PERSISTENCE OF HYPERBOLIC TORI IN HAMILTONIAN SYSTEMS

YONG LI AND YINGFEI YI

Abstract. We generalize the well-known result of Graff and Zehnder on the persistence of hyperbolic invariant tori in Hamiltonian systems by considering non-Floquet, frequency varying normal forms and allowing the degeneracy of the unperturbed frequencies. The preservation of part or full frequency components associated to the degree of non-degeneracy is considered. As applications, we consider the persistence problem of hyperbolic tori on a submanifold of a nearly integrable Hamiltonian system and the persistence problem of a fixed invariant hyperbolic torus in a non-integrable Hamiltonian system.

Dedicated to Professor George R. Sell on the occasion of his 65th birthday

1. Introduction

In [14], Moser considered the following Hamiltonian system:

$$H = e + \langle \omega_0, y\rangle + \frac{1}{2}\langle y, Ay\rangle + \frac{1}{2}\langle z, Mz\rangle + P(x, y, z), \tag{1.1}$$

where $(x, y, z) \in T^n \times R^n \times R^{2m}$, $\omega_0 \in R^n$ is a fixed Diophantine toral frequency, A, M are $n \times n$, $2m \times 2m$ non-singular, constant matrices, respectively, JM is hyperbolic with all eigenvalues being real and distinct, and P is a small perturbation. The persistence of the unperturbed Diophantine hyperbolic torus $T^n \times \{0\} \times \{0\}$ was shown along with the preservation of the toral frequency ω_0. By considering a symplectic reduction of M into the form

$$M = \begin{pmatrix} O & B_0 \\ B_0^\top & O \end{pmatrix}, \tag{1.2}$$

Graff [9] generalized Moser's result by allowing multiple eigenvalues of M. An alternative proof of Graff's result was later given by Zehnder in [25] using implicit function technique. This result has played a fundamental role in analyzing global branches of invariant tori and Arnold diffusion in Hamiltonian systems ([10, 23]). For the Lindstedt series approach to the persistence of hyperbolic tori in Hamiltonian systems, we refer the readers to [8, 11].

The present paper concerns a generalization of the Graff-Zehnder result to the non-Floquet, frequency varying cases in which certain degeneracy of the unperturbed frequencies is allowed. More precisely, we consider the following Hamiltonian

---

1991 *Mathematics Subject Classification.* 37J40.

Key words and phrases. Hamiltonian system, Hyperbolicity, Invariant tori, KAM theory.

The first author is partially supported by NSFC 19971042, National 973 Key Project of China: Nonlinearity, the outstanding young's project of Ministry of Education of China, and National outstanding young's award of China. The second author was partially supported by NSF grant DMS0204119.

systems

$$H = e(\lambda) + \langle \omega(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}(x,\lambda)\begin{pmatrix} y \\ z \end{pmatrix}\rangle + h(x,y,z,\lambda) + P(x,y,z,\lambda), \tag{1.3}$$

where $(x,y,z) \in T^n \times R^n \times R^{2m}$, λ is a parameter in a bounded, closed, connected region $\mathcal{O} \subset R^k$, e and ω are of class C^{l_0} on $\mathcal{O}$ for some fixed $l_0 \geq \max\{n, 2\}$, $\mathcal{M}$ is symmetric, real analytic in $x \in \mathcal{D}(r) = \{x \in C^n/Z^n : \ |\mathrm{Im}x| < r\}$ and C^{l_0} in $\lambda \in \mathcal{O}$, $h(x,y,z,\lambda) = O(|(y,z)|^3)$ is real analytic, and, P, viewed as a perturbation, is real analytic in a complex neighborhood $D(r,s) = \{(x,y,z) : \ |\mathrm{Im}x| < r, |y| < s, |z| < s\}$ of $T^n \times \{0\} \times \{0\}$ and C^{l_0} in $\lambda \in \mathcal{O}$.

Throughout the paper, we shall use the same symbol $|\cdot|$ to denote the sup-norm of vectors and its induced matrix norm, the standard l_1 norm in a lattice Z^p, absolute value of functions, and Lebesgue measure of sets etc. Thus, for any matrix $Q = (q_{ij})$, $|Q| = \max_i \sum_j |q_{ij}|$. Also, $[\cdot]$ will denote both the average of a matrix valued function on a torus and the integral part of a real number. For any (vector, matrix valued) function f defined on a domain D, $|f|_D$ stands for $\sup_D |f|$, and, for any two complex column vectors ξ, ζ of the same dimension, $\langle \xi, \zeta\rangle$ means the transpose of ξ times ζ.

Write $\mathcal{M}$ in (1.3) into blocks:

$$\mathcal{M} = \begin{pmatrix} A & B \\ B^\top & M \end{pmatrix}, \tag{1.4}$$

where $A = A(x,\lambda)$, $B = B(x,\lambda)$, $M = M(x,\lambda)$ are $n \times n$, $n \times 2m$, $2m \times 2m$ blocks of $\mathcal{M} = \mathcal{M}(x,\lambda)$ respectively. With respect to the standard symplectic form

$$\sum_{i=1}^{n} \mathrm{d}x_i \wedge \mathrm{d}y_i + \sum_{j=1}^{m} \mathrm{d}z_j \wedge \mathrm{d}z_{j+m},$$

the equation of motion associated to (1.3) reads

$$\begin{cases} \dot{x} & = \omega(\lambda) + Ay + Bz + \partial_y h + \partial_y P, \\ \dot{y} & = -\frac{1}{2}\partial_x \langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}(x,\lambda)\begin{pmatrix} y \\ z \end{pmatrix}\rangle - \partial_x h - \partial_x P, \\ \dot{z} & = JMz + JB^\top y + J\partial_z h + J\partial_z P, \end{cases}$$

where J denotes the $2m \times 2m$ symplectic matrix. Thus, the unperturbed system associated to (1.3) admits a smooth family of invariant n-tori $T_\lambda = T^n \times \{0\} \times \{0\}$ with toral frequencies $\omega(\lambda)$ parameterized by $\lambda \in \mathcal{O}$. We first assume that $J[M]$ is hyperbolic on $\mathcal{O}$, i.e., if $\lambda_i(\lambda)$, $i = 1, 2, \cdots, 2m$, are eigenvalues of $J[M](\lambda)$, then

H) *there is a $\sigma_0 > 0$ such that*

$$|\mathrm{Re}\lambda_i(\lambda)| \geq \sigma_0,$$

for all $\lambda \in \mathcal{O}$ and $i = 1, 2, \cdots, 2m$.

We note that if both $|M - [M]|_{\mathcal{D}(r)\times\mathcal{O}}$ and $|B|_{\mathcal{D}(r)\times\mathcal{O}}$ are sufficiently small, then H) implies that the invariant tori T_λ, $\lambda \in \mathcal{O}$, are hyperbolic in the z-direction.

Next, we assume the Rüssmann condition on the frequency map, i.e.,

R)

$$\max_{\lambda\in\mathcal{O}} \mathrm{rank}\{\partial^\alpha \omega(\lambda) : \ \forall |\alpha| \leq n-1\} = n.$$

The Rüssmann condition is known to be the weakest non-degenerate condition for the persistence of maximal dimensional invariant tori in nearly integrable Hamiltonian systems ([3, 19, 20, 22, 24]).

Define

$$\eta_0 = \frac{2}{\sqrt{\rho_0^2 + 4\alpha_0\rho_0} + \rho_0},$$

where,

$$\begin{aligned} \alpha_0 &= (1+2m)|[M]^{-1}|_{\mathcal{O}}, \\ \rho_0 &= \frac{4m}{\sigma_0}(1 + \frac{2m}{\sigma_0}|[M]|_{\mathcal{O}})^{2m-1}. \end{aligned} \tag{1.5}$$

Our main result is stated as follows.

Theorem 1. *Consider* (1.3)*. Assume conditions* H)*,* R) *and that*

$$|M - [M]|_{\mathcal{D}(r)\times\mathcal{O}},\ |B - [B]|_{\mathcal{D}(r)\times\mathcal{O}} < \eta_0. \tag{1.6}$$

Then there is a $\mu = \mu(r, s, l_0, \sigma_0) > 0$ sufficiently small such that if

$$|\partial_\lambda^l P|_{D(r,s)\times\mathcal{O}} < s^2\gamma^{3l_0+4}\mu,\ |l| \le l_0, \tag{1.7}$$

then there is a $0 < r_0 = r_0(r, \sigma_0) \le r$ and a Cantor-like set $\mathcal{O}_\gamma \subset \mathcal{O}$, with $|\mathcal{O}\setminus\mathcal{O}_\gamma| = O(\gamma^{\frac{1}{n_*-1}})$, where $n_* = \max\{2, n\}$, for which the following holds. There is a C^{l_0-1} Whitney smooth family of real analytic, symplectic transformations

$$\Psi_\lambda : D(r_0, \frac{s}{2}) \to D(r_0, s), \quad \lambda \in \mathcal{O}_\gamma,$$

which are C^{l_0} uniformly close to the identity such that

$$H \circ \Psi_\lambda = e_* + \langle \Omega_*(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}_*(x,\lambda)\begin{pmatrix} y \\ z \end{pmatrix}\rangle + h(x,y,z,\lambda) + P_*(x,y,z,\lambda),$$

where

$$\begin{aligned} &|\partial_\lambda^l e_* - \partial_\lambda^l e|_{\mathcal{O}_\gamma} = O(\gamma^{2(l_0+1)}\mu^{\frac{1}{2}}), \\ &|\partial_\lambda^l \Omega_* - \partial_\lambda^l \omega|_{\mathcal{O}_\gamma} = O(\gamma^{2(l_0+1)}\mu^{\frac{1}{2}}), \\ &|\partial_\lambda^l \mathcal{M}_* - \partial_\lambda^l \mathcal{M}|_{\mathcal{D}(r_0)\times\mathcal{O}_\gamma} = O(\gamma^{l_0+1}\mu^{\frac{1}{4}}), \end{aligned}$$

for all $|l| \le l_0$. Moreover,

$$\partial_y^p\partial_z^q P_*|_{(y,z)=(0,0)} \equiv 0,\ |p| + |q| \le 2.$$

Thus, all unperturbed tori T_λ with $\lambda \in \mathcal{O}_\gamma$ will persist and give rise to a C^{l_0-1} Whitney smooth family of slightly deformed, analytic, quasi-periodic, invariant n-tori of the perturbed system with the Diophantine toral frequencies $\Omega_*(\lambda)$.

Our next result concerns the preservation of toral frequencies in connection with the degree of non-degeneracy of the matrix $[A]$. More precisely, we assume that

- **ND)** *there is an $1 \le n_0 \le n$ such that both the $n_0 \times n_0$ ordered principal block U of $[A]$ and $Y \equiv [M] - [B]^\top \mathrm{diag}(U^{-1}, O)[B]$ are non-singular on $\mathcal{O}$, where O denotes the zero matrix.*

It is clear that ND) holds automatically if $[A]$ is non-singular on $\mathcal{O}$ and $|[B]|_{\mathcal{O}}$ is sufficiently small (in particular when $[B] \equiv 0$).

Define

$$\eta = \frac{2}{\sqrt{\rho_0^2 + 4\alpha\rho_0} + \rho_0}, \tag{1.8}$$

where ρ_0 is as in (1.5) and

$$\alpha = (1+2m)(|Y^{-1}| + |U^{-1}| + (|Y^{-1}||U^{-1}|)(1 + |[B]| + |[B]||U^{-1}|))_{\mathcal{O}}. \tag{1.9}$$

Theorem 2. *Consider* (1.3)*. Assume conditions* H), ND)*, the condition* (1.6) *with* η *in place of* η_0*, and the condition* (1.7) *with respect to a sufficiently small positive number* $\mu = \mu(r, s, l_0, \sigma_0, U)$*.*

1) *If* R) *holds, then there is a* $r_0 = r_0(r, \sigma_0, U)$ *such that* Theorem 1 *holds with*

$$(\Omega_*(\lambda))_i = \omega_i(\lambda),\ \ \lambda \in \mathcal{O}_\gamma,\ \ i = 1, 2, \cdots, n_0,$$

i.e., the first n_0 *components of a perturbed toral frequency* $\Omega_*(\lambda)$ *coincide with the corresponding ones of the unperturbed toral frequency* $\omega(\lambda)$*.*

2) *If* $n_0 = n$*, i.e.,* $U = [A]$ *is non-singular on* $\mathcal{O}$*, then every hyperbolic Diophantine tori* T_λ *with Diophantine type* (γ, τ) *for a fixed* $\tau > n - 1$ *will persist with unchanged toral frequencies.*

The roughness condition (1.6) in Theorem 1 and the similar roughness condition in Theorem 2 hold automatically if both M and B are small perturbations of x-independent matrices. Since both roughness conditions are independent of the size of the perturbation P and there is no restriction on the smallness of $|A - [A]|$, the above theorems can be applied to Hamiltonians of form (1.3) which may be far from being integrable. In particular, if $M = [M], B = [B]$, then Theorem 1 holds for arbitrary $A(x, \lambda)$.

