$

By using the standard Cauchy estimate, we have

$$\begin{aligned} |\partial_\lambda^l I|_{\hat{D}(s)} &\leq \sum_{|k|>K_+} |\partial_\lambda^l P|_{D(r,s)} e^{-|k|\frac{r-r_+}{8}} \leq \gamma^{d+m+5} s^m \mu \sum_{\kappa=K_+}^{\infty} \kappa^{d+m} e^{-\kappa\frac{r-r_+}{8}} \\ &\leq \gamma^{d+m+5} s^m \mu \int_{K_+}^{\infty} t^{d+m} e^{-t\frac{r-r_+}{16}} \mathrm{d}t \leq \gamma^{d+m+5} s^m \mu^2. \end{aligned}$$

It follows that

$$|\partial_\lambda^l (P - I)|_{\hat{D}(s)} \leq |\partial_\lambda^l P|_{D(r,s)} + |\partial_\lambda^l I|_{\hat{D}(s)} \leq 2\gamma^{d+m+5} s^m \mu.$$

For $|q| = m+1$, let $\int$ be the obvious anti-derivative of $\dfrac{\partial^q}{\partial y^q}$. Then the Cauchy estimate of $\partial_\lambda^l (P - I)$ on $\hat{D}(s)$ yields

$$\begin{aligned} |\partial_\lambda^l II|_{D_{\frac{7}{8}\alpha}} &= |\partial_\lambda^l \int \frac{\partial^q}{\partial y^q} \sum_{|k|\leq K_+, |\jmath|\geq m+1} p_{k\jmath} e^{\sqrt{-1}\langle k,x\rangle} y^\jmath \mathrm{d}y|_{D_{\frac{7}{8}\alpha}} \\ &\leq | \frac{c}{s^{m+1}} \int |\partial_\lambda^l (P - I - R)|_{\hat{D}(s)} \mathrm{d}y|_{D_{\frac{7}{8}\alpha}} \\ &\leq |\int c\gamma^{d+m+5} s^m \mu \cdot \frac{1}{s^{m+1}} \mathrm{d}y|_{D_{\frac{7}{8}\alpha}} \\ &\leq c(\alpha s)^{m+1} \gamma^{d+m+5} \frac{\mu}{s} = c\gamma^{d+m+5} s^m \mu^2. \end{aligned}$$

Thus,

$$|\partial_\lambda^l (P - R)|_{D_{\frac{7}{8}\alpha}} \leq c\gamma^{d+m+5} s^m \mu^2, \tag{3.8}$$

and therefore,

$$|\partial_\lambda^l R|_{D_{\frac{7}{8}\alpha}} \leq |P - R|_{D_{\frac{7}{8}\alpha}} + |P|_{D(r,s)} \leq c\gamma^{d+m+5} s^m \mu.$$

■

3.2 Averaging and quasi-linear equations

As usual, we shall construct the averaging transformation as the time 1-map ϕ_F^1 of the flow generated by a Hamiltonian F. Let F have the following form:

$$F = \sum_{0<|k|\leq K_+, |\jmath|\leq m} f_{k\jmath} y^\jmath e^{\sqrt{-1}\langle k,x\rangle}, \tag{3.9}$$

where $f_{k\jmath}$ are (matrix valued) functions of y.

Let $[R] = \int_{T^n} R(x,\cdot)\mathrm{d}x$ be the average of the truncation R defined in (3.7). Substituting F into the equation

$$\{N, F\} + R - [R] = 0 \tag{3.10}$$

yields

$$\begin{aligned} &- \sum_{0<|k|\le K_+,|\jmath|\le m} \sqrt{-1}\langle k, \omega(\lambda) + \partial_y h\rangle f_{k\jmath} y^\jmath e^{\sqrt{-1}\langle k,x\rangle} \\ &= - \sum_{0<|k|\le K_+,|\jmath|\le m} p_{k\jmath} y^\jmath e^{\sqrt{-1}\langle k,x\rangle}. \end{aligned}$$

By equating the coefficients above, we then obtain the following *quasi-linear equations*:

$$\sqrt{-1}\langle k, \omega(\lambda) + \partial_y h\rangle f_{k\jmath} = p_{k\jmath}, \quad |\jmath| \le m, \ 0 < |k| \le K_+. \tag{3.11}$$

Lemma 3.2 *Assume that*

H2) $\max_{|l|\le d, \ |j|\le m+5} |\partial_\lambda^l \partial_y^j h - \partial_\lambda^l \partial_y^j h_0|_{D(s)\times\Lambda_+} \le \mu_0^{\frac{1}{2}}$,

H3) $2s < \frac{\gamma-\gamma_+}{(M^*+1)K_+^{\tau+1}}$.

Then the quasi-linear equations (3.11) *can be uniquely solved on* $D(s)\times\Lambda_+$ *to obtain a family of functions* $f_{k\jmath}$ *which are analytic in* y*, smooth in* λ*, and satisfy the following properties:*

$$\bar f_{k\jmath}(\bar y, \lambda) = f_{-k\jmath}(y, \lambda), \tag{3.12}$$

$$|\partial_\lambda^l \partial_y^j f_{k\jmath}|_{D(s)\times\Lambda_+} \le c_2 |k|^{(|l|+|j|+1)\tau+|l|+|j|+1} \gamma^{d+m+4-|l|-|j|} s^{m-|\jmath|} e^{-|k|r} \tag{3.13}$$

for all $|\jmath| \le m, \ 0 < |k| \le K_+, \ |l| \le d, \ |j| \le m+4$*, where* c_2 *is a constant.*

Proof: Let $(y,\lambda) \in D(s)\times\Lambda_+$. By H2), H3),

$$|\partial_y h(y)| \le (M^*+1)|y| < (M^*+1)s < \frac{\gamma}{2|k|^{\tau+1}}.$$

It follows that

$$|\langle k, \omega(\lambda) + \partial_y h(y)\rangle| > \frac{\gamma}{|k|^\tau} - \frac{\gamma}{2|k|^\tau} = \frac{\gamma}{2|k|^\tau}.$$

Hence

$$L_k = \sqrt{-1}\langle k, \omega(\lambda) + \partial_y h(y)\rangle$$

is non-vanishing on Λ_+, and,

$$f_{k\jmath} = f_{k\jmath}(y,\lambda) = L_k^{-1} p_{k\jmath}$$

for all $(y,\lambda) \in D(s)\times\Lambda_+$, $0 < |k| \le K_+$, $|\jmath| \le m$, from which (3.12) clearly follows.

Let $0 < |k| \le K_+$. We note by the Cauchy estimate that

$$|\partial_\lambda^l p_{k\jmath}|_{\Lambda_+} \le |\partial_\lambda^l \partial_y^\jmath P|_{D(r,s)\times\Lambda_+} e^{-|k|r} \le \gamma^{d+m+5} s^{m-|\jmath|} \mu e^{-|k|r}, \quad |l| \le d, \quad |\jmath| \le m, \tag{3.14}$$

and by H2) that

$$|\partial_\lambda^l \partial_y^j L_k^{-1}|_{D(s)\times\Lambda_+} \le c|k|^{|j|+1}|L_k^{-1}|^{|j|+2} \le c\frac{|k|^{(|l|+|j|+1)\tau+|l|+|j|+1}}{\gamma^{|l|+|j|+1}}, \quad |l|\le d,\ |j|\le m+4.$$

Therefore,

$$\begin{aligned}|\partial_\lambda^l \partial_y^j f_{kj}|_{D(s)\times\Lambda_+} &\le c\frac{|k|^{(|l|+|j|+1)\tau+|l|+|j|+1}}{\gamma^{|l|+|j|+1}}\gamma^{d+m+5}s^{m-|j|}e^{-|k|r} \\ &= c|k|^{(|l|+|j|+1)\tau+|l|+|j|+1}\gamma^{d+m+4-|l|-|j|}s^{m-|j|}e^{-|k|r}, \quad |l|\le d,\ |j|\le m+4.\end{aligned}$$

■

Let F be the Hamiltonian (3.9) with coefficients given by Lemma 3.2. If ϕ_F^t denotes the flow generalized by F, then

$$H\circ\phi_F^1 = \bar{N}_+ + \bar{P}_+,$$

where

$$\begin{aligned}\bar{N}_+ &= N+[R], \\ \bar{P}_+ &= \int_0^1 \{R_t, F\}\circ\phi_F^t \mathrm{d}t + (P-R)\circ\phi_F^1,\end{aligned} \tag{3.15}$$

with

$$R_t = (1-t)[R] + tR.$$

This completes the averaging process.