The above theorems extend the results of Graff and Zehnder even for the case of (1.1), by allowing the degeneracy of A, a general matrix $\mathcal{M}$, and the variation of toral frequencies. It is clear that if $B = 0$ then Theorem 2 2) coincides with the results of Graff and Zehnder. We remark that, in the frequency varying case, a smoothly varying, symplectic reduction of (1.3) to (1.1), in particular with M being reduced to the form (1.2), is generally not available especially when eigenvalues of $J[M](\omega)$ change their multiplicities.

The proof of the above theorems will use a linear iterative scheme which follows the traditional KAM framework but only deals with the elimination of all resonant terms in each KAM step.

The paper is organized as follows. In section 2, we consider applications of the above results to problems such as the persistence of hyperbolic lower dimensional tori on a submanifold in a partially integrable system and the persistence of a hyperbolic invariant torus in a non-integrable system. In Section 3, we describe the linear iterative scheme for one KAM step with respect to (1.3) and also give estimates for the symplectic transformation and the new Hamiltonian. In Section 4, we prove an iteration lemma which checks the validity of all KAM steps. Proofs of both theorems will be completed in Section 5.

We would like to thank the referee for valuable comments and suggestions.

2. Applications

2.1. **Persistence of hyperbolic tori on submanifolds.** A natural way to introduce parameters in the application of Theorems 1 and 2 is to consider persistence of hyperbolic tori on submanifolds in a partially integrable Hamiltonian system. To be more precise, let us consider the Hamiltonian

$$H(x,y,z) = H_0(y,z) + \varepsilon P(x,y,z), \quad x \in T^n, \ y \in R^n, \ z \in R^{2m}, \tag{2.1}$$

endorsed with the standard symplectic structure. We assume that H_0 is partially integrable in the sense that the surface $\{z = 0\}$ is invariant with respect to the Hamiltonian flow associated to H_0. Hence H_0 admits the following Taylor expansion

$$H_0(y,z) = H_0(y,0) + \frac{1}{2}\langle z, \tilde{M}(y)z\rangle + O(|z|^3),$$

where $\tilde{M}(y) = (\partial^2 H_0/\partial z^2)(y,0)$. Now consider a k-dimensional $(1 \le k \le n)$ submanifold Ξ of R^n defined by

$$\Xi = \{y = y(\lambda), z = 0 : \lambda \in \mathcal{O}\},$$

where $\mathcal{O}$ is a closed bounded region in R^k and $y : \mathcal{O} \to R^n$ is a smooth imbedding. Then with the translation $y - y(\lambda) \mapsto y, z = z$, the Hamiltonian H_0 becomes

$$\begin{aligned} H_0(y,z,\lambda) &= H_0(y(\lambda),0) + \langle \tilde{\omega}(y(\lambda)), y\rangle \\ &\quad + \frac{1}{2}\langle y, \tilde{A}(y(\lambda))y\rangle + \frac{1}{2}\langle z, \tilde{M}(y(\lambda))z\rangle + O(|z|^3) + O(|y||z|^2), \end{aligned}$$

where $\tilde{\omega}(y) = (\partial H_0/\partial y)(y,0)$, $\tilde{A}(y) = (\partial^2 H_0/\partial y^2)(y,0)$. Thus by letting $e(\lambda) = H_0(y(\lambda),0)$, $\omega(\lambda) = \tilde{\omega}(y(\lambda))$, $A(\lambda) = \tilde{A}(y(\lambda))$, $M(\lambda) = \tilde{M}(y(\lambda))$, the Hamiltonian H has the canonical form

$$H = e(\lambda) + \langle \omega(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}(\lambda)\begin{pmatrix} y \\ z \end{pmatrix}\rangle + h(y,z,\lambda) + P(x,y,z,\lambda),$$

with $\mathcal{M} = \mathrm{diag}(A, M)$, and,

$$h(y,z,\lambda) = O(|z|^3) + O(|y||z|^2), \ P(x,y,z,\lambda) = \varepsilon P(y + y(\lambda), z, x).$$

Since P and its derivatives with respect to λ will be sufficiently small in a complex neighborhood of $T^n \times \{0\} \times \{0\}$, an application of Theorems 1 and 2 yields the following.

Corollary 1. *Consider* (2.1) *on the submanifold* Ξ *and assume that* $\tilde{M}$ *is hyperbolic on* Ξ *in the sense of* H).

1) *If* $\tilde{\omega}$ *satisfies the Rüssmann condition on* Ξ*, i.e.,* ω *satisfies* R) *on* $\mathcal{O}$*, then there is an* $\varepsilon_0 > 0$ *and a family of Cantor-like sets* $\Xi_\varepsilon \subset \Xi$*,* $0 < \varepsilon \le \varepsilon_0$*, with* $|\Xi \setminus \Xi_\varepsilon| \to 0$*, as* $\varepsilon \to 0$*, such that for all* $y \in \Xi_\varepsilon$*, the unperturbed* n*-tori* $T_y = \{y\} \times \{0\} \times T^n$ *persist and give rise to a Whitney smooth family of hyperbolic, analytic, Diophantine* n*-tori* $T_{y,\varepsilon}$ *of the perturbed system.*

   Moreover, if for some $1 \le n_0 \le n$*, the* $n_0 \times n_0$ *ordered principal block* $U(y)$ *of* $\tilde{A}(y)$ *is non-singular on* Ξ*, then the first* n_0 *components of the toral frequency of each* $T_{y,\varepsilon}$ *are the same as those of* T_y*.*
2) *If* $\tilde{A}(y)$ *itself is non-singular on* Ξ*, then all unperturbed hyperbolic Diophantine* n*-tori* T_y *on* Ξ *with Diophantine type* (γ,τ)*, where* $0 < \gamma < \varepsilon^{\frac{1}{6n+8}}$ *is arbitrary and* $\tau > n-1$ *is fixed, will persist as* $\varepsilon \to 0$ *with unchanged toral frequencies.*

To give an example, let us consider

$$H(x,y,z)=\frac{1}{2}y_1^2+y_2+\frac{1}{2}(u^2-v^2)+\varepsilon P(x,y,z),$$

where $x=(x_1,x_2)^\top\in T^2, y=(y_1,y_2)^\top\in R^2, z=(u,v)^\top\in R^2$, associated to the standard symplectic structure. Let Ξ be the unit circle in R^2 which we parameterize by $\Xi=\{(\cos\lambda,\sin\lambda):0\le\lambda<2\pi\}$. With the notion above, it is clear that

$$\begin{aligned}\omega(\lambda) &= (\cos\lambda,1)^\top,\\ A(\lambda) &= \begin{pmatrix}1&0\\0&0\end{pmatrix},\\ M(\lambda) &= \begin{pmatrix}1&0\\0&-1\end{pmatrix}.\end{aligned}$$

Hence the condition H) is clearly satisfied on Ξ. Since

$$\mathrm{rank}\{\partial^\alpha\omega(\lambda):\ \forall|\alpha|\le 1\}=\mathrm{rank}\begin{pmatrix}\cos\lambda & -\sin\lambda\\ 1 & 0\end{pmatrix}=2,\ \ \lambda\neq 0,\ \pi,$$

R) is also satisfied. An application of Corollary 2 yields the persistence of the majority of the unperturbed 2-tori on Ξ after a small perturbation. Moreover, since the 1×1 principal block of A equals 1, the first component of the frequency of a perturbed torus $T_{\lambda,\varepsilon}$ is simply $\cos\lambda$.

2.2. **Non-integrable KAM problem.** Let $H_0(p,q)$, $(p,q)\in R^d\times R^d$ be a real analytic Hamiltonian on R^{2d} endorsed with the standard symplectic structure. We assume that H_0 admits an analytic, invariant, Diophantine n-torus T_0 for some $1\le n<d$, i.e., there are real analytic functions $\hat P,\hat Q:T^n\to R^d$, and a Diophantine vector $\omega_0\in R^n$, such that $T_0=\mathrm{cl}\{(\hat P(\omega_0 t),\hat Q(\omega_0 t)):t\in R\}$. Let $x\in T^n$ be the standard coordinate on T^n. One can apply the Frobenius' theorem ([16]) to obtain a symplectic coordinate system $(x,y,z)\in T^n\times R^n\times R^{2d-2n}$ in the vicinity of T_0 such that $T_0=\{y=0,z=0\}$ and

$$H_0=e+\langle\omega_0,y\rangle+\frac{1}{2}\langle\begin{pmatrix}y\\z\end{pmatrix},\mathcal{M}(x)\begin{pmatrix}y\\z\end{pmatrix}\rangle+O((|y|+|z|)^3).$$

Write $\mathcal{M}$ into blocks like (1.4) and assume that a) $[A]$ is non-singular; b) $[M]$ is hyperbolic in the sense of H); c) $|M-[M]|,|B-[B]|$ are sufficiently small. Then an application of Theorem 2 2) yields the persistence as well as the frequency preservation of T_0 after a small perturbation of H_0. In the case that a smooth family of such invariant n-tori T_λ of H are given, one can show the persistence of the majority of invariant tori in the family under either the Rüssmann condition R) or the non-degeneracy of $[A_\lambda]$ according to Theorems 1 and 2.

The non-integrable persistence problem is even more significant when the persistence of a smooth family of maximal dimensional invariant tori $T_\lambda\simeq T^d$, $\lambda\in\mathcal{O}$, is considered for H_0, where $\mathcal{O}\subset R^k$ is a bounded closed region.

By introducing a symplectic coordinate $(x,y)\in D(r,s)=\{(x,y):\ |\mathrm{Im}x|<r,|y|<s\}$ in the vicinity of the invariant tori, the perturbed Hamiltonian $H_0+\varepsilon P$ can be written in the form

$$H=e(\lambda)+\langle\omega(\lambda),y\rangle+\frac{1}{2}\langle y,A(x,\lambda)y\rangle+O(|y|^3)+\varepsilon P(x,y,\lambda). \tag{2.2}$$

For a fixed Diophantine toral frequency $\omega = \omega_0$, persistence of an invariant n-torus of (2.2) for small ε has been shown for the non-degenerate case (i.e., $[A]$ is a non-singular constant matrix) by Salamon and Zehnder ([21]) using the Lagrangian formalism and by Eliasson ([6]) using Lindstedt series. A generalization of these works is recently made in [5] without using action-angle variables.

In the frequency varying case with possible degeneracy of $\omega(\lambda)$, an immediate consequence of Theorems 1 and 2 is the following.

Corollary 2. *Consider* (2.2).

1) *If $\omega(\lambda)$ satisfies the Rüssmann condition* R) *for $n = d$, then there is a sufficiently small $\varepsilon_0 > 0$ and a family of Cantor-like sets $\mathcal{O}_\varepsilon \subset \mathcal{O}$, $0 < \varepsilon \leq \varepsilon_0$, with $|\mathcal{O} \setminus \mathcal{O}_\varepsilon| \to 0$, as $\varepsilon \to 0$, such that for all $\lambda \in \mathcal{O}_\varepsilon$, the unperturbed d-tori $T_\lambda = T^d \times \{0\}$ persist and give rise to a Whitney smooth family of slightly deformed, analytic, Diophantine, invariant d-tori of the perturbed system.*

   Moreover, if for some $1 \leq d_0 \leq d$, the $d_0 \times d_0$ ordered principal block of $[A]$ is non-singular on $\mathcal{O}$, then the first d_0 components of the toral frequency of each perturbed torus coincide with those of the corresponding unperturbed toral frequency.
2) *If $[A]$ itself is non-singular on $\mathcal{O}$, then all Diophantine tori T_λ of Diophantine type (γ, τ), where $0 < \gamma < \varepsilon^{\frac{1}{6d+8}}$ is arbitrary and $\tau > n - 1$ is fixed, will persist as $\varepsilon \to 0$ with unchanged toral frequencies.*

Part 2) of Corollary 3 particularly holds when the persistence of a fixed Diophantine torus in a non-integrable Hamiltonian system is considered. Assume that the Hamiltonian H_0 admits a Diophantine invariant d-torus T_0 with toral frequency ω_0. Let $(x, y) \in T^d \times R^d$ be a symplectic coordinate system in the vicinity of T_0 such that $T_0 = \{y = 0\}$. Then with respect to the new coordinate the Hamiltonian H_0 becomes

$$H_0 = e + \langle \omega_0, y \rangle + \frac{1}{2} \langle y, A(x) y \rangle + O(|y|^3),$$

where $A(x) = (\partial^2 H / \partial y^2)(x, 0)$. Applying part 2) of Corollary 3, it is clear that if H_0 is non-degenerate on T_0, i.e., A is non-singular on T^d, then not only does T_0 persist under a small perturbation of H_0 but also the perturbed toral frequency is kept unchanged.

3. KAM Step

In this section, we describe our linear iterative scheme with respect to (1.3) for one KAM step, under the conditions of the first part of Theorem 2. As we shall see in the sequel, Theorem 1 and the second part of Theorem 2 can be more or less treated as special cases of the first part of Theorem 2 by taking $n_0 = 0$ and n respectively (to unify the notation, $n_0 = 0$ means the omission of all U-related terms in the assumptions of Theorems 1 and 2). Below, we let $\tau > \max\{n(n-1) - 1, 0\}$ be fixed. Also, for simplicity, we set $l_0 = n$.

Initially, set $e_0 = e$, $\Omega_0 = \omega$, $\mathcal{M}^0 = \mathcal{M}$, $A^0 = A$, $B^0 = B$, $M^0 = M$, $h_0 = h$, $P_0 = P$, $\mathcal{O}_0 = \mathcal{O}$, $\beta_0 = s$, $r_* = r$, $\gamma_0 = \gamma$, $s_0 = \gamma_0^{n+1+\frac{\alpha_0}{2}} \mu^{\frac{1}{4}}$, $\mu_0 = s^2 \gamma_0^{n+1+\frac{\alpha_0}{2}} \mu^{\frac{1}{2}}$,

$\mu_* = \mu$. We also write $[A^0](= [A])$ into blocks:

$$[A^0] = \begin{pmatrix} U^0 & E^0 \\ (E^0)^\top & V^0 \end{pmatrix},$$

where $U^0 = U$. Without loss of generality, assume that $0 < r_*, \beta_0, \mu_* \le 1$, $s_0 \le \beta_0$. By (1.7), we have that

$$|\partial_\lambda^l P_0|_{D(r_0,s_0)} \le \gamma_0^{n+1} s_0^2 \mu_0, \ |l| \le n, \tag{3.1}$$

where $r_0 \in (0, r_*]$ will be specified in Section 3.3.

Suppose at a KAM step, say the νth step, we have arrived at a Hamiltonian

$$\begin{aligned} &H = H_\nu = N + P, \\ &N = N_\nu(x,y,z,\lambda) = e + \langle \Omega(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}(x,\lambda) \begin{pmatrix} y \\ z \end{pmatrix}\rangle + h_0(x,y,z,\lambda), \end{aligned} \tag{3.2}$$

where $(x,y,z) \in D = D_\nu = D(r,s)$, $0 < r = r_\nu \le r_0$, $0 < s = s_\nu \le s_0$, $\gamma = \gamma_\nu \le \gamma_0$, $\lambda \in \mathcal{O} = \mathcal{O}_\nu \subset \mathcal{O}_0$, $e(\lambda) = e_\nu(\lambda)$, $\Omega(\lambda) = \Omega_\nu(\lambda)$ are smooth on $\mathcal{O}$ with $(\Omega(\lambda))_i = \omega_i(\lambda)$, $1 \le i \le n_0$, $\mathcal{M}(x,\lambda) = \mathcal{M}^\nu(x,\lambda)$ is real symmetric over $\mathcal{D} \times \mathcal{O} = \{x : |\mathrm{Im} x| < r\} \times \mathcal{O}$ which is smooth in $\lambda \in \mathcal{O}$ and real analytic in $x \in \mathcal{D} = \mathcal{D}_\nu = \mathcal{D}(r)$, $P = P_\nu(x,y,z,\lambda)$ is real analytic in $(x,y,z) \in D$, smooth in $\lambda \in \mathcal{O}$, and moreover,

$$|\partial_\lambda^l P|_{D\times\mathcal{O}} \le \gamma^{n+1} s^2 \mu, \ |l| \le n,$$

for some $\mu = \mu_\nu > 0$.

We shall construct a symplectic transformation $\Phi = \Phi_{\nu+1}$ which transforms the Hamiltonian (3.2), in smaller phase and frequency domains, to the desired Hamiltonian in the next KAM cycle (the $(\nu+1)$th KAM step).

Below, for simplicity, quantities (domains, normal form, perturbation, etc.) in the next KAM cycle will be simply indexed by " $+$ " ($=\nu+1$) and we shall often suspend the dependence of functions on their arguments. Also, all constants $c_1 - c_7$ in this section are positive and independent of the iteration process. We shall also use $c = c(r_0, \beta_0, l_0, \sigma_0)$ to denote any intermediate positive constant which is independent of the iteration process.

Let b, σ, d be sufficiently small positive constants such that

$$\begin{aligned} &\sigma - (b+\sigma)(2b+3\sigma) > 0, \ \delta(1+b+\sigma) > 1, \\ &(n+1+\frac{a_0}{2})(1-2b-3\sigma) > n+1, \end{aligned}$$

where $\delta = 1 - d$, $a_0 \in (0, 1/3)$.

Define

$$\begin{aligned}
\gamma_+ &= \frac{\gamma_0}{4}+\frac{\gamma}{2},\\
r_+ &= \delta r + d(1-\frac{\delta^2}{2})r_0,\\
s_+ &= s^{1+b+\sigma},\\
\beta_+ &= \frac{\beta}{2}+\frac{\beta_0}{4},\\
K_+ &= ([\log\frac{1}{s}]+1)^3,\\
D(a) &= D(r_+ + \frac{6}{8}(r-r_+),a),\ \ a>0,\\
\mathcal{D}(a) &= \{x: |\mathrm{Im} x|<a\},\ \ a>0,\\
\Gamma(a) &= e^{\frac{r_0(1-\delta)\delta^2}{16}}\sum_{0<|k|\le K_+}|k|^{n+(n+1)\tau+6}e^{-|k|\frac{a}{8}},\ \ a>0,\\
D_+ &= D(r_+,s_+),\\
\mathcal{D}_+ &= \mathcal{D}(r_+)=\{x: |\mathrm{Im} x|<r_+\},\\
\tilde{D}_+ &= D(r_+ + \frac{5}{8}(r-r_+),\beta_+),\\
D_i &= D(r_+ + \frac{i-1}{8}(r-r_+), is_+),\ \ i=1,2,\cdots,8.
\end{aligned}$$

3.1. **Truncation.** We express P into Taylor-Fourier series

$$P=\sum_{i\in Z_+^n,j\in Z_+^{2m},k\in Z^n} p_{kij}y^iz^je^{\sqrt{-1}\langle k,x\rangle}$$

and consider the truncation

$$\begin{aligned}
R &= \sum_{|i|+|j|<3,|k|\le K_+} p_{kij}y^iz^je^{\sqrt{-1}\langle k,x\rangle}=\sum_{|k|\le K_+}(P_{k00}+\langle P_{k10},y\rangle\\
&\quad +\langle P_{k01},z\rangle+\langle y,P_{k20}y\rangle+\langle y,P_{k11}z\rangle+\langle z,P_{k02}z\rangle)e^{\sqrt{-1}\langle k,x\rangle}. \qquad (3.3)
\end{aligned}$$

Lemma 3.1. *Assume that*

H1) $s_+\le \frac{s}{16}$;

H2) $\int_{K_+}^{\infty}\lambda^n e^{-\lambda\frac{r-r_+}{16}}\mathrm{d}\lambda\le s.$

Then there is a constant c_1 such that

$$|\partial_\lambda^l(P-R)|_{D_8}\le c_1\gamma^{n+1}(s^3+\frac{s_+^3}{s})\mu,\ \ |l|\le n.$$

Proof. Denote

$$
\begin{aligned}
I &= \sum_{|k|>K_+} p_{kij} y^i z^j e^{\sqrt{-1}\langle k,x\rangle},\\
II &= \sum_{|k|\le K_+,|i|+|j|\ge 3} p_{kij} y^i z^j e^{\sqrt{-1}\langle k,x\rangle}\\
&= \int \frac{\partial^{(p,q)}}{\partial y^p \partial z^q} \sum_{|k|\le K_+,|i|+|j|\ge 3} p_{kij} e^{\sqrt{-1}\langle k,x\rangle} y^i z^j \mathrm{d}y\mathrm{d}z,\\
D_* &= D(r_+ + \frac{7}{8}(r-r_+), s),
\end{aligned}
$$

where $\int$ is the obvious anti-derivative of $\frac{\partial^{(p,q)}}{\partial y^p \partial z^q}$ for $|p|+|q|=3$. Clearly,

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

We note by H1) that $D_8 \subset D_*$.

Since, by Cauchy's estimate,

$$|\sum_{i\in Z_+^n, j\in Z_+^{2m}} \partial_\lambda^l p_{kij} y^i z^j| \le |\partial_\lambda^l P|_{D(r,s)} e^{-|k|r} \le \gamma^{n+1} s^2 \mu e^{-|k|r},\ \ |l|\le n,$$

H2) implies that

$$
\begin{aligned}
|\partial_\lambda^l I|_{D_*} &\le \sum_{|k|>K_+} \gamma^{n+1} s^2 \mu e^{-|k|r} e^{|k|(r_+ + \frac{7}{8}(r-r_+))}\\
&\le \gamma^{n+1} s^2 \mu \sum_{\kappa=K_+}^{\infty} \kappa^n e^{-\kappa\frac{r-r_+}{8}} \le \gamma^{n+1} s^2 \mu \int_{K_+}^{\infty} \lambda^n e^{-\lambda\frac{r-r_+}{16}} \mathrm{d}\lambda\\
&\le \gamma^{n+1} s^3 \mu,\ \ |l|\le n.
\end{aligned}
$$

It follows that

$$|\partial_\lambda^l (P-I)|_{D_*} \le |\partial_\lambda^l P|_{D(r,s)} + |\partial_\lambda^l I|_{D_*} \le 2\gamma^{n+1} s^2 \mu,\ \ |l|\le n.$$

By Cauchy's estimate of $\partial_\lambda^l(P-I)$ on D_*, we then have

$$
\begin{aligned}
|\partial_\lambda^l II|_{D_8} &\le |\int \frac{\partial^{(p,q)}}{\partial y^p \partial z^q} \sum_{|k|\le K_+,|i|+|j|\ge 3} \partial_\lambda^l p_{kij} e^{\sqrt{-1}\langle k,x\rangle} y^i z^j \mathrm{d}y\mathrm{d}z|_{D_8}\\
&\le |\int |\frac{\partial^{(p,q)}}{\partial y^p \partial z^q} \partial_\lambda^l (P-I-R)|_{D_*} \mathrm{d}y\mathrm{d}z|_{D_8}\\
&\le 2\left(\frac{1}{s-8s_+}\right)^3 \gamma^{n+1} s^2 \mu |\int \mathrm{d}y\mathrm{d}z|_{D_8} \le c\gamma^{n+1} \frac{s_+^3}{s}\mu,\ \ |l|\le n.
\end{aligned}
$$

Thus,

$$|\partial_\lambda^l (P-R)|_{D_8} \le c\gamma^{n+1}(s^3 + \frac{s_+^3}{s})\mu,\ \ |l|\le n.$$

□

3.2. **The linear homological equations.** Write $\mathcal{M}$ into blocks

$$\mathcal{M}(x,\lambda) = \begin{pmatrix} A & B \\ B^\top & M \end{pmatrix},$$

where

$$\begin{aligned} A(x,\lambda) &= \sum_{k\in Z^n} A_k e^{\sqrt{-1}\langle k,x\rangle}, \\ B(x,\lambda) &= \sum_{k\in Z^n} B_k e^{\sqrt{-1}\langle k,x\rangle}, \\ M(x,\lambda) &= \sum_{k\in Z^n} M_k e^{\sqrt{-1}\langle k,x\rangle} \end{aligned}$$

are $n\times n$, $n\times 2m$, $2m\times 2m$ blocks of $\mathcal{M}$ respectively.