3.3 Translation and partial non-degeneracy

Let Y, P_{01} be the vectors formed by the first n components of y, p_{01} respectively and denote $\hat{H}(Y) = \hat{h}(\binom{Y}{0})$. Then by the implicit function theorem, the equation

$$\mathcal{A}Y + \partial_Y \hat{H}(Y) = -P_{01} \tag{3.16}$$

admits a unique solution Y^* on $D(s)$ which also smoothly depends on λ. Define

$$y^* = \begin{pmatrix} Y^* \\ 0 \end{pmatrix}.$$

By (3.16), we clearly have

$$Ay^* + \partial_y \hat{h}(y^*) = -\begin{pmatrix} P_{01} \\ 0 \end{pmatrix}. \tag{3.17}$$

Consider the translation

$$\phi:\ x\to x, \quad y\to y+y^*$$

and let

$$\Phi_+ = \phi_F^1\circ\phi.$$

Then

$$\begin{aligned}
H \circ \Phi_+ &= N_+ + P_+, \\
N_+ &= \bar{N}_+ \circ \phi - \psi = e_+ + \langle \omega_+, y \rangle + h_+(y), \\
P_+ &= \bar{P}_+ \circ \phi + \psi, \qquad (3.18)
\end{aligned}$$

where

$$e_+ = e + \langle \omega, y^* \rangle + \frac{1}{2}\langle y^*, Ay^* \rangle + \hat{h}(y^*) + [R](y^*), \qquad (3.19)$$

$$\omega_+ = \omega + p_{01} - \begin{pmatrix} P_{01} \\ 0 \end{pmatrix}, \qquad (3.20)$$

$$h_+(y) = \frac{1}{2}\langle y, A_+ y \rangle + \hat{h}_+(y), \qquad (3.21)$$

$$A_+ = A + \partial_y^2 \hat{h}(y^*) + \partial_y^2 [R](y^*), \qquad (3.22)$$

$$\begin{aligned}
\hat{h}_+(y) &= \hat{h}(y+y^*) - \hat{h}(y^*) - \langle \partial_y \hat{h}(y^*), y \rangle - \frac{1}{2}\langle y, \partial_y^2 \hat{h}(y^*) y \rangle \qquad (3.23) \\
&\quad + [R](y+y^*) - [R](y^*) - \langle \partial_y [R](y^*), y \rangle - \frac{1}{2}\langle y, \partial_y^2 [R](y^*) y \rangle,
\end{aligned}$$

$$\psi = \langle \partial_y [R](y^*), y \rangle - \langle p_{01}, y \rangle = \sum_{2 \le |\jmath| \le m, |\jmath - \jmath'| \le m-1, |\jmath'| = 1} \begin{pmatrix} \jmath \\ 1 \end{pmatrix} p_{0\jmath} y^{*\jmath - \jmath'} y. \qquad (3.24)$$

3.4 Estimate on N_+

Lemma 3.3 *Assume* H2), H3). *Then there is a constant c_3 such that the following holds for all $|l| \le d$:*

$$|\partial_\lambda^l y^*|_{\Lambda_+} \le c_3 \gamma^{d+m+5} s^{m-1} \mu; \qquad (3.25)$$

$$|\partial_\lambda^l e_+ - \partial_\lambda^l e|_{\Lambda_+} \le c_3 \gamma^{d+m+5} s^{m-1} \mu; \qquad (3.26)$$

$$|\partial_\lambda^l \omega_+ - \partial_\lambda^l \omega|_{\Lambda_+} \le c_3 \gamma^{d+m+5} s^{m-1} \mu; \qquad (3.27)$$

$$|\partial_\lambda^l \partial_y^j h_+ - \partial_\lambda^l \partial_y^j h|_{D(s_+) \times \Lambda_+} \le \begin{cases} c_3 \gamma^{d+m+5} s^{m-|j|} \mu, & |j| \le m; \\ c_3 \gamma^{d+m+5} \mu, & |j| > m. \end{cases} \qquad (3.28)$$

Proof: Denote $M_* = \max_{\lambda \in \Lambda_0} |\mathcal{A}_0^{-1}(\lambda)| + 1$. By (3.1), we can make μ_0 small, say $\mu_0 < \frac{1}{8M_*^2(M^*+1)}$, such that $M_*(M^*+1)s_0^2 < \frac{1}{4}$.

Let $\lambda \in \Lambda_+$. To prove (3.25), we denote

$$B(y) = \mathcal{A} + (\int_0^1 \partial_y^2 \hat{H}(\theta y) \mathrm{d}\theta) y.$$

Then by (3.16),

$$B(Y^*)Y^* = -P_{01}. \qquad (3.29)$$

Since, by H2), $|A - A_0|_\Lambda \le \mu_0^{\frac{1}{2}}$ and $|\partial_y^2 \hat{H}|_{D(s)} \le (M^*+1)s$, we have that

$$\begin{aligned}
|\mathcal{A}_0 - B(Y^*)| &\le |A - A_0| + |B(Y^*) - \mathcal{A}_0| \le \mu_0^{\frac{1}{2}} + (M^*+1)s^2 \\
&\le \mu_0^{\frac{1}{2}} + (M^*+1)s_0^2 \le \frac{1}{2M_*}.
\end{aligned}$$

It follows that $B(Y^*)$ is non-singular and

$$|B^{-1}(Y^*)| \leq \frac{|\mathcal{A}_0^{-1}|}{1 - |\mathcal{A}_0 - B(Y^*)||\mathcal{A}_0^{-1}|} \leq 2M_*.$$

Hence

$$|y^*| = |Y^*| \leq 2M_*|P_{01}| = 2M_*|p_{01}| \leq 2M_*|\partial_y P|_{D(s)} \leq 2M_*\gamma^{d+m+5}s^{m-1}\mu. \tag{3.30}$$

Differentiating (3.29) with respect to λ yields

$$B(Y^*)\partial_\lambda Y^* + \partial_y B(Y^*)(\partial_\lambda Y^*)Y^* + \partial_\lambda B(Y^*)Y^* = -\partial_\lambda P_{01}.$$

Therefore,

$$|\partial_\lambda Y^*| \leq 4M_*^2(M^*+1)\gamma^{d+m+5}s^m\mu|\partial_\lambda Y^*| + M_*c|Y^*| + M_*c|\partial_\lambda\partial_y P|_{D(s)}.$$

The estimate (3.25) now follows from (3.6), (3.30) and induction. Using (3.19) ((3.20) respectively), (3.26) ((3.27) respectively) easily follows from H2), (3.25) and (3.14). Also, it follows from (3.22) that

$$|\partial_\lambda^l A_+ - \partial_\lambda^l A|_{\Lambda_+} \leq c_3\gamma^{d+m+5}s^{m-2}\mu. \tag{3.31}$$

Note by (3.5) that

$$\begin{aligned}
\hat{h}_+ &= \sum_{|\jmath|\geq 3}\frac{1}{\jmath!}\partial_y^\jmath\hat{h}(y^*)y^\jmath + \sum_{3\leq|\jmath|\leq m}\frac{1}{\jmath!}\partial_y^\jmath[R](y^*)y^\jmath \\
&= \sum_{|\jmath|\geq 3}\frac{1}{\jmath!}\partial_y^\jmath\hat{h}(y^*)y^\jmath + \sum_{3\leq|\jmath|\leq|\imath|\leq m}\binom{\imath}{\jmath}p_{0\imath}(y^*)^{\imath-\jmath}y^\jmath.
\end{aligned}$$