To transform (3.2) into the Hamiltonian in the next KAM cycle, a symplectic transformation should at least eliminate all its first order resonant terms

$$P_{k00}e^{\sqrt{-1}\langle k,x\rangle}, \langle P_{k10}, y\rangle e^{\sqrt{-1}\langle k,x\rangle}, \langle P_{k01}, z\rangle e^{\sqrt{-1}\langle k,x\rangle}, \langle P_{001}, z\rangle,\ 0<|k|\le K_+.$$

An essential idea of our linear iterative scheme is to find a Hamiltonian F of the form

$$F = \sum_{0<|k|\le K_+} (f_{k0} + \langle f_{k1}, y\rangle + \langle F_{k1}, z\rangle) e^{\sqrt{-1}\langle k,x\rangle} + \langle F_{01}, z\rangle \tag{3.4}$$

such that the time-1 map ϕ_F^1 of the flow generated by F, as a symplectic transformation, will precisely eliminate the above resonant terms. To be able to fix the first n_0 components of the toral frequencies as stated in Theorem 2 1), we shall also find a $Y_*\in R^{n_0}$ so that the translation of coordinate

$$\phi : x\to x,\ y\to y+\begin{pmatrix} Y_* \\ 0\end{pmatrix},\ z\to z$$

removes all possible drifts among the first n_0 components of the new toral frequencies.

Denote

$$\begin{aligned} [A] &= \begin{pmatrix} U & E \\ E^\top & V\end{pmatrix}, \\ R' &= [R] + \sum_{0<|k|\le K_+} (\langle y, P_{k20}y\rangle + \langle y, P_{k11}z\rangle + \langle z, P_{k02}z\rangle) e^{\sqrt{-1}\langle k,x\rangle} \\ &\quad -\langle P_{001}, z\rangle + \sum_{|k|\le K_+} \langle B_{-k}JF_{k1}, y\rangle, && (3.5) \\ R_t &= (1-t)\{N,F\} + R, && (3.6) \\ y_* &= \begin{pmatrix} Y_* \\ 0\end{pmatrix}, \end{aligned}$$

where U, E, V are the $n_0\times n_0$, $n_0\times(n-n_0)$, $(n-n_0)\times(n-n_0)$ blocks of $[A]$ respectively.

Let

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

Then it is easy to see that

$$\begin{aligned} H_+ &= H\circ\Phi_+ = H\circ\phi_F^1\circ\phi = (N+R)\circ\phi_F^1\circ\phi + (P-R)\circ\phi_F^1\circ\phi \\ &= (N+R')\circ\phi - \langle y_*, (A-[A])y\rangle - \langle y_*, Bz\rangle \\ &\quad +(\{N,F\}+R-R')\circ\phi + \langle y_*, (A-[A])y\rangle + \langle y_*, Bz\rangle \\ &\quad + \int_0^1 \{R_t, F\}\circ\phi_F^t\circ\phi \mathrm{d}t + (P-R)\circ\phi_F^1\circ\phi. \end{aligned}$$

Since the Taylor-Fourier series of $R-R'$ consists of terms of Fourier modes $e^{\sqrt{-1}\langle k,x\rangle}$, $0<|k|\le K_+$, but that of $\{N,F\}$ contains some high modes $e^{\sqrt{-1}\langle k,x\rangle}$ for $|k|>K_+$, we need to choose a function Q of high order such that both equations

$$(\{N,F\}+R-R')\circ\phi - Q + \langle y_*, (A-[A])y\rangle + \langle y_*, Bz\rangle = 0, \tag{3.7}$$

$$\mathrm{diag}(U,O)y_* = \mathrm{diag}(I_{n_0},O)(-P_{010} - \sum_{|j|\le K_+} B_{-j}JF_{j1}) \tag{3.8}$$

are solvable. If this is the case, then it is easy to see that

$$\begin{aligned} H_+ &= N_+ + P_+, \\ N_+ &= e_+ + \langle \Omega_+(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z\end{pmatrix}, \mathcal{M}^+ \begin{pmatrix} y \\ z\end{pmatrix}\rangle + h_0(x,y,z,\lambda) \\ &= e_+ + \langle \Omega_+(\lambda), y\rangle + \frac{1}{2}\langle \begin{pmatrix} y \\ z\end{pmatrix}, \begin{pmatrix} A^+ & B^+ \\ B^{+\top} & M^+ \end{pmatrix} \begin{pmatrix} y \\ z\end{pmatrix}\rangle + h_0(x,y,z,\lambda), \end{aligned}$$

where

$$e_+ = e + P_{000} + \langle \Omega, y_*\rangle, \tag{3.9}$$

$$\Omega_+ = \Omega + \mathrm{diag}(O, I_{n-n_0})([A]y_* + P_{010} + \sum_{|k|\le K_+} B_{-k}JF_{k1}), \tag{3.10}$$

$$A^+ = A + \sum_{|k|\le K_+} 2P_{k20}e^{\sqrt{-1}\langle k,x\rangle}, \tag{3.11}$$

$$B^+ = B + \sum_{|k|\le K_+} P_{k11}e^{\sqrt{-1}\langle k,x\rangle}, \tag{3.12}$$

$$M^+ = M + \sum_{|k|\le K_+} 2P_{k02}e^{\sqrt{-1}\langle k,x\rangle}, \tag{3.13}$$

$$\begin{aligned} P_+ &= \int_0^1 \{R_t, F\}\circ\phi_F^t\circ\phi \mathrm{d}t + (P-R)\circ\phi_F^1\circ\phi \\ &\quad + \frac{1}{2}\langle y_*, (A - 2\mathrm{diag}(U,O))y_*\rangle + h_0(x, y+y_*, z,\lambda) - h_0(x,y,z,\lambda) \\ &\quad + \sum_{|k|\le K_+} (\langle y_*, P_{k20}y_*\rangle + \langle y_*, 2P_{k20}y\rangle + \langle y_*, P_{k11}z\rangle)e^{\sqrt{-1}\langle k,x\rangle} + Q. \end{aligned} \tag{3.14}$$

We now explore equations (3.7) and (3.8). Let

$$\begin{aligned} Q = (&\sum_{0<|k|\le K_+} (\langle \frac{1}{2}\partial_x\langle y, A(x,\lambda)y\rangle + \partial_x\langle y, B(x,\lambda)z\rangle + \frac{1}{2}\partial_x\langle z, M(x,\lambda)z\rangle \\ &+\partial_x h_0(x,y,z,\lambda), f_{k1}\rangle + (-\sqrt{-1}\langle k, A(x,\lambda)y + B(x,\lambda)z\rangle \\ &+\partial_y h_0(x,y,z,\lambda))(f_{k0} + \langle f_{k1}, y\rangle + \langle F_{k1}, z\rangle))e^{\sqrt{-1}\langle k,x\rangle} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{|k|>K_+}(\langle B_kJF_{01},y\rangle+\langle M_kJF_{01},z\rangle)e^{\sqrt{-1}\langle k,x\rangle}\\
&+\sum_{|k|>K_+,0<|j|\le K_+}(\langle B_{k-j}JF_{j1},y\rangle+\langle M_{k-j}JF_{j1},z\rangle)e^{\sqrt{-1}\langle k,x\rangle}\\
&+\sum_{0<|k|\le K_+}\langle\partial_z h_0(x,y,z,\lambda)JF_{k1},z\rangle e^{\sqrt{-1}\langle k,x\rangle})\circ\phi\\
&+\sum_{0<|k|\le K_+}(-\sqrt{-1}\langle f_{k1},y_*\rangle+\langle B^\top(x,\lambda)y_*,JF_{k1}\rangle+\langle P_{k10},y_*\rangle)e^{\sqrt{-1}\langle k,x\rangle}\\
&+\sum_{|k|>K_+}(\langle y_*,A_ky\rangle+\langle y_*,B_kz\rangle)e^{\sqrt{-1}\langle k,x\rangle}+\langle B^\top(x,\lambda)y_*,JF_{01}\rangle.
\end{aligned}
\tag{3.15}
$$

Substituting (3.3)-(3.5) and (3.15) into (3.7) yields

$$
\begin{aligned}
&-\sum_{0<|k|\le K_+}\sqrt{-1}\langle k,\Omega(\lambda)\rangle(f_{k0}+\langle f_{k1},y\rangle+\langle F_{k1},z\rangle)e^{\sqrt{-1}\langle k,x\rangle}\\
&+\sum_{|j|\le K_+}\langle z,M_{-j}JF_{j1}\rangle+\sum_{0<|k|\le K_+,|j|\le K_+}(\langle y,B_{k-j}JF_{j1}\rangle\\
&+\langle z,M_{k-j}JF_{j1}\rangle)e^{\sqrt{-1}\langle k,x\rangle}+\sum_{0<|k|\le K_+}(\langle y_*,A_ky\rangle+\langle y_*,B_kz\rangle+P_{k00}+\langle P_{k10},y\rangle\\
&+\langle P_{k01},z\rangle)e^{\sqrt{-1}\langle k,x\rangle}+\langle P_{001}+[B]^\top y_*,z\rangle=0.
\end{aligned}
$$

By comparing coefficients above, equations (3.7), (3.8) give rise to the following linear homological equations for all $0<|k|\le K_+$:

$$\sqrt{-1}\langle k,\Omega(\lambda)\rangle f_{k0}=P_{k00},\tag{3.16}$$

$$\sqrt{-1}\langle k,\Omega(\lambda)\rangle f_{k1}=P_{k10}+A_ky_*+\sum_{|j|\le K_+}B_{k-j}JF_{j1},\tag{3.17}$$

$$
\begin{aligned}
\sqrt{-1}\langle k,\Omega(\lambda)\rangle F_{k1}-[M]JF_{k1}=&\sum_{0<|j|\le K_+,j\ne k}M_{k-j}JF_{j1}\\
&+P_{k01}+B_k^\top y_*+M_kJF_{01},
\end{aligned}
\tag{3.18}
$$

$$[M]JF_{01}=-P_{001}-\sum_{0<|j|\le K_+}M_{-j}JF_{j1}-[B]^\top y_*,\tag{3.19}$$

$$\mathrm{diag}(U,O)y_*=\mathrm{diag}(I_{n_0},O)(-P_{010}-\sum_{0<|j|\le K_+}B_{-j}JF_{j1}-[B]JF_{01}).\tag{3.20}$$

Let

$$
\begin{aligned}
\mathcal{O}_+&=\{\lambda\in\mathcal{O}:|\langle k,\Omega(\lambda)\rangle|>\frac{\gamma}{|k|^\tau},\ 0<|k|\le K_+\},\\
Y&=[M]J-[B]^\top\mathrm{diag}(U^{-1},O)[B]J,\\
Y^0&=[M^0]J-[B^0]^\top\mathrm{diag}((U^0)^{-1},O)[B^0]J.
\end{aligned}
\tag{3.21}
$$

Then $\langle k,\Omega(\lambda)\rangle$ is invertible on $\mathcal{O}_+$ for all $0<|k|\le K_+$. If we assume that

H3) $|\partial_\lambda^l(\mathcal{M}-\mathcal{M}^0)|_{\mathcal{D}(r)\times\mathcal{O}}\le\gamma_0^{n+1}\mu_*^{\frac14}$, $|l|\le n$, then as μ_* small, both U and Y are non-singular on $\mathcal{O}$. Thus, for all $0<|k|\le K_+$, $\lambda\in\mathcal{O}_+$, (3.16) is immediately solvable, and, (3.17), (3.19) and (3.20) can be also solved in terms of F_{k1}, $0<|k|\le K_+$. More precisely, we have

$$f_{k0} = -\sqrt{-1}\langle k,\Omega(\lambda)\rangle^{-1}P_{k00}, \tag{3.22}$$

$$\begin{aligned} f_{k1} &= -\sqrt{-1}\langle k,\Omega(\lambda)\rangle^{-1}(P_{k10} - A_k \mathrm{diag}(U^{-1},O)(P_{010} \\ &\quad + \sum_{0<|j|\le K_+} B_{-j}JF_{j1}) + (B_k - A_k\mathrm{diag}(U^{-1},O)[B])JY^{-1}(-P_{001} \\ &\quad - \sum_{0<|j|\le K_+} M_{-j}JF_{j1} + [B]^\top \mathrm{diag}(U^{-1},O)P_{010} \\ &\quad +\mathrm{diag}(U^{-1},O)\sum_{0<|j|\le K_+} B_{-j}JF_{j1}) + \sum_{0<|j|\le K_+} B_{k-j}JF_{j1}), \end{aligned} \tag{3.23}$$

$$\begin{aligned} F_{01} &= Y^{-1}(-P_{001} - \sum_{0<|j|\le K_+} M_{-j}JF_{j1} + [B]^\top\mathrm{diag}(U^{-1},O)P_{010} \\ &\quad +\mathrm{diag}(U^{-1},O)\sum_{0<|j|\le K_+} B_{-j}JF_{j1}), \end{aligned} \tag{3.24}$$

$$\begin{aligned} y_* &= \mathrm{diag}(U^{-1},O)(-P_{010} - \sum_{0<|j|\le K_+} B_{-j}JF_{j1} - [B]JY^{-1}(-P_{001} \\ &\quad - \sum_{0<|j|\le K_+} M_{-j}JF_{j1} + [B]^\top\mathrm{diag}(U^{-1},O)P_{010} \\ &\quad +\mathrm{diag}(U^{-1},O)\sum_{0<|j|\le K_+} B_{-j}JF_{j1})). \end{aligned} \tag{3.25}$$

Substituting (3.24), (3.25) into (3.18), it is easy to see that F_{k1}, $0<|k|\le K_+$, satisfy the following equations:

$$L_kF_{k1} = \sum_{0<|j|\le K_+} M_{kj}F_{j1} + \mathcal{P}_k, \tag{3.26}$$

where

$$L_k = \sqrt{-1}\langle k,\Omega(\lambda)\rangle I_{2m} - [M]J, \tag{3.27}$$

$$M_{kj} = \begin{cases} \bar{M}_{kk}, & j=k, \\ M_{k-j}J + \bar{M}_{kj}, & j\ne k, \end{cases} \tag{3.28}$$

$$\begin{aligned} \mathcal{P}_k &= P_{k01} - B_k^\top\mathrm{diag}(U^{-1},O)P_{010} + B_k^\top\mathrm{diag}(U^{-1},O)[B]JY^{-1}P_{001} \\ &\quad -B_k^\top\mathrm{diag}(U^{-1},O)[B]JY^{-1}[B]^\top\mathrm{diag}(U^{-1},O)P_{010} \\ &\quad -M_kJY^{-1}P_{001} + M_kJY^{-1}[B]^\top\mathrm{diag}(U^{-1},O)P_{010}, \end{aligned} \tag{3.29}$$

with

$$\begin{aligned} \bar{M}_{kj} &= M_kJY^{-1}(-M_{-j}J + \mathrm{diag}(U^{-1},O)B_{-j}J) \\ &\quad +B_k^\top\mathrm{diag}(U^{-1},O)(-B_{-j}J + [B]JY^{-1}(M_{-j}J - \mathrm{diag}(U^{-1},O)B_{-j}J)). \end{aligned}$$

Thus, equations (3.16)-(3.20) are solvable on $\mathcal{O}_+$ if and only if (3.26) is. In fact, for the sake of convergence of the symplectic transformations to be constructed, not only do these equations need to be solvable, but also their solutions should satisfy certain exponential decay properties. This motivates us to consider the following weighted functions:

$$\begin{aligned}\tilde{F}_{k1} &= e^{|k|(r_+ + \frac{7}{8}(r-r_+))}F_{k1},\ 0<|k|\le K_+,\\ \tilde{M}_{kj} &= e^{(|k|-|j|)(r_+ + \frac{7}{8}(r-r_+))}M_{kj},\ 0<|k|,\ |j|\le K_+,\\ \tilde{\mathcal{P}}_k &= e^{|k|(r_+ + \frac{7}{8}(r-r_+))}\mathcal{P}_k,\ 0<|k|\le K_+.\end{aligned}$$

Let $\mathcal{T}$ = " $<$ " be a fixed ordering on Z_+^n with the property that whenever $k, k_* \in Z_+^n$ with $|k| < |k_*|$, then $k < k_*$. We define

$$\mathcal{F} = \begin{pmatrix} \vdots \\ \tilde{F}_{j1} \\ \vdots \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} \vdots \\ \tilde{\mathcal{P}}_j \\ \vdots \end{pmatrix}$$

as the column vectors which vertically line up all $\tilde{F}_{j1}$, $\tilde{\mathcal{P}}_j$, $0 < |j| \le K_+$, respectively, according to the ordering $\mathcal{T}$. We also form the matrices

$$\Lambda = \mathrm{diag}(L_k), \quad \mathcal{A} = (\tilde{M}_{kj}), \tag{3.30}$$

according to the same ordering $\mathcal{T}$. Then it is clear that the equations in (3.26) can be putted into the following system form:

$$(\Lambda - \mathcal{A})\mathcal{F} = \mathcal{P}. \tag{3.31}$$

3.3. **Invertibility of $\Lambda - \mathcal{A}$.** The invertibility of $\Lambda - \mathcal{A}$ will be shown by a perturbation argument.

Let $\{M_k^0\}$, $\{B_k^0\}$ be the Fourier coefficients of M^0, B^0 respectively, and let M_{kj}^0, $0 < |k|, |j| \le K_+$, be defined as in (3.28) with M_k^0, B_k^0, U^0, Y^0 in place of M_k, B_k, U, Y, for all $0 < |k| \le K_+$, respectively. Denote

$$L_k^0 = \sqrt{-1}\langle k, \omega\rangle - [M^0]J, \quad \tilde{M}_{kj}^0 = e^{(|k|-|j|)(r_+ + \frac{7}{8}(r-r_+))}M_{kj}^0,\ 0<|j|,\ |k|\le K_+,$$

and let Λ^0, $\mathcal{A}^0$ be defined as in (3.30) with L_k^0, $\tilde{M}_{kj}^0$ in place of L_k, $\tilde{M}_{kj}$, $0 < |j|,\ |k| \le K_+$, respectively. We first show the invertibility of $\Lambda^0 - \mathcal{A}^0$ on $\mathcal{O}_0$, along with the estimate of an upper bound for its inverse.

By the hyperbolicity of $[M^0]J$, it is easy to see that all L_k^0, $0 < |k| \le K_+$, are non-singular on $\mathcal{O}_0$. This implies that Λ^0 is non-singular on $\mathcal{O}_0$. To obtain an upper bound of $|(\Lambda^0)^{-1}|$ on $\mathcal{O}_0$, we need to estimate a uniform bound for all $|L_k^{0^{-1}}|_{\mathcal{O}_0}$, $0 < |k| \le K_+$.

To do so, let $L = L(\lambda)$ be non-singular on $\mathcal{O}_0$ such that

$$L^{-1}[M^0]JL = E,$$

where $E = E(\lambda)$ denotes the Jordan canonical form of $[M^0]J = [M^0(\lambda)]J$ for each $\lambda \in \mathcal{O}_0$. Due to the possible change of multiplicities of eigenvalues of $[M^0]J$, L is in general not continuous on $\mathcal{O}_0$. However, by the standard Q-R decomposition procedure, for each $\lambda \in \mathcal{O}_0$, there is an upper triangular matrix $S = S(\lambda)$ such that LS is orthogonal. Denote $\tilde{E} = \tilde{E}(\lambda) \equiv S^{-1}ES$, then $\tilde{E}$ is also upper triangular, and,

$$(LS)^\top [M^0]J(LS) = (LS)^{-1}[M^0]J(LS) = S^{-1}ES = \tilde{E}.$$

Let $|\cdot|_2$ denote the matrix norm induced by the Euclidean vector norm. Then for any $n \times 2m$ matrix (a_{ij}),

$$\frac{1}{\sqrt{2m}}|(a_{ij})| \le |(a_{ij})|_2 \le \sqrt{2m}|(a_{ij})|.$$

It follows that

$$|\tilde{E}| \le \sqrt{2m}|\tilde{E}|_2 = \sqrt{2m}|[M^0]J|_2 \le 2m|[M^0]|, \tag{3.32}$$

$$\begin{aligned} |(L_k^0)^{-1}| &\le \sqrt{2m}|(L_k^0)^{-1}|_2 = \sqrt{2m}|(\sqrt{-1}\langle k,\omega\rangle I_{2m} - \tilde{E})^{-1}|_2 \\ &\le 2m|(\sqrt{-1}\langle k,\omega\rangle I_{2m} - \tilde{E})^{-1}|. \end{aligned} \tag{3.33}$$

Since the eigenvalues of $[M^0]J$ coincide with those of $\tilde{E}$, eigenvalues of $\sqrt{-1}\langle k,\omega\rangle I_{2m} - \tilde{E}$ are bounded below in absolute value by σ_0. Using H), (3.32), (3.33) and the standard inverse formula for upper triangular matrices, we have that

$$\begin{aligned} |(L_k^0)^{-1}| &\le 2m|(\sqrt{-1}\langle k,\omega\rangle I_{2m} - \tilde{E})^{-1}| \le \frac{2m}{\sigma_0}(1 + \frac{|\tilde{E}|_{\mathcal{O}_0}}{\sigma_0})^{2m-1} \\ &\le \frac{2m}{\sigma_0}(1 + \frac{2m}{\sigma_0}|[M^0]|_{\mathcal{O}_0})^{2m-1} = \frac{\rho_0}{2} \end{aligned}$$

for all $\lambda \in \mathcal{O}_0$, $0 < |k| \le K_+$, i.e.,

$$|(\Lambda^0)^{-1}|_{\mathcal{O}_0} \le \frac{\rho_0}{2}. \tag{3.34}$$

Next, we give an estimate for $|\mathcal{A}^0|_{\mathcal{O}_0}$.

Let $\mathcal{U}(x) = (u_{ij}(x))$ be a real analytic, matrix valued function defined on $\mathcal{D}(a)$ $(a > 0)$ and denote $\mathcal{U}_k = (u_{kij})$, $k \in Z^n \setminus \{0\}$, as the kth Fourier coefficient of $\mathcal{U}(x)$. Since

$$\mathcal{U}_k = \int_{T^n} (\mathcal{U} - [\mathcal{U}])e^{-\sqrt{-1}\langle k,x\rangle}\mathrm{d}x,$$

Cauchy's estimate yields that

$$|\mathcal{U}_k| \le |\mathcal{U} - [\mathcal{U}]|_{\mathcal{D}(a)}e^{-|k|a}. \tag{3.35}$$

It also follows from Parseval's identity that

$$\begin{aligned} \sum_{0<|k|\le\tilde{K}} |\mathcal{U}_k| &= \sum_{0<|k|\le\tilde{K}} \max_i \sum_j |u_{kij}| \le \sqrt{2m} \sum_{0<|k|\le\tilde{K}} \sum_{i,j} |u_{kij}|^2 \\ &= \sqrt{2m}\sum_{i,j}(\sum_{0<|k|\le\tilde{K}} |u_{kij}|^2)^{\frac{1}{2}} \le 2m \max_i \sum_j \|u_{ij} - [u_{ij}]\|_{L^2(\mathcal{D}(a))} \\ &\le 2m \max_i \sum_j |u_{ij} - [u_{ij}]|_{\mathcal{D}(a)} = 2m|\mathcal{U} - [\mathcal{U}]|_{\mathcal{D}(a)} \end{aligned} \tag{3.36}$$

for any $\tilde{K} > 0$.