We have that

$$\hat{h}_+ - \hat{h} = \sum_{|\jmath|\geq 3}\frac{1}{\jmath!}(\partial_y^\jmath\hat{h}(y^*) - \partial_y^\jmath\hat{h}(0))y^\jmath + \sum_{3\leq|\jmath|\leq|\imath|\leq m}\binom{\imath}{\jmath}p_{0\imath}(y^*)^{\imath-\jmath}y^\jmath.$$

Therefore,

$$|\partial_\lambda^l\partial_y^j\hat{h}_+ - \partial_\lambda^l\partial_y^j\hat{h}|_{D(s)\times\Lambda_+} \leq \begin{cases} c_3\gamma^{d+m+5}s^{m-|j|}\mu, & |j|\leq m; \\ c_3\gamma^{d+m+5}\mu, & |j| > m. \end{cases}$$

Combining the above with (3.31), we obtain (3.28). ■

3.5 Estimate on Φ_+

Let F be as in (3.9) with coefficients given by Lemma 3.2. By (3.12), F is real analytic in $(x, y) \in D(r, s)$.

Lemma 3.4 *Assume* H2), H3). *Then the following holds.*

1) *There is a constant c_4 such that for all $|l| \le d,\ |i| \le m+4$,*

$$|\partial_\lambda^l \partial_x^i \partial_y^j F|_{\hat{D}(s)\times\Lambda_+} \le \begin{cases} c_4 \gamma^{d+m+4-|l|-|j|} s^{m-|j|} \mu \Gamma(r-r_+), & |j| \le m; \\ c_4 \gamma^{d+m+4-|l|-|j|} \mu \Gamma(r-r_+), & m < |j| \le m+4. \end{cases}$$

2) *F, y^* can be extended to functions of Hölder class $C^{m+3,d-1+\sigma_0}(\hat{D}(\beta_0)\times\Lambda_0)$, $C^{d-1+\sigma_0}(\Lambda_0)$, respectively, where $0<\sigma_0<1$ is fixed. Moreover, there is a constant c_5 such that*

$$\begin{aligned} \|F\|_{C^{m+3,d-1+\sigma_0}(\hat{D}(\beta_0)\times\Lambda_0)} &\le c_5 \mu \Gamma(r-r_+), \\ \|y^*\|_{C^{d-1+\sigma_0}(\Lambda_0)} &\le c_5 \mu \Gamma(r-r_+). \end{aligned}$$

Proof: By (3.9), (3.13), we have

$$\begin{aligned} |\partial_\lambda^l \partial_x^i \partial_y^j F| &\le c \sum_{|\jmath|\le m, 0<|k|\le K_+} |k|^i |\partial_y^j (\partial_\lambda^l f_{k\jmath} y^\jmath)| e^{|k|(r_+ + \frac{7}{8}(r-r_+))} \\ &\le c \sum_{0<|k|\le K_+} |k|^{(|l|+|j|+1)\tau+|l|+|i|+|j|+1} \gamma^{d+m+4-|l|-|j|} s^{a(|j|)} \mu e^{-|k|\frac{r-r_+}{8}} \\ &\le c \gamma^{d+m+4-|l|-|j|} s^{a(|j|)} \mu \Gamma(r-r_+), \end{aligned}$$

where

$$a(|j|) = \begin{cases} m-|j|, & \text{if } |j| \le m, \\ 0, & \text{if } m<|j|\le m+4. \end{cases}$$

This proves 1).

2) follows from the standard Whitney extension theorem (see [22], [26]). ■

Lemma 3.5 *In addition to* H2), H3)*, assume that*

H4) $c_4 s^{m-1} \mu \Gamma(r-r_+) < \frac{1}{8}(r-r_+)$;

H5) $c_4 s^m \mu \Gamma(r-r_+) < \frac{1}{8}\alpha s$;

H6) $c_3 s^{m-1}\mu < \frac{1}{8}\alpha s$.

Then for all $0 \le t \le 1$,

$$\phi_F^t : D_{\frac{1}{4}\alpha} \longrightarrow D_{\frac{1}{2}\alpha}, \tag{3.32}$$

$$\phi : D_{\frac{1}{8}\alpha} \to D_{\frac{1}{4}\alpha} \tag{3.33}$$

are well defined, real analytic and depend smoothly on $\lambda \in \Lambda_+$.

Proof: (3.33) follows immediately from Lemma 3.3 and H6).

To show (3.32), we write $\phi_F^t = (\phi_1^t, \phi_2^t)^\top$, where ϕ_1^t, ϕ_2^t are components of ϕ_F^t in the directions x, y respectively. Let $(x,y) \in D_{\frac{1}{4}\alpha}$ and let $t_* = \mathrm{Sup}\{t \in [0,1] : \phi_F^t(x,y) \in D_\alpha\}$. Then for any $0 \le t \le t_*$,

$$\begin{aligned} |\phi_{F1}^t(x,y) - x| &\le \int_0^t |F_y \circ \phi_F^u|_{D_\alpha} \mathrm{d}u| \le |F_y|_{\hat{D}(s)} \le c_4 s^{m-1}\mu\Gamma(r-r_+) < \frac{1}{16}(r-r_+), \\ |\phi_{F2}^t(x,y) - y| &\le \int_0^t |F_x \circ \phi_F^u|_{D_\alpha} \mathrm{d}u| \le |F_x|_{\hat{D}(s)} \le c_4 s^m \mu\Gamma(r-r_+) < \frac{1}{16}\alpha s. \end{aligned}$$

It follows that $|\phi_{F1}^t(x,y)| < r_+ + \frac{3}{8}(r-r_+)$, $|\phi_{F2}^t(x,y)| < \frac{3}{8}\alpha s$, i.e., $\phi_F^t(x,y) \in D_{\frac{1}{2}\alpha} \subset D_\alpha$. Thus, $t_* = 1$ and (3.32) holds. ■

The above lemma implies that $\Phi_+ : D_+ \to D_{\frac{1}{2}\alpha}$ is well defined, symplectic and real analytic for all $\lambda \in \Lambda_+$. We now consider Φ_+ on the domain $\tilde{D}_+$.

Lemma 3.6 *Assume* H2), H3) *and also the following:*

H7) $c_5\mu\Gamma(r-r_+) < \frac{1}{8}(r-r_+)$;

H8) $c_5\mu\Gamma(r-r_+) + c_3\delta\mu < \beta - \beta_+$.

Let F, y^* be the extended functions defined in Lemma 3.4 2). *Then*

$$\Phi_+ = \phi_F^1 \circ \phi : \hat{D}_+ \to D(r,\beta) \tag{3.34}$$

is of class C^{m+2} and also depends $C^{d-1+\sigma_0}$ smoothly on $\lambda \in \Lambda_0$, where σ_0 is as in Lemma 3.4 2). *Moreover, there is a constant c_6 such that*

$$\|\Phi_+ - id\|_{C^{m+2,d-1+\sigma_0}(\tilde{D}_+\times\Lambda_0)} \le c_6\mu\Gamma(r-r_+). \tag{3.35}$$

Proof: By a similar argument as in Lemma 3.5, it is easy to see that Φ_+ maps $\hat{D}_+$ into $D(r,\beta)$ for all $\lambda \in \Lambda_0$.