Applying (3.36) to the formula for M^0_{kj} similar to (3.28), we see that, for any $0 < |k| \le K_+$, $\lambda \in \mathcal{O}_0$,

$$\begin{aligned}
&\sum_{0<|j|\le K_+} |M^0_{kj}| = |M^0_{kk}| + \sum_{0<|j|\le K_+, j\ne k} |M^0_{kj}| \\
&\le |M^0 - [M^0]|_{\mathcal{D}(r_*)} + (1+2m)|(Y^0)^{-1}||M^0 - [M^0]|^2_{\mathcal{D}(r_*)} \\
&+(1+2m)|(Y^0)^{-1}||(U^0)^{-1}||M^0 - [M^0]|_{\mathcal{D}(r_*)}|B^0 - [B^0]|_{\mathcal{D}(r_*)} \\
&+(1+2m)|(Y^0)^{-1}||(U^0)^{-1}||B^0 - [B^0]|^2_{\mathcal{D}(r_*)} \\
&+(1+2m)|(U^0)^{-1}||B^0 - [B^0]|^2_{\mathcal{D}(r_*)} + (1+2m)|[B^0]||(Y^0)^{-1}||(U^0)^{-1}| \\
&|M^0 - [M^0]|_{\mathcal{D}(r_*)}|B^0 - [B^0]|_{\mathcal{D}(r_*)} \\
&+(1+2m)|[B^0]||(Y^0)^{-1}||(U^0)^{-1}|^2|B^0 - [B^0]|^2_{\mathcal{D}(r_*)} < \eta + \alpha\eta^2 = \frac{1}{\rho_0},
\end{aligned} \tag{3.37}$$

where η, α are as in (1.8), (1.9) respectively, i.e.,

$$\begin{aligned}
\eta &= \frac{2}{\sqrt{\rho_0^2 + 4\alpha\rho_0} + \rho_0}, \\
\alpha &= (1+2m)(|(Y^0)^{-1}| + |(U^0)^{-1}| \\
&+ (|(Y^0)^{-1}||(U^0)^{-1}|)(1 + |[B^0]| + |[B^0]||(U^0)^{-1}|))_{\mathcal{O}_0}.
\end{aligned}$$

We now choose r_0 in (3.1). Define

$$\mathcal{A}^0(a) = (e^{(|k-j|)a} M^0_{kj})$$

as the one parameter family of matrices which are of the same dimension as $\mathcal{A}$ and are formed according to the ordering $\mathcal{T}$. Then (3.37) clearly implies that

$$|\mathcal{A}^0(0)|_{\mathcal{O}_0} < \frac{1}{\rho_0}.$$

Note that, by (3.35),

$$\begin{aligned}
\sum_{|j|>0} e^{|k-j|a}|\partial^l_\lambda M^0_{kj}| &\le c\sum_{|j|>0} e^{-|k-j|(r_*-a)} \le c\sum_{|j|\ge 0} e^{-|j|(r_*-a)} < \infty, \\
\sum_{|j|>0} |k-j|e^{|k-j|a}|\partial^l_\lambda M^0_{kj}| &\le c\sum_{|j|>0} |k-j|e^{-|k-j|(r_*-a)} \\
&\le c\sum_{|j|\ge 0} |j|e^{-|j|(r_*-a)} < \infty
\end{aligned}$$

for all $|l| \le n$. It follows that the family of functions

$$\eta_{k,\tilde{K}}(a) = \sum_{0<|j|\le \tilde{K}} e^{|k-j|a}|M^0_{kj}|$$

is uniformly bounded and equi-continuous. Hence

$$|\mathcal{A}^0(a)|_{\mathcal{O}_0} = \max_{0<|k|\le K_+} \sum_{0<|j|\le K_+} e^{|k-j|a}|M^0_{kj}|_{\mathcal{O}_0}$$

is continuous in $a \in [0, r_*)$ uniformly in K_+.

Let $r_0 = r_0(r_*, \sigma_0, U^0) \in (0, \frac{r_*}{2}]$ be fixed such that

$$|\mathcal{A}^0((1 - \frac{d\delta^2}{16})r_0)|_{\mathcal{O}_0} < \frac{1}{\rho_0}.$$

Denote

$$\xi(a) = (\frac{7}{8} + \frac{\delta}{8})a + \frac{d}{8}(1 - \frac{\delta^2}{2})r_0. \tag{3.38}$$

Then

$$\begin{aligned} \xi(r) &= r_+ + \frac{7}{8}(r - r_+), \\ \xi(r_0) &= (1 - \frac{d\delta^2}{16})r_0. \end{aligned}$$

Since

$$e^{(|k|-|j|)\xi(r)}|M^0_{kj}| \le e^{|k-j|\xi(r)}|M^0_{kj}| < e^{|k-j|\xi(r_0)}|M^0_{kj}|,\ 0 < |k|,\ |j| \le K_+,$$

we have that

$$|\mathcal{A}^0|_{\mathcal{O}_0} \le |\mathcal{A}^0(\xi(r_0))|_{\mathcal{O}_0} < \frac{1}{\rho_0}. \tag{3.39}$$

Thus, by (3.34),

$$|(\Lambda^0 - \mathcal{A}^0)^{-1}|_{\mathcal{O}_0} \le \frac{|(\Lambda^0)^{-1}|_{\mathcal{O}_0}}{1 - |\mathcal{A}^0|_{\mathcal{O}_0}|(\Lambda^0)^{-1}|_{\mathcal{O}_0}} < 2\rho_0.$$

Lemma 3.2. *Assume* H3) *and also that*

H4) $|\partial^l_\lambda \mathcal{A} - \partial^l_\lambda \mathcal{A}^0|_{\mathcal{O}} < \mu_*^{\frac{1}{4}}$.

Then as μ_* sufficiently small, $\mathcal{L} = \Lambda - \mathcal{A}$ is non-singular on $\mathcal{O}$, and moreover, there is a constant c_2 such that

$$|\partial^l_\lambda \mathcal{L}|_{\mathcal{O}} \le c_2 K_+,\ |l| \le n.$$

Proof. Given $\epsilon_0 > 0$ small. Since, by H3),

$$|U - U^0|_{\mathcal{O}},\ |Y - Y^0|_{\mathcal{O}},\ |M - M^0|_{\mathcal{O}} \le |\mathcal{M} - \mathcal{M}^0|_{\mathcal{O}} \le \mu_*^{\frac{1}{4}},$$

we can choose μ_* small, say, $\mu_* \le \dfrac{\epsilon_0^4}{1+\epsilon_0} \min\{\dfrac{1}{|(U^0)^{-1}|_{\mathcal{O}_0}}, \dfrac{1}{|(Y^0)^{-1}|_{\mathcal{O}_0}}\}$, such that

$$|U^{-1}|_{\mathcal{O}} \le (1+\epsilon_0)|(U^0)^{-1}|_{\mathcal{O}_0}, \tag{3.40}$$

$$|Y^{-1}|_{\mathcal{O}} \le (1+\epsilon_0)|(Y^0)^{-1}|_{\mathcal{O}_0}, \tag{3.41}$$

$$|M|_{\mathcal{O}} \le (1+\epsilon_0)|M^0|_{\mathcal{O}_0}.$$

Define ρ_* similar to ρ_0 with M, $\mathcal{O}$, $\sigma_* = (1-\epsilon_0)\sigma_0$ in place of M^0, $\mathcal{O}_0$, σ_0 respectively. Then, $\rho_* \le \frac{4}{3}\rho_0$ as ϵ_0 small, and, as μ_* small the real parts of all eigenvalues of $[M]J$ are bounded below in absolute value by σ_0.

On one hand, by a similar argument as for (3.34), we have that

$$|\Lambda^{-1}|_{\mathcal{O}} \le \frac{1}{2}\rho_* \le \frac{2}{3}\rho_0.$$

On the other hand, by (3.39) and H4), we can make μ_* even smaller if necessary so that

$$|\mathcal{A}|_{\mathcal{O}} < \frac{2}{3\rho_0}.$$

It follows that $\mathcal{L}$ is non-singular on $\mathcal{O}$ and

$$|\mathcal{L}^{-1}|_{\mathcal{O}} \le 4\rho_0.$$

Since, by H3) and H4),

$$|\partial_\lambda \mathcal{L}|_{\mathcal{O}} \le |\partial_\lambda \Lambda|_{\mathcal{O}} + |\partial_\lambda \mathcal{A}|_{\mathcal{O}} \le cK_+,$$

we also have

$$|\partial_\lambda \mathcal{L}^{-1}|_{\mathcal{O}} = |\mathcal{L}^{-1}(\partial_\lambda \mathcal{L})\mathcal{L}^{-1}|_{\mathcal{O}} \le |\partial_\lambda \mathcal{L}|_{\mathcal{O}} |\mathcal{L}^{-1}|^2_{\mathcal{O}} \le cK_+.$$

By induction,

$$|\partial^l_\lambda \mathcal{L}^{-1}|_{\mathcal{O}} \le cK_+, \ |l| \le n.$$

□

Above all, with the hypotheses H3), H4), the linear system (3.31) can be uniquely solved on $\mathcal{O}_+$ to yield smooth functions $f_{k0}, f_{k1}, F_{k1}, F_{01}, y_*$, $0 < |k| \le K_+$. Hence, the Hamiltonian F in (3.4) is well defined, smooth in $\lambda \in \mathcal{O}_+$, and real analytic in $(x, y, z) \in D$.

3.4. **Estimates on the transformation.** We first give some estimates on F and its derivatives. Denote

$$\zeta = K_+\Gamma(r - r_+).$$

Lemma 3.3. *Assume* H3), H4) *and also that*

H5) $|\partial^l_\omega(\Omega - \omega)|_{\mathcal{O}} \le \mu_*^{\frac{1}{4}}, \ |l| \le n.$

Then there is a constant c_3 such that the following holds for all $|l| \le n$.

1) $|\partial^l_\lambda y_*|_{\mathcal{O}_+} \le c_3\gamma^{n+1}s\mu\zeta.$
2) *On* $D(s) \times \mathcal{O}_+$,
$$|\partial^l_\lambda F|, \ |\partial^l_\lambda F_x|, \ s|\partial^l_\lambda F_y|, \ s|\partial^l_\lambda F_z| \le c_3 s^2\mu\zeta.$$
3) *On* $D(\beta) \times \mathcal{O}_+$,
$$|\partial^l_\lambda D^i F| \le c_3\mu\zeta, \ |i| \le 4,$$
where $D = \partial_{(x,y,z)}$.

Proof. Let $|l| \le n$. First, we observe by Cauchy's estimate that

$$\begin{aligned} |\partial^l_\lambda P_{kij}|_{\mathcal{O}} &\le cs^{-(i+j)}|\partial^l_\lambda P|_{D(r,s)\times\mathcal{O}} e^{-|k|r} \\ &\le c\gamma^{n+1}s^{2-i-j}\mu e^{-|k|r}, \ |k| \ge 0, \ i, j = 0, 1, 2, \end{aligned} \tag{3.42}$$

and by (3.35) and H3) that

$$\begin{aligned} |\partial^l_\lambda M_k|_{\mathcal{O}}, \ |\partial^l_\lambda B_k|_{\mathcal{O}} &\le |\partial^l_\lambda \mathcal{M} - \partial^l_\lambda \mathcal{M}^0|_{\mathcal{D}(r)\times\mathcal{O}} e^{-|k|r} \\ &\le \gamma_0^{n+1}\mu_*^{\frac{1}{4}} e^{-|k|r} \le 2^{n+1}\gamma^{n+1}\mu_*^{\frac{1}{4}} e^{-|k|r}, \ |k| > 0. \end{aligned} \tag{3.43}$$

Let $\lambda \in \mathcal{O}_+$, $|k| > 0$. Then, by (3.29), (3.40)-(3.43), we have that

$$|\partial^l_\lambda \mathcal{P}_k| \le c\gamma^{n+1}s\mu e^{-|k|r},$$

i.e.,

$$|\partial^l_\lambda \mathcal{P}| \le c\gamma^{n+1}s\mu.$$

Using (3.31) and Lemma 3.2, it follows that

$$|\partial^l_\lambda \tilde{F}_{k1}| \le |\partial^l_\lambda \mathcal{F}| = |\partial^l_\lambda(\mathcal{L}^{-1}\mathcal{P})| \le c\gamma^{n+1}s\mu K_+.$$

Hence,

$$|\partial^l_\lambda F_{k1}| \le c\gamma^{n+1}s\mu K_+ e^{-|k|(r_+ + \frac{7}{8}(r - r_+))}. \tag{3.44}$$

By straightforward applications of (3.40)-(3.44) to (3.24) and (3.25), we then obtain 1) and also that

$$|\partial_\lambda^l F_{01}| \le c\gamma^{n+1} s\mu K_+. \tag{3.45}$$

Next, we note by H5), (3.21) and a direct calculation that

$$|\partial_\lambda^l \langle k, \Omega(\lambda)\rangle^{-1}| \le c\frac{|k|^{|l|+(|k|+1)\tau}}{\gamma^{|l|+1}}.$$

This together with (3.40)-(3.44) implies that

$$|\partial_\lambda^l f_{k0}| \le c|k|^{|l|+(|l|+1)\tau} s^2\mu e^{-|k|r}, \tag{3.46}$$

$$|\partial_\lambda^l f_{k1}| \le c|k|^{|l|+(|l|+1)\tau} s\mu K_+ e^{-|k|(r_+ + \frac{7}{8}(r-r_+))}. \tag{3.47}$$

Now, using the expression of F, y_* in (3.4) and (3.25) respectively, parts 2), 3) of the lemma follow directly from (3.43)-(3.47). □

Lemma 3.4. *Assume* H1), H3)-H5) *and also that*

- **H6)** $c_3\mu\zeta < \frac{1}{8}(r - r_+)$;
- **H7)** $c_3 s\mu\zeta < s_+$;
- **H8)** $c_3\mu\zeta < \frac{\beta-\beta_+}{2}$.

Let ϕ_F^t be the flow generated by F. Then the following holds.

1) *For all $0 \le t \le 1$,*

$$\begin{aligned} \phi_F^t : D_3 &\to D_4, \\ \phi : D_1 &\to D_3 \end{aligned}$$

are well defined, real analytic and depend smoothly on $\lambda \in \mathcal{O}_+$.

2) *Let $\Phi_+ = \phi_F^1 \circ \phi$. Then for all $\lambda \in \mathcal{O}_+$,*

$$\Phi_+ : \begin{array}{l} D_+ \to D, \\ \tilde{D}_+ \to D(r,\beta). \end{array}$$

3) *There is a constant c_4 such that*

$$\begin{aligned} |\partial_\lambda^l(\phi_F^t - id)|_{D(s)\times\mathcal{O}_+} &\le c_4 s\mu\zeta, \\ |\partial_\lambda^l D^i(\Phi_+ - id)|_{\tilde{D}_+\times\mathcal{O}_+} &\le c_4\mu\zeta, \end{aligned}$$

for all $|l| \le n,\ 0 \le i \le 3,\ 0 \le t \le 1$, where $D = \partial_{(x,y,z)}$.

Proof. Let $\lambda \in \mathcal{O}_+$.

1) We note that

$$\phi_F^t = \mathrm{id} + \int_0^t X_F \circ \phi_F^\xi \mathrm{d}\xi, \tag{3.48}$$

where $X_F = (F_y, -F_x, JF_z)^\top$.

Denote $\phi_{F1}^t, \phi_{F2}^t, \phi_{F3}^t$ as components of ϕ_F^t in x, y, z planes respectively. For any $(x,y,z) \in D_3$, let $t_* = \sup\{t \in [0,1] : \phi_F^t(x,y,z) \in D_4$. By H1), we have that

$D_4 \subset D(s)$. It follows from H6), H7) and Lemma 3.3 that

$$\begin{aligned}
|\phi^t_{F1}(x,y,z)| &= |x| + |\int_0^t F_y \circ \phi_F^\xi \mathrm{d}\xi| \le |x| + |F_y|_{D(s)} \le r_+ + \frac{2}{8}(r-r_+) + c_3 s\mu\zeta \\
&< r_+ + \frac{3}{8}(r-r_+), \\
|\phi^t_{F2}(x,y,z)| &= |y| + |-\int_0^t F_x \circ \phi_F^\xi \mathrm{d}\xi| \le |y| + |F_x|_{D(s)} \le 3s_+ + c_3 s^2\mu\zeta < 4s_+, \\
|\phi^t_{F3}(x,y,z)| &= |z| + |\int_0^t JF_z \circ \phi_F^\xi \mathrm{d}\xi| \le |z| + |F_z|_{D(s)} \le 3s_+ + c_3 s\mu\zeta < 4s_+,
\end{aligned}$$

i.e., $\phi^t_F(x,y,z) \in D_4$ for all $0 \le t \le t_*$. Thus, $t_* = 1$ and 1) holds.

2) follows from Lemma 3.3, H8) and a similar argument as 1).

3) Using Lemma 3.3 and (3.48), we immediately have

$$|\phi^t_F - id|_{D(s)} \le c_3 s\mu\zeta.$$

Differentiating (3.48) with respect to λ yields

$$\partial_\lambda \phi^t_F = \int_0^t DX_F \circ \phi_F^\xi \partial_\lambda \phi_F^\xi \mathrm{d}\xi + \int_0^t (\partial_\lambda F_y, -\partial_\lambda F_x, J\partial_\lambda F_z)^\top \circ \phi_F^\xi \mathrm{d}\xi.$$

It follows from Lemma 3.3 and Gronwall's inequality that

$$|\partial_\lambda \phi^t_F|_{D(s)} \le cs\mu\zeta.$$

By induction, we have

$$|\partial^l_\lambda \phi^t_F|_{D(s)} \le cs\mu\zeta, \ |l| \le n.$$

The estimates for Φ_+ follow from a similar application of Lemma 3.3 and Gronwall's inequality, and the identity

$$\Phi_+ - id = (\phi^1_F - id) \circ \phi + \begin{pmatrix} 0 \\ y_* \\ 0 \end{pmatrix}.$$

We omit the details. $\square$

3.5. **Estimates on the new Hamiltonian.** We first estimate the new normal form.

Lemma 3.5. *Assume* H3), H5)*. Then there is a constant* c_5 *such that the following holds for all* $|l| \le n$*:*

$$\begin{aligned}
|\partial^l_\lambda(e_+ - e)|_{\mathcal{O}_+} &\le c_5\gamma^{n+1}s\mu\zeta, \\
|\partial^l_\lambda(\Omega_+ - \Omega)|_{\mathcal{O}_+} &\le c_5\gamma^{n+1}s\mu\zeta, \\
|\partial^l_\lambda(\mathcal{M}^+ - \mathcal{M})|_{\mathcal{D}_+\times\mathcal{O}_+} &\le c_5\gamma^{n+1}\mu\Gamma(r-r_+).
\end{aligned}$$

Proof. The estimates for $|\partial^l_\lambda(e^+ - e)|_{\mathcal{O}_+}$ follows immediately from (3.9), (3.42), H5) and Lemma 3.3 1). Also, it follows from (3.10), (3.42)-(3.44) and Lemma 3.3 1) that

$$|\partial^l_\lambda(\Omega_+ - \Omega)|_{\mathcal{O}_+} \le c\gamma^{n+1}s\mu\zeta + c\gamma^{n+1}s\mu K_+ \sum_{|k|>0} e^{-\frac{|k|r_0}{2}} \le c\gamma^{n+1}s\mu\zeta,$$

and from (3.11)-(3.13) and (3.42) that

$$|\partial_\lambda^l(\mathcal{M}^+ - \mathcal{M})|_{\mathcal{D}(r_+ + \frac{1}{8}(r-r_+))\times\mathcal{O}_+} \le c\sum_{|k|>0}|k|\sum_{i+j=2}|\partial_\lambda^l P_{kij}|_{\mathcal{O}} e^{|k|(r_+ + \frac{7}{8}(r-r_+))}$$

$$\le c\gamma^{n+1}\mu\sum_{|k|>0}|k|e^{-|k|\frac{r-r_+}{8}} \le c\gamma^{n+1}\mu\Gamma(r-r_+),\ |l|\le n. \tag{3.49}$$

This proves the lemma. $\square$

Lemma 3.6. *Assume that*

H9) $c_5\gamma^{n+1}s\mu\zeta K_+^{\tau+1} < \gamma - \gamma_+$.

Then

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

for all $\lambda \in \mathcal{O}_+$ *and* $0 < |k| \le K_+$.

Proof. By H9) and Lemma 3.5, we have

$$\begin{aligned} |\langle k, \Omega_+(\lambda)\rangle| &\ge |\langle k, \Omega(\lambda)| - c_5\gamma^{n+1}s\mu\zeta K_+^{\tau+1} \\ &\ge \frac{\gamma}{|k|^\tau} - c_5\gamma^{n+1}s\mu\zeta K_+^{\tau+1} > \frac{\gamma_+}{|k|^\tau}, \end{aligned}$$

as desired. $\square$

Let U^+ be the $n_0\times n_0$ ordered principal block of $[A^+]$. By (3.49) and H7), we see that

$$\begin{aligned} |U^+ - U|_{\mathcal{O}_+} &\le c\gamma^{n+1}\mu\Gamma(r-r_+) \le c\mu\zeta \le \frac{c}{c_3}\frac{s_+}{s} \\ &= \frac{c}{c_3}s^{b+\sigma} \le \frac{c}{c_3}s_0^{b+\sigma} \le \frac{c}{c_3}\mu_*^{\frac{b+\sigma}{4}}. \end{aligned}$$

It follows that U^+ is non-singular on $\mathcal{O}_+$ as long as μ_* is small. A similar argument shows the same for

$$Y^+ = [M^+]J - [B^+]^\top \operatorname{diag}((U^+)^{-1}, O)[B^+]J.$$

Now, let M_k^+, B_k^+ denote the Fourier coefficients of M^+, B^+, respectively, and define M_{kj}^+, $|k|, |j| \ne 0$, as in (3.28) with M_k^+, B_k^+, U^+, Y^+ in place of M_k, B_k, U, Y, for all $|k|\ne 0$, respectively. We have the following.

Lemma 3.7. *Assume* H3), H5) and H7). *Then there is a constant* c_6 *such that*

$$\max_{|k|>0}\sum_{|j|>0} e^{(|k|-|j|)\xi(r_+)}|\partial_\lambda^l M_{kj}^+ - \partial_\lambda^l M_{kj}|_{\mathcal{O}_+} \le c_6\gamma^{n+1}\mu\Gamma^2(r-r_+),\ |l|\le n,$$

where ξ *is as in* (3.38).

Proof. First of all, we note by H3) and (3.49) that M_k^+, B_k^+, M_k, B_k, $(U^+)^{-1}$, $(Y^+)^{-1}$, U^{-1}, Y^{-1} are uniformly bounded on $\mathcal{O}_+$ by a constant which is independent of the iteration process. Secondly, by (3.49) and Cauchy's estimate, we have that

$$\begin{aligned}
|\partial_\lambda^l M_j^+ - \partial_\lambda^l M_j|_{\mathcal{O}_+} &\le |\partial_\lambda^l M^+ - \partial_\lambda^l M|_{\mathcal{D}(r_+ + \frac{1}{8}(r-r_+))\times\mathcal{O}_+} e^{-|j|(r_+ + \frac{7}{8}(r-r_+))} \\
&\le |\partial_\lambda^l \mathcal{M}^+ - \partial_\lambda^l \mathcal{M})|_{\mathcal{D}(r_+ + \frac{1}{8}(r-r_+))\times\mathcal{O}_+} e^{-|j|(r_+ + \frac{7}{8}(r-r_+))} \\
&\le c\gamma^{n+1}\mu\Gamma(r-r_+)e^{-|j|\xi(r)}, \\
|\partial_\lambda^l B_j^+ - \partial_\lambda^l B_j|_{\mathcal{O}_+} &\le |\partial_\lambda^l B^+ - \partial_\lambda^l B|_{\mathcal{D}(r_+ + \frac{1}{8}(r-r_+))\times\mathcal{O}_+} e^{-|j|(r_+ + \frac{7}{8}(r-r_+))} \\
&\le |\partial_\lambda^l \mathcal{M}^+ - \partial_\lambda^l \mathcal{M})|_{\mathcal{D}(r_+ + \frac{1}{8}(r-r_+))\times\mathcal{O}_+} e^{-|j|(r_+ + \frac{7}{8}(r-r_+))} \\
&\le c\gamma^{n+1}\mu\Gamma(r-r_+)e^{-|j|\xi(r)}.
\end{aligned}$$

It follows from (3.28) and a similar formular for M_{kj}^+ that

$$|\partial_\lambda^l M_{kj}^+ - \partial_\lambda^l M_{kj}|_{\mathcal{O}_+} \le c\gamma^{n+1}\mu\Gamma(r-r_+)e^{-|k-j|\xi(r)}$$

for all $|k|,\ |j| > 0$.