Let $X_F = (F_y, -F_x)^\top$ be the vector field generated by F. We note that

$$\begin{aligned} \phi_F^t &= \mathrm{id} + \int_0^t X_F \circ \phi_F^u \mathrm{d}u, \quad 0 \le t \le 1, \\ \|X_F\|_{C^{m+2,d-1+\sigma_0}(\hat{D}(\beta_0)\times\Lambda_0)} &\le c\|F\|_{C^{m+3,d-1+\sigma_0}(\hat{D}(\beta_0)\times\Lambda_0)}. \end{aligned}$$

By applying Lemma 3.4 2) and the Gronwall inequality inductively, we have that, on $\tilde{D}_+ \times \Lambda_0$,

$$|\phi^t - id|,\ |\partial_y\phi_F^t - I_{2n}|,\ |\partial_y^j\phi_F^t| \le c\mu\Gamma(r-r_+), \quad 2 \le |j| \le m+2,\ 0 \le t \le 1. \tag{3.36}$$

The lemma now follows from Lemma 3.4 2) and the identity

$$\Phi_+ - id = (\phi_F^1 - id)\circ\phi + \begin{pmatrix} 0 \\ y^* \end{pmatrix}. \tag{3.37}$$

■

3.6 Frequency property

Lemma 3.7 *Assume H2),H3),H6). Then*

$$|\langle k, \omega_+(\lambda)\rangle| > \frac{\gamma_+}{|k|^\tau},$$

for all $\lambda \in \Lambda_+$ and $0 < |k| \le K_+$.

Proof: By H3), H6), we have

$$c_3 s^{m+1}\mu K_+^{\tau+1} < \gamma - \gamma_+.$$

It follows from Lemma 3.3 that

$$\begin{aligned} |\langle k, \omega_+(\lambda)\rangle| &\ge |\langle k, \omega(\lambda)\rangle| - c_3\gamma_0 s^{m-1}\mu K_+ \\ &\ge \frac{\gamma}{|k|^\tau} - c_3\gamma_0 s^{m-1}\mu K_+ > \frac{\gamma_+}{|k|^\tau}, \end{aligned}$$

as desired. ■

3.7 Estimate on P_+

Lemma 3.8 *Assume* H1)-H6). *Then, there is a constant c_7 such that, on $D_+ \times \Lambda_+$,*

$$|\partial_\lambda^l P_+| \leq c_7 \gamma^{d+m+5} s^m \mu^2 (\Gamma^2(r-r_+)+1), \quad |l| \leq d. \tag{3.38}$$

Proof: By Lemmas 3.1, 3.4 1) and (3.36), we see that, for all $|l| \leq d$, $0 \leq t \leq 1$,

$$\begin{aligned} |\partial_\lambda^l \{R_t, F\} \circ \phi_F^t|_{D_{\frac{1}{4}\alpha} \times \Lambda_+} &\leq c\gamma^{d+m+5} s^m \mu^2 \Gamma^2 (r-r_+), \\ |\partial_\lambda^l (P-R) \circ \phi_F^1|_{D_{\frac{1}{4}\alpha} \times \Lambda_+} &\leq c\gamma^{d+m+5} s^m \mu^2 \Gamma (r-r_+). \end{aligned}$$

Hence, by (3.15),

$$|\partial_\lambda^l \bar{P}_+|_{D_{\frac{1}{4}\alpha} \times \Lambda_+} \leq c\gamma^{d+m+5} s^m \mu^2 (\Gamma^2(r-r_+)+1), \ |l| \leq d.$$

Since, by (3.14),

$$|\partial_\lambda^l p_{0\jmath}|_{\Lambda_+} \leq c\gamma^{d+m+5} s^m \mu, \quad |l| \leq d,$$

it follows from (3.24), (3.25) that

$$|\partial_\lambda^l \psi|_{D_+ \times \Lambda_+} \leq \gamma^{d+m+5} s^{m-|\jmath|} \mu^2, \quad |l| \leq d.$$

By (3.25), we also have

$$|\partial_\lambda^l \phi|_{D_+ \times \Lambda_+} \leq c\gamma^{d+m+5} s^{m-1} \mu, \quad |l| \leq d.$$

The lemma now follows from (3.18) and the above estimates. ∎

Let $c_0 = \max\{1, c_1, \cdots, c_7\}$ and define

$$\mu_+ = 8^m c_0 \mu^{1+\sigma}.$$

If we assume that

H9) $\mu^\sigma (\Gamma^2(r-r_+)+1) \leq \frac{\gamma_+^{d+m+5}}{\gamma^{d+m+5}}$,

then, on $D_+ \times \Lambda_+$,

$$\begin{aligned} |\partial_\lambda^l P_+| &\leq 8^m c_0 s_+^m \mu^{1+\sigma} \mu^{1-2\sigma-\frac{m}{m+1}} (\mu^\sigma \gamma^{d+m+5} (\Gamma^2(r-r_+)+1)) \\ &\leq \gamma_+^{d+m+5} s_+^m \mu_+, \quad |l| \leq d. \end{aligned}$$

This completes one cycle of KAM steps.

4 Proof of Main Results

4.1 Iteration Lemma

Consider (2.1) and let $r_0, s_0, \gamma_0, \beta_0, \mu_0, \Lambda_0, H_0, N_0, e_0, \omega_0, h_0, A_0, \hat{h}_0, P_0$ be given at the beginning of Section 3 and let $\hat{D}_0 = D(r_0, \beta_0), K_0 = 0$. We define the following sequences inductively for all $\nu = 1, 2, \cdots$:

$$
\begin{aligned}
r_\nu &= r_0(1 - \sum_{i=1}^{\nu} \frac{1}{2^{i+1}}), \\
s_\nu &= \frac{1}{8}\alpha_{\nu-1}s_{\nu-1}, \\
\alpha_\nu &= \mu_\nu^{2\sigma} = \mu_\nu^{\frac{1}{m+1}}, \\
\mu_\nu &= 8^m c_0 \mu_{\nu-1}^{1+\sigma}, \\
\beta_\nu &= \beta_0(1 - \sum_{i=1}^{\nu} \frac{1}{2^{i+1}}), \\
\gamma_\nu &= \gamma_0(1 - \sum_{i=1}^{\nu} \frac{1}{2^{i+1}}), \\
K_\nu &= ([\log(\frac{1}{\mu_{\nu-1}})] + 1)^{3\eta}, \\
\Lambda_\nu &= \{\lambda \in \Lambda_{\nu-1} : |\langle k, \omega_{\nu-1}(\lambda)\rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau} \quad \text{for all } 0 < |k| \le K_\nu\}, \\
D_\nu &= D(r_\nu, s_\nu), \\
\tilde{D}_\nu &= D(r_\nu + \frac{3}{4}(r_{\nu-1} - r_\nu), \beta_\nu).
\end{aligned}
$$

Lemma 4.1 *If* (3.4) *holds for a sufficiently small* $\mu_0 = \mu_0(r_0, \beta_0, m, d, \tau)$*, or equivalently,* $\mu = \mu(r, s, m, d, \tau)$*, then the KAM step described in* Section 3 *is valid for all* $\nu = 0, 1, \cdots$*, resulting in sequences:*

$$\Lambda_\nu,\ H_\nu,\ N_\nu,\ e_\nu,\ \omega_\nu,\ h_\nu,\ A_\nu,\ \hat{h}_\nu,\ P_\nu,\ \Phi_\nu,$$

$\nu = 1, 2, \cdots$*, with the following properties.*

1) $\Phi_\nu : \hat{D}_\nu \times \Lambda_0 \longrightarrow \hat{D}_{\nu-1}$, $D_\nu \times \Lambda_\nu \longrightarrow D_{\nu-1}$ *is symplectic for each* $\lambda \in \Lambda_0$ *or* Λ_ν*, and is of class* $C^{m+2,d-1+\sigma_0}$, $C^{\alpha,d}$*, respectively, where* α *stands for real analyticity and* $0 < \sigma_0 < 1$ *is fixed, and,*

$$\|\Phi_\nu - id\|_{C^{m+2,d-1+\sigma_0}(\hat{D}_\nu \times \Lambda_0)} \le \frac{\mu^{\frac{1}{2}}}{2^\nu}. \tag{4.1}$$

Moreover, on $\hat{D}_\nu \times \Lambda_\nu$*,*

$$H_\nu = H_{\nu-1} \circ \Phi_\nu = N_\nu + P_\nu,$$

where

$$
\begin{aligned}
H_\nu &= N_\nu + P_\nu, \\
N_\nu &= e_\nu + \langle \omega_\nu, y\rangle + h_\nu, \\
h_\nu &= \frac{1}{2}\langle y, A_\nu y\rangle + \hat{h}_\nu,
\end{aligned}
$$

A_ν *is real symmetric with its* $n \times n$ *ordered principal minor* $\mathcal{A}_\nu$ *being non-singular on* Λ_ν*,* $\hat{h}_\nu = O(|y|^3)$*.*

2) $(\omega_\nu(\lambda))_q = (\omega_\nu(\lambda))_q$ *for all* $q = 1, 2, \cdots, n$ *and* $\lambda \in \Lambda_\nu$.

3) *For all* $|l| \le d$,

$$|\partial_\lambda^l e_\nu - \partial_\lambda^l e_{\nu-1}|_{\Lambda_\nu} \le \gamma_0^{d+m+4} \frac{\mu}{2^\nu}, \tag{4.2}$$

$$|\partial_\lambda^l e_\nu - \partial_\lambda^l e_0|_{\Lambda_\nu} \le \gamma_0^{d+m+4} \mu, \tag{4.3}$$

$$|\partial_\lambda^l \omega_\nu - \partial_\lambda^l \omega_{\nu-1}|_{\Lambda_\nu} \le \gamma_0^{d+m+4} \frac{\mu}{2^\nu}, \tag{4.4}$$

$$|\partial_\lambda^l \omega_\nu - \partial_\lambda^l \omega_0|_{\Lambda_\nu} \le \gamma_0^{d+m+4} \mu, \tag{4.5}$$

$$|\partial_\lambda^l \partial_y^j h_\nu - \partial_\lambda^l \partial_y^j h_{\nu-1}|_{D(s_\nu)\times\Lambda_\nu} \le \gamma_0^{d+m+4} \frac{\mu^{\frac{1}{2}}}{2^\nu}, \quad |j| \le m+1, \tag{4.6}$$

$$|\partial_\lambda^l \partial_y^j h_\nu - \partial_{D(s_\nu)\times\lambda}^l \partial_y^j h_0|_{\Lambda_\nu} \le \gamma_0^{d+m+4} \mu^{\frac{1}{2}}, \quad |j| \le m+1, \tag{4.7}$$

$$|\partial_\lambda^l P_\nu|_{D_\nu\times\Lambda_\nu} \le \gamma_\nu^{d+m+5} s_\nu^m \mu_\nu. \tag{4.8}$$

4) $\Lambda_\nu = \{\lambda \in \Lambda_{\nu-1} : |\langle k, \omega_{\nu-1}(\lambda)\rangle| > \dfrac{\gamma_{\nu-1}}{|k|^\tau} \quad \text{for all } K_{\nu-1} < |k| \le K_\nu\}$.

Proof: The proof amounts to the verification of H1)-H9) for all ν. For simplicity, we let $r_0 = \beta_0 = 1$.

First, it is obvious from (3.1) that H3) holds for $\nu = 0$. By choosing μ_0 small, we also see that H2), H4)-H9) hold for $\nu = 0$ and H6) holds for all ν.

By the definition of μ_ν, we have that

$$\mu_\nu = (8^m c_0)^{(1+\sigma)^\nu - 1} \mu_0^{(1+\sigma)^\nu}. \tag{4.9}$$

Let $\zeta \gg 1$ be fixed and μ_0 be sufficiently small such that

$$\mu_0 < \left(\frac{1}{8^m c_0 \zeta}\right)^\sigma < 1. \tag{4.10}$$

Then

$$\begin{aligned}
\mu_1 &= 8^m c_0 \mu_0^{1+\sigma} < \frac{1}{\zeta}\mu_0 < 1, \\
\mu_2 &= 8^m c_0 \mu_1^{1+\sigma} < \frac{1}{\zeta}\mu_1 < \frac{1}{\zeta^2}\mu_0, \\
&\cdots\cdots \\
\mu_\nu &= 8^m c_0 \mu_{\nu-1}^{1+\sigma} < \cdots < \frac{1}{\zeta^\nu}\mu_0.
\end{aligned} \tag{4.11}$$

Denote

$$\Gamma_\nu = \Gamma(r_\nu - r_{\nu+1}).$$

We note that

$$r_\nu - r_{\nu+1} = \frac{1}{2^{\nu+2}} = \frac{\beta_\nu - \beta_{\nu+1}}{\beta_0}. \tag{4.12}$$

Since

$$\begin{aligned}
\Gamma_\nu &\le \int_1^\infty \lambda^{(d+m+6)\tau+d+m+6} e^{-\frac{\lambda}{2^{\nu+6}}} \mathrm{d}\lambda \\
&\le ([(d+m+6)\tau] + d + m + 7)! 2^{(\nu+6)((d+m+6)\tau+d+m+6)},
\end{aligned}$$

it is clear that if ζ is sufficiently large, then

$$\mu_\nu^\sigma \Gamma_\nu^i < \mu_\nu^\sigma(\Gamma_\nu^i + 1) < \frac{\gamma_{\nu+1}^{d+m+5}}{\gamma_\nu^{d+m+5}}, \quad i = 1, 2. \tag{4.13}$$

In particular, H9) holds for all $\nu \geq 1$, and,

$$\mu_\nu \Gamma_\nu \leq \mu_\nu^{1-\sigma} \leq \frac{\mu_0^{1-\sigma}}{\zeta^{(1-\sigma)\nu}}. \tag{4.14}$$

By (4.12) and (4.14), it is easy to see that if ζ is sufficiently large and μ_0 is sufficiently small, then H4), H5), H7), H8) hold for all $\nu \geq 1$.

Since

$$\int_{K_{\nu+1}}^{\infty} t^{d+m} e^{-\frac{t}{2^{\nu+3}}} \mathrm{d}t \leq (d+m+1)! 2^{(\nu+6)(d+m)} K_{\nu+1}^n e^{-\frac{K_{\nu+1}}{2^{\nu+2}}},$$

it follows from (4.9) and the inequality $(1+\sigma)^\eta > 2$ that H1) holds for all $\nu \geq 0$ as μ_0 small.

For the verification of H3), we observe by (4.11) that

$$\frac{1}{4}(M^* + 1)\mu_{\nu-1}^{2\sigma} K_{\nu+1}^{\tau+1} < \frac{1}{2^{\nu+2}},$$

as μ_0 small. Then

$$2(M^*+1)s_\nu K_{\nu+1}^{\tau+1} \leq \frac{s_{\nu-1}}{4}(M^*+1)\mu_{\nu-1}^{2\sigma} K_{\nu+1}^{\tau+1} \leq \frac{s_0}{2^{\nu+2}} < \frac{\gamma_0}{2^{\nu+2}} < \gamma_\nu - \gamma_{\nu+1},$$

which verifies H3) for all $\nu \geq 1$.

Let $\zeta^{1-\sigma} \geq 2$ in (4.10), (4.11). We have by (3.1)-(3.3) that if μ_0 is sufficiently small, then the following holds for all $\nu \geq 1$:

$$c_0 \mu_\nu \leq \frac{\mu_0}{2^\nu} \leq \frac{\mu^{\frac{1}{2}}}{2^\nu}, \tag{4.15}$$

$$c_0 \mu_\nu \Gamma_\nu \leq \frac{\mu_0^{1-\sigma}}{2^\nu} \leq \frac{\mu^{\frac{1}{2}}}{2^\nu}, \tag{4.16}$$

$$c_0 s_\nu^{m-1} \mu_\nu \leq \frac{\mu_0^{1+2\sigma(m-1)} s_0^{m-1}}{2^{\nu+3}} \leq \frac{(\mu_0 s_0^m)}{2^\nu} \frac{\mu_0^{2\sigma(m-1)}}{8 s_0} \leq \frac{\mu}{2^\nu}. \tag{4.17}$$

The verification of H2) follows from (4.15) and an inductive application of (3.28) for all $\nu = 0, 1, \cdots$.

Above all, the KAM steps described in Section 3 are valid for all ν, which gives the desired sequences stated in the lemma.

Now, 1) follows from Lemma 3.7, 2) follows from (3.20) and induction, (4.2), (4.4), (4.6) follow from (4.15), (4.17) and Lemma 3.3, and (4.8) follows from Lemma 3.8 and H9). By adding up (4.2), (4.4), (4.6) for all $\nu = 1, 2, \cdots$, we also obtain (4.3), (4.5), (4.7) respectively.

4) clearly holds for $\nu = 0$. We now assume that $\nu > 0$. Then by Lemma 3.6,

$$\Lambda_\nu = \{\lambda \in \lambda_\nu : |\langle k, \omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau}, \ 0 < |k| \leq K_\nu\}.$$

Hence

$$\begin{aligned}\Lambda_{\nu+1} &= \{\lambda\in\Lambda_\nu : |\langle k,\omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ 0<|k|\le K_{\nu+1}\}\\ &= \{\lambda\in\Lambda_\nu : |\langle k,\omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ 0<|k|\le K_{\nu}\}\\ &\qquad \cap\{y_0\in\Lambda_\nu : |\langle k,\omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ K_\nu<|k|\le K_{\nu+1}\}\\ &= \Lambda_\nu\cap\{y_0\in\Lambda_\nu : |\langle k,\omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ K_\nu<|k|\le K_{\nu+1}\}\\ &= \{y_0\in\Lambda_\nu : |\langle k,\omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ K_\nu<|k|\le K_{\nu+1}\}.\end{aligned}$$

The lemma is now complete. ■

4.2 Convergence

Let

$$\Psi^\nu = \Phi_1\circ\Phi_2\circ\cdots\circ\Phi_\nu,\quad \nu=1,2,\cdots.$$

Then $\Psi^\nu : \tilde{D}_\nu\times\Lambda_0\to\tilde{D}_0$, and,

$$\begin{aligned}&H_0\circ\Psi^\nu = H_\nu = N_\nu + P_\nu,\\ &N_\nu = e_\nu + \langle\omega_\nu(\lambda), y\rangle + h_\nu(y,\lambda),\end{aligned}$$

$\nu=0,1,\cdots$, where $\Psi_0 = id$. Using (4.1) and the identity

$$\Psi^\nu = id + \sum_{i=1}^{\nu}(\Psi_i - \Psi_{i-1}),$$

it is easy to see that Ψ^ν converges in $C^{m+1,d-1+\sigma_0}(D(\frac{r_0}{2},\frac{\beta_0}{2})\times\Lambda_0)$ norm to a function $\Psi^\infty\in C^{m,d-1}(D(\frac{r_0}{2},\frac{\beta_0}{2})\times\Lambda_0)$ such that $\Psi_\lambda = \Psi^\infty(\cdot,\lambda)$, $\lambda\in\Lambda_0$, are symplectic and C^m uniformly close to the identity. Let

$$\Lambda_* = \bigcap_{\nu\ge 0}\Lambda_\nu.$$

Then $\{\Psi_\lambda : \lambda\in\Lambda_*\}$ is a C^{d-1} Whitney smooth family of analytic symplectic transformations on $D(\frac{r_0}{2},\frac{s_0}{2})$. By Lemma 4.1, it is also clear that e_ν,ω_ν converge uniformly on Λ_* and h_ν converge uniformly on $D(\frac{s_0}{2})\times\Lambda_*$. Denote $e_\infty,\omega_\infty,h_\infty$ as the limit of e_ν,ω_ν,h_ν respectively. Then, on $D(\frac{s_0}{2})\times\Lambda_*$, N_ν converge uniformly to

$$N_\infty = e_\infty + \langle\omega_\infty(\lambda), y\rangle + h_\infty(y,\lambda).$$

Hence, on $D(\frac{r_0}{2},\frac{s_0}{2})\times\Lambda_*$,

$$P_\nu = H_0\circ\Psi^\nu - N_\nu,$$

converge uniformly to

$$P_\infty = H_0\circ\Psi^\infty - N_\infty.$$

Since P_ν is real analytic on D_ν and

$$|P_\nu|_{D_\nu} \le \gamma_\nu^{d+m+5} s_\nu^m \mu_\nu,$$

the Cauchy estimate yields that

$$|\partial_y^j P_\nu|_{D(r_{\nu+m},\frac{1}{2}s_\nu)} \le \left(\frac{m}{r_0}\right)^m 2^{m\nu+\frac{m}{2}+2}\gamma_\nu^{d+m+5}\mu_\nu, \quad |j| \le m.$$

By (4.9), the right hand side of the above converges to 0 as $\nu \to \infty$, provided that μ (hence μ_0) is sufficiently small. Thus, on $D(\frac{r_0}{2}, 0) \times \Lambda_*$,

$$\partial_y^j P_\infty = 0, \quad |j| \le m.$$

Hence for each $\lambda \in \Lambda_*$, $T^d \times \{0\}$ is an analytic invariant torus of H_∞ with the toral frequency $\omega_\infty(\lambda)$, which, by definition of Λ_ν and Lemma 4.1 2), satisfies

$$\begin{aligned} |\langle k, \omega_\infty(\lambda)\rangle| &> \frac{\gamma}{2|k|^\tau}, \quad \text{for all } k \in Z^d \setminus \{0\}, \\ (\omega_\infty(\lambda))_q &\equiv (\omega_0(\lambda))_q, \quad \text{for all } 1 \le q \le n. \end{aligned}$$

Following the Whitney extension of Ψ^ν's, all $e_\nu, \omega_\nu, h_\nu, P_\nu$, $\nu = 0, 1, \cdots$, admit uniform $C^{d-1+\sigma_0}$ extensions in $\lambda \in \Lambda_0$ with derivatives in λ up to order $d-1$ satisfying the same estimates (4.2)-(4.8). Thus, $e_\infty, \omega_\infty, h_\infty, P_\infty$ are C^{d-1} Whitney smooth in $\lambda \in \Lambda_*$, and, the derivatives of $(e_\infty - e_0), (\omega_\infty - \omega_0), (h_\infty - h_0)$ satisfy similar estimates as in (4.3),(4.5),(4.7). Consequently, the perturbed tori form a C^{d-1} Whitney smooth family on Λ_*.

4.3 Measure estimate

Lemma 4.2 *Let $\Lambda \subset R^d$, $d > 1$, be a bounded closed region and let $g : \Lambda \to R^d$ be such that*

$$\text{rank}\{\frac{\partial^\alpha g}{\partial \lambda^\alpha} : |\alpha| \le d-1\} = d.$$

Then for a fixed $\tau > d(d-1) - 1$

$$\left|\{\lambda \in \Lambda : |\langle g(\lambda), k\rangle| \le \frac{\gamma}{|k|^\tau}\}\right| \le c(\Lambda, d, \tau)\left(\frac{\gamma}{|k|^{\tau+1}}\right)^{\frac{1}{d-1}}, \quad k \in Z^d \setminus \{0\}, \ \gamma > 0.$$

Proof: See Theorem B in [27]. We note that the constant c above does not depend on g but rather on a lower bound of the derivatives of g up to order $d-1$. ■

The following measure estimate is adopted from [19]. We consider the following three cases.

Case 1: $d_0 = d$. Let

$$\begin{aligned} R_k^{\nu+1} &= \{\lambda \in \Lambda_\nu : |\langle k, \omega_\nu(\lambda)\rangle| \le \frac{\gamma_\nu}{|k|^\tau}\}, \quad k \in R^d \setminus \{0\}, \\ \hat{R}_k^{\nu+1} &= \{\lambda \in \Lambda_0 : |\langle k, \omega_\nu(\lambda)\rangle| \le \frac{\gamma_\nu}{|k|^\tau}\}, \quad k \in R^d \setminus \{0\}, \end{aligned}$$

for all $\nu = 0, 1, \cdots$. Then by Lemma 4.1 4),

$$\Lambda_{\nu+1} = \Lambda_\nu \setminus \bigcup_{K_\nu < |k| \le K_{\nu+1}} R_k^{\nu+1},$$

and,

$$\Lambda_0 \setminus \Lambda_* = \bigcup_{\nu=0}^{\infty} \bigcup_{K_\nu < |k| \le K_{\nu+1}} R_k^{\nu+1}.$$

Since (4.5) is also satisfied by the extended toral frequencies ω_ν on Λ_0, **A1)** implies that if μ is sufficiently small, then

$$\mathrm{rank}\{\frac{\partial^\alpha \omega_\nu}{\partial \lambda^\alpha} : |\alpha| \le d-1\} = d$$

for all $\lambda \in \Lambda_0$, $\nu = 0, 1, \cdots$. It follows from Lemma 4.2 that

$$|R_k^{\nu+1}| \le |\hat{R}_k^{\nu+1}| \le c \left(\frac{\gamma}{|k|^{\tau+1}} \right)^{\frac{1}{d-1}},$$

for all $k \in Z^d \setminus \{0\}$ and $\nu = 0, 1, \cdots$, where c is a constant independent of ν. Hence

$$\begin{aligned} |\Lambda_0 \setminus \Lambda_*| &\le \sum_{\nu=0}^{\infty} \sum_{K_\nu < |k| \le K_{\nu+1}} |R_k^{\nu+1}| \le c\gamma^{\frac{1}{d-1}} \sum_{\nu=0}^{\infty} \sum_{K_\nu < |k| \le K_{\nu+1}} \frac{1}{|k|^{\frac{\tau+1}{d-1}}} \\ &= O(\gamma^{\frac{1}{d-1}}) = O(\gamma^{\frac{1}{d_*-1}}), \end{aligned}$$

as desired.

Case 2: $d_0 < d$. Let $\bar{\Lambda} = [1, 2]^{d-d_0}$ and define

$$\begin{aligned} &\tilde{\Lambda} = \Lambda_0 \times \bar{\Lambda}, \\ &\tilde{\Lambda}_* = \Lambda_* \times \bar{\Lambda}, \\ &\tilde{\lambda} = (\lambda, \bar{\lambda})^\top, \quad \bar{\lambda} \in \bar{\Lambda}, \\ &\tilde{\omega}_\nu(\tilde{\lambda}) = \omega_\nu(\lambda), \quad \nu = 0, 1, \cdots, \ \tilde{\lambda} \in \tilde{\Lambda}. \end{aligned}$$

Then it is clear that

$$\mathrm{rank}\{\frac{\partial^\alpha \tilde{\omega}_\nu}{\partial \tilde{\lambda}^\alpha} : |\alpha| \le d-1\} = d$$

on $\tilde{\Lambda}$ for all $\nu = 0, 1, \cdots$, as μ sufficiently small. Similar to Case 1, we have that

$$|\tilde{\Lambda} \setminus \tilde{\Lambda}_*| = O(\gamma^{\frac{1}{d-1}}).$$

By Fubini's theorem,

$$|\Lambda_0 \setminus \Lambda_*| = O(\gamma^{\frac{1}{d-1}}) = O(\gamma^{\frac{1}{d_*-1}}),$$

as desired.

Case 3: $d_0 > d$. For any $\lambda \in \Lambda_0$, **A1)** implies that there exist indexes

$$\alpha^i \in \{\alpha \in Z_+^{d_0} : |\alpha| \le d-1\}, \quad i = 0, 1, \cdots, d-1,$$

such that

$$\mathrm{rank}\{\frac{\partial^{\alpha^i}\omega}{\partial\lambda^{\alpha^i}}(\lambda) : i = 0, 1, \cdots, d-1\} = d.$$

Since $\mathrm{rank}\{\frac{\partial\omega}{\partial\lambda}(\lambda)\} \le d$, there are $\lambda_{i_1}, \lambda_{i_2}, \cdots, \lambda_{i_{d_0-d}}$ such that

$$\frac{\partial\omega}{\partial\lambda_{i_j}}(\lambda) \notin \{\frac{\partial^{\alpha^i}\omega}{\partial\lambda^{\alpha^i}}(\lambda) : i = 0, 1, \cdots, d-1\}, \quad j = 1, 2, \cdots, d_0 - d.$$

Define

$$\begin{aligned}
&\Omega(\lambda) = (\lambda_{i_1}, \lambda_{i_2}, \cdots, \lambda_{i_{d_0-d}})^\top, \quad \lambda \in \Lambda_0,\\
&\tilde\omega_\nu(\lambda) = (\omega_\nu(\lambda), \Omega(\lambda))^\top, \quad \nu = 0, 1, \cdots, \ \lambda \in \Lambda_0,\\
&\tilde R_k^{\nu+1} = \{\lambda \in \Lambda_\nu : |\langle k, \tilde\omega_\nu(\lambda)\rangle| \le \frac{\gamma_\nu}{|k|^\tau}\}, \quad k \in Z^{d_0} \setminus \{0\}, \ \nu = 0, 1, \cdots,\\
&\tilde\Lambda_{\nu+1} = \tilde\Lambda_\nu \setminus \bigcup_{K_\nu < |k| \le K_{\nu+1}} \tilde R_k^{\nu+1}, \quad \nu = 0, 1, \cdots,\\
&\tilde\Lambda_* = \bigcap_{\nu\ge 0} \tilde\Lambda_\nu.
\end{aligned}$$

Then

$$\mathrm{rank}\{\frac{\partial^{\alpha^i}\tilde\omega_\nu}{\partial\lambda^{\alpha^i}}(\lambda) : i = 0, 1, \cdots, d-1; \frac{\partial\tilde\omega_\nu}{\partial\lambda_{i_j}}(\lambda) : j = 1, \cdots, d_0 - d\} = d_0$$

on Λ_0 for all $\nu = 0, 1, \cdots$. It follows that

$$\mathrm{rank}\{\frac{\partial^\alpha\tilde\omega_\nu}{\partial\lambda^\alpha} : \forall|\alpha| \le d_0 - 1\} = d_0$$

on Λ_0 for all $\nu = 0, 1, \cdots$. Similar to Case 1), we have that

$$|\Lambda \setminus \tilde\Lambda_*| = O(\gamma^{\frac{1}{d_0-1}}) = O(\gamma^{\frac{1}{d_*-1}}).$$

Since $\tilde\Lambda_* \subset \Lambda_*$,

$$|\Lambda_0 \setminus \Lambda_*| \le |\Lambda_0 \setminus \tilde\Lambda_*| = O(\gamma^{\frac{1}{d_*-1}}),$$

as desired. This proves part 1) of Theorem A.

Given the convergence in Section 4.2, part 2) of Theorem A clearly follows from Lemma 4.1 2).

4.4 Proof of Corollary

Without loss of generality, we assume that M admits a global coordinate, i.e., there is a bounded closed region $\Lambda \in R^{d_0}$ and a C^{l_0} diffeomorphism $y : \Lambda \to M$ such that $M = y(\Lambda)$. Let $\lambda \in \Lambda$ and consider the transformation

$$y \mapsto y + y(\lambda).$$

Then (1.1) gives rise to

$$H(x, y, \lambda, \varepsilon) = e(\lambda) + \langle\omega(\lambda), y\rangle + h(y, \lambda) + P(x, y, \lambda, \varepsilon),$$

where

$$\begin{aligned}
e(\lambda) &= N(y(\lambda)), \\
\omega(\lambda) &= \frac{\partial N}{\partial y}(y(\lambda)), \\
h(y,\lambda) &= \frac{1}{2}\langle y, A(\lambda)y\rangle + \hat{h}(y,\lambda), \\
A(\lambda) &= \frac{\partial^2 N}{\partial y^2}(y(\lambda)), \\
\hat{h}(y,\lambda) &= O(|y|^3), \\
P(x,y,\lambda,\varepsilon) &= \varepsilon P(x, y+y(\lambda), \varepsilon).
\end{aligned}$$

Let r be fixed and take

$$s = \varepsilon^{\frac{1}{2m+l_0+5}}, \quad \gamma = \varepsilon^{\frac{1}{2(2m+l_0+5)}}, \quad \mu = \varepsilon^{\frac{2}{2m+l_0+5}}.$$

Then (2.2) holds and the Corollary follows immediately from the theorem as ε sufficiently small.

4.5 Proof of Theorem B

By choosing λ, Λ as in the Section 4.4 above with the present M, the proof of Theorem B essentially follows from that of Theorem A, except that the translation

$$\phi: \; x \to x, \quad y \to y + y^*$$

in Section 3.3 should be defined for the purpose of eliminating the energy drift at each KAM step.

In the case of part 1) of Theorem B, y_* is defined so that $e_+ = e = E$. Hence, instead of (3.17), we consider the equation

$$\langle \omega, y^* \rangle + \frac{1}{2}\langle y^*, Ay^* \rangle + \hat{h}(y^*) + [R](y^*) = 0,$$

which, by the implicit function theorem, clearly admits a local smooth solution y^* on M.

In the case of part 2) of Theorem B, y_* is defined so that $e_+ = e = E$, and,

$$[\omega_{+,i_1} : \cdots : \omega_{+,i_n}] = [\omega_{i_1} : \cdots : \omega_{i_n}].$$

Hence, instead of (3.17), we consider the equations

$$\Big(\mathcal{A} + \frac{\partial \hat{h}}{\partial(y_{i_1}, \cdots, y_{i_n})}(y^*)\Big)(y^*_{i_1}, \cdots, y^*_{i_n})^\top - t^*(\omega_{i_1} : \cdots : \omega_{i_n})^\top = -(p_{01,i_1}, \cdots, p_{01,i_n})^\top,$$
$$\langle (\omega_{i_1}, \cdots, \omega_{i_n})^\top, (y^*_{i_1}, \cdots, y^*_{i_n})^\top \rangle + \frac{1}{2}\langle y^*, Ay^* \rangle + \hat{h}(y^*) + [R](y^*) = 0,$$

which, by the sub-isoenergetic nondegenerate condition **A1)**$''$ and the implicit function theorem, admits a local smooth solution (y^*, t^*), $y^* \in M$, $t^* \in R^1$, such that $y^*_j = 0$ if $j \notin \{i_1, \cdots, i_n\}$.

Let ϕ_F^1 be as in Section 3 and ϕ be as in the above. Then under the symplectic transformation

$$\Phi_+ = \phi_F^1 \circ \phi,$$

the new Hamiltonian reads

$$\begin{aligned} H \circ \Phi_+ &= N_+ + P_+, \\ N_+ &= \bar{N}_+ \circ \phi - \psi = E + \langle \omega_+, y \rangle + h_+(y), \\ P_+ &= \bar{P}_+ \circ \phi + \psi, \end{aligned}$$

where, with respect to y^* defined above,

$$\omega_+ = \omega + p_{01} + Ay^* + \partial_y \hat{h}(y^*),$$

and, $h_+(y), A_+, \hat{h}_+(y), \psi$ have the same forms as in (3.21)-(3.24). Thus, with estimates on the present y^* similar to those in Sections 3.4, 3.5, the remaining proof of Theorem A is valid.

Part 3) of Theorem B is a special case of part 3) of the Corollary.

References

[1] V.I. Arnold, Proof of A. N. Kolmogorov's theorem on the preservation of quasi periodic motions under small perturbations of the Hamiltonian, *Usp. Math. USSR* **18** (1963), 13-40

[2] V.I. Arnold, Mathematical methods of classical mechanics, Springer-Verlag, New York, 1980

[3] V.I. Arnold and A. Avez, Problm̀es ergodiques de la mécanique classique, Gauthier-Villars, Paris, 1967

[4] G. Benettin, L. Galgani, A Giorgilli and J. Strelcyn, A proof of Kolmogorov's theorem on invariant tori using canonical transformations defined by the Lie method, *Nuovo. Cimento B.* **79** (1984), 201-223

[5] J. Bricmont, K. Gawedzki and A. Kupiainen, KAM theorem and quantum field theory, *Comm. Math. Phys.* **201** (1999), 699-727

[6] H. Broer and G. Huitema, A proof of the isoenergetic KAM-theorem from the "ordinary" one, *J. Diff. Eqs.* **90** (1991), 52-60

[7] H. Broer, G. Huitema and M. Sevryuk, Quasi-periodic motions in families of dynamical systems, *Lecture Notes in Math.,* **1645**, Springer-Verlag, 1996

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

[9] C-Q. Cheng and Y-S. Sun, Existence of KAM tori in degenerate Hamiltonian systems, *J. Diff. Eqs.* **114** (1994), 288-335

[10] L. Chierchia and C. Falcolini, A direct proof of a theorem by Kolmogorov in Hamiltonian systems, *Ann. Sc. Super. Pisa. CI Sci. Serie 4* **21** (1994) 541-593

[11] A. Delshams and P. Gutierrez, Estimates on invariant tori near an elliptic equilibrium point of a Hamiltonian systems, *J. Diff. Eqs.* **131** (1996), 277-303

[12] L. H. Eliasson, Absolutely convergent series expansions for quasi-periodic motions, *Math. Phy. Elect. J.* **2** (1996), 1-33

[13] G. Gallavotti, Twistless KAM tori, *Comm. Math Phy.* **164** (1994), 145-156

[14] G. Gallavotti, G. Gentile, and V. Mastropietro, Field theory and KAM tori, *Math. Phy. Elect. J.* **1** (1995), 1-13

[15] A. Giorgilli and L. Galgani, Formal integrals for an autonomous Hamiltonian system near an equilibrium point, *Celest. Mech.* **17** (1978), 267-280

[16] A. Giorgilli and L. Galgani, Rigorous estimates for the series expansions of Hamiltonian perturbation theory, *Celest. Mech.* **37** (1985), 95-112

[17] A. N. Kolmogorov, On quasi-periodic motions under small perturbations of the Hamiltonian, *Doklady Akad. Nauk. USSR* **98** (1954), 527-530

[18] Y. Li and Y. Yi, A quasi-periodic Poincaré's theorem, to appear in *Math. Ann.*

[19] Y. Li and Y. Yi, Persistence of invariant tori in generalized Hamiltonian systems, to appear in *Erg. Th. Dyn. Sys.*

[20] J. Moser, On invariant curves of area preserving mappings of an annulus, *Nachr. Akad. Wiss. Gött. Math. Phys.* **K1** (1962), 1-20

[21] A. I. Neishtadt, Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions, *J. Appl. Math. Mech.* **45** (1982), 766-772

[22] J. Pöschel, Integrability of Hamiltonian systems on Cantor sets, *Comm. Pure Appl. Math.* **35** (1982), 653-696

[23] J. Pöschel, On the elliptic lower dimensional tori in Hamiltonian systems, *Math. Z.* **202** (1989), 559-608

[24] H. Rüssmann, On twist Hamiltonians, Talk on the Colloque International: Mécanique céleste et syst. hamiltoniens (1990)

[25] M. B. Sevryuk, KAM-stable Hamiltonians, *J. Dyn. Control Syst.* **1**(1995), 351-366

[26] E. M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, New Jersey, 1970

[27] J. X. Xu, J. G. You and Q. J. Qiu, Invariant tori for nearly integrable Hamiltonian systems with degeneracy, *Math. Z.* **226** (1997), 375-387