Hence,

$$\begin{aligned}
&\sum_{|j|>0} e^{(|k|-|j|)\xi(r_+)}|\partial_\lambda^l M_{kj}^+ - \partial_\lambda^l M_{kj}|_{\mathcal{O}_+} \le \sum_{|j|>0} e^{(|k-j|)\xi(r_+)}|\partial_\lambda^l M_{kj}^+ - \partial_\lambda^l M_{kj}|_{\mathcal{O}_+} \\
&\quad \le c\gamma^{n+1}\mu\Gamma(r-r_+)\sum_{|j|>0} e^{-|k-j|(\xi(r)-\xi(r_+))} \\
&\quad \le c\gamma^{n+1}\mu\Gamma(r-r_+)\sum_{|j|>0} e^{-|k-j|\frac{r-r_+}{8}} \le c\gamma^{n+1}\mu\Gamma^2(r-r_+).
\end{aligned}$$

□

Let

$$\Delta = \frac{\zeta^3}{r-r_+}(s^3\mu + \gamma^{n+1}s^2\mu^2 + \gamma^{n+1}\frac{s_+^3}{s}\mu). \tag{3.50}$$

Lemma 3.8. *Assume* H2)-H7). *Then there is a constant* c_7 *such that*

$$|\partial_\lambda^l P_+|_{D_+} \le c_7\Delta,\ |l| \le n.$$

Thus, if

H10) $c_7\Delta \le \gamma_+^{n+1}s_+^2\mu_+$, *and* $2s\mu K_+^{\tau+1} < \gamma - \gamma_+$,

then

$$|\partial_\lambda^l P_+|_{D_+} \le \gamma_+^{n+1}s_+^2\mu_+. \tag{3.51}$$

Proof. Let $|l| \le n$, $\lambda \in \mathcal{O}_+$. By Lemma 3.1 and Cauchy's estimate, we have

$$|\partial_\lambda^l D(P-R)|_{D_4} \le \frac{c}{(r-r_+)s_+}|\partial_\lambda^l (P-R)|_{D_8} \le \frac{c\gamma^{n+1}\mu}{r-r_+}(\frac{s^3}{s_+} + \frac{s_+^2}{s}),$$

where $D = \partial_{(x,y,z)}$. This together with Lemma 3.3 1) implies that

$$\begin{aligned}
|\partial_\lambda^l (P-R)\circ\Phi_+|_{D_+} &\le c\frac{\gamma^{n+1}\mu^2}{r-r_+}(\frac{s^4}{s_+} + s_+^2)\zeta + c\mu(s^3 + \frac{s_+^3}{s}) \\
&\le c\frac{\gamma^{n+1}}{r-r_+}(s^2\mu^2 + \frac{s_+^3}{s}\mu)\zeta.
\end{aligned} \tag{3.52}$$

Also, a direct estimate using (3.42) yields that

$$
\begin{aligned}
|\partial_\lambda^l D^\alpha R|_{D(s)} &\le \sum_{0<|k|\le K_+, 0\le i,j\le 2} |\partial_\lambda^l P_{kij}| s^{i+j-1}|k|e^{|k|(r_++\frac{6}{8}(r-r_+)} \\
&\le c\gamma^{n+1}s\mu \sum_{0<|k|\le K_+} |k|e^{-|k|\frac{r-r_+}{4}} \le c\gamma^{n+1}s\mu\Gamma(r-r_+),\ \alpha=0,1. \qquad (3.53)
\end{aligned}
$$

Similarly, by (3.5), (3.42)-(3.44),

$$
|\partial_\lambda^l D^\alpha R'|_{D(s)} \le c\gamma^{n+1}s\mu K_+\Gamma(r-r_+) = c\gamma^{n+1}s\mu\zeta,\ \alpha=0,1. \qquad (3.54)
$$

We now estimate Q in (3.15). Let

$$
\begin{aligned}
Q_1 &= \sum_{0<|k|\le K_+} ((\langle\frac{1}{2}\partial_x\langle y, A(x,\lambda)y\rangle + \partial_x\langle y, B(x,\lambda)z\rangle \\
&+\frac{1}{2}\partial_x\langle z, M(x,\lambda)z\rangle + \partial_x h_0(x,y,z,\lambda), f_{k1}\rangle \\
&+(-\sqrt{-1}\langle k, A(x,\lambda)y + B(x,\lambda)z\rangle + \partial_y h_0(x,y,z,\lambda))(f_{k0} + \langle f_{k1}, y\rangle + \langle F_{k1}, z\rangle) \\
&+\langle\partial_z h_0(x,y,z,\lambda)JF_{k1}, z\rangle)e^{\sqrt{-1}\langle k,x\rangle}, \\
Q_2 &= \sum_{|k|>K_+} ((\langle B_k JF_{01}, y\rangle + \langle M_k JF_{01}, z\rangle)e^{\sqrt{-1}\langle k,x\rangle}, \\
Q_3 &= \sum_{|k|>K_+, |j|\le K_+} ((\langle B_{k-j}JF_{j1}, y\rangle + \langle M_{k-j}JF_{j1}, z\rangle)e^{\sqrt{-1}\langle k,x\rangle}, \\
q_0 &= \langle B^\top(x)y_*, JF_{01}\rangle, \\
q_1 &= -\sum_{0<|k|\le K_+} \sqrt{-1}\langle f_{k1}, y_*\rangle + \langle B^\top(x)y_*, JF_{k1}\rangle + \langle P_{k10}, y_*\rangle)e^{\sqrt{-1}\langle k,x\rangle}, \\
q_2 &= \sum_{|k|>K_+} ((\langle y_*, A_k y\rangle + \langle y_*, B_k z\rangle)e^{\sqrt{-1}\langle k,x\rangle}, \\
q &= h_0(x, y+y_*, z, \lambda) - h_0(x,y,z,\lambda).
\end{aligned}
$$

Then

$$
Q = (Q_1+Q_2+Q_3)\circ\phi + q_0 + q_1 + q_2.
$$

By H3), it is clear that

$$
|\partial_\lambda^l A|_{\mathcal{D}(r)},\ |\partial_\lambda^l B|_{\mathcal{D}(r)},\ |\partial_\lambda^l M|_{\mathcal{D}(r)},\ s^{-3}|\partial_\lambda^l\partial_x h_0|_{\mathcal{D}(r)},\ s^{-2}|\partial_\lambda^l\partial_{(y,z)}h_0|_{\mathcal{D}(r)},\ |U|\le c.
$$

It follows from the above and (3.44)-(3.47) that

$$
\begin{aligned}
|\partial_\lambda^l Q_1|_{D_3} &\le c\sum_{0<|k|\le K_+} (s_+^2|\partial_\lambda^l f_{k1}| + s_+|k|(|\partial_\lambda^l f_{k0}| + s_+|\partial_\lambda^l f_{k1}| \\
&+ s_+|\partial_\lambda^l F_{k1}|)e^{|k|(r_++\frac{1}{4}(r-r_+))} \\
&\le cs^2s_+\mu K_+ \sum_{0<|k|\le K_+} |k|^{n+(n+1)\tau+6}e^{-|k|\frac{5(r-r_+)}{8}} \\
&\le cs^2s_+\mu K_+\Gamma(r-r_+) = cs^2s_+\mu\zeta
\end{aligned}
$$

and from (3.42), (3.47) and Lemma 3.3 1) that

$$\begin{aligned}
|\partial_\lambda^l q_0|_{D_1} &\le c\sum_{0<|k|\le K_+}(|\partial_\lambda^l y_*||F_{k01}|+|y_*||\partial_\lambda^l F_{k01}|)\le c\gamma^{n+1}s^2\mu^2K_+\zeta,\\
|\partial_\lambda^l q_1|_{D_1} &\le \sum_{0<|k|\le K_+}(|\partial_\lambda^l y_*|(|f_{k1}|+|F_{k1}|+|P_{k10}|)\\
&\quad +|y_*|(|\partial_\lambda^l f_{k1}|+|\partial_\lambda^l F_{k1}|+|\partial_\lambda^l P_{k10}|))e^{|k|(r_++\frac14(r-r_+))}\\
&\le c\gamma^{n+1}s^2\mu^2\zeta K_+\sum_{0<|k|\le K_+}|k|^{1+2\tau}e^{-|k|\frac{3(r-r_+)}{4}}\le c\gamma^{n+1}s^2\mu^2\zeta^2.
\end{aligned}$$

Similarly,

$$\begin{aligned}
|\partial_\lambda^l Q_1|_{D(s)} &\le cs^3\mu\zeta,\\
|\partial_\lambda^l q_0|_{D(s)} &\le c\gamma^{n+1}s^2\mu^2K_+\zeta,\\
|\partial_\lambda^l q_1|_{D(s)} &\le cs^2\mu^2\zeta^2,\\
|\partial_\lambda^l q|_{D(s)} &\le c\gamma^{n+1}s_+s^3\mu\zeta.
\end{aligned}\tag{3.55}$$

By (3.43)-(3.45) and H2), we also have

$$\begin{aligned}
|\partial_\lambda^l Q_2|_{D_3} &\le c\sum_{|k|>K_+}s_+((|\partial_\lambda^l B_k|+|\partial_\lambda^l M_k|)|F_{01}|\\
&\quad +(|B_k|+|M_k|)|\partial_\lambda^l F_{01}|)e^{|k|(r_++\frac14(r-r_+))}\\
&\le c\gamma^{n+1}ss_+\mu K_+\sum_{|k|>K_+}e^{-|k|\frac{3(r-r_+)}{4}}\le c\gamma^{n+1}ss_+\mu^2K_+,\\
|\partial_\lambda^l Q_3|_{D_3} &\le c\sum_{|k|>K_+,0<|j|\le K_+}s_+((|\partial_\lambda^l B_{k-j}|+|\partial_\lambda^l M_{k-j}|)|F_{j1}|\\
&\quad +(|B_{k-j}|+|M_{k-j}|)|\partial_\lambda^l F_{j1}|)e^{|k|(r_++\frac14(r-r_+))}\\
&\le c\gamma^{n+1}ss_+\mu K_+(\sum_{|k|>K_+}e^{-|k|\frac{(r-r_+)}{2}})(\sum_{0<|j|\le K_+}e^{-|j|\frac{(r-r_+)}{8}})\\
&\le c\gamma^{n+1}ss_+\mu^2K_+\Gamma(r-r_+)=c\gamma^{n+1}ss_+\mu^2\zeta.
\end{aligned}$$

Similarly,

$$\begin{aligned}
|\partial_\lambda^l Q_2|_{D(s)} &\le c\gamma^{n+1}s^2\mu^2K_+,\\
|\partial_\lambda^l Q_3|_{D(s)} &\le c\gamma^{n+1}s^2\mu^2\zeta.
\end{aligned}$$

Note that, similar to (3.43),

$$|\partial_\lambda^l A_k|_{\mathcal O}\le|\partial_\lambda^l\mathcal M-\partial_\lambda^l\mathcal M_0|_{\mathcal D(r)\times\mathcal O_0}e^{-|k|r}\le\gamma_0^{n+1}\mu_*^{\frac14}e^{-|k|r}.$$

This together with (3.43), H2) and Lemma 3.3 1) implies that

$$\begin{aligned}
|\partial_\lambda^l q_2|_{D_3} &\le c\sum_{|k|>K_+}s_+((|\partial_\lambda^l A_k|+|\partial_\lambda^l B_k|)|y_*|\\
&+\ (|A_k|+|B_k|)|\partial_\lambda^l y_*|)e^{|k|(r_++\frac14(r-r_+))}\\
&\le c\gamma_0^{n+1}s_+s\mu K_+\sum_{|k|>K_+}e^{-|k|\frac{3(r-r_+)}{4}}\le 2^{n+1}c\gamma^{n+1}s_+s\mu^2\zeta.
\end{aligned}$$

Similarly,

$$|\partial_\lambda^l q_2|_{D(s)} \le c\gamma^{n+1}s^2\mu^2\zeta.$$

It now follows from Lemma 3.3 1) that

$$|\partial_\lambda^l Q|_{D_+} \le c(s^2 s_+\mu + \gamma^{n+1}s^2\mu^2)\zeta^2, \tag{3.56}$$

$$|\partial_\lambda^l Q\circ\phi^{-1}|_{D(s)} \le c(s^3\mu + \gamma^{n+1}s^2\mu^2)\zeta^2. \tag{3.57}$$

Applying Cauchy's estimate to (3.57), we then have

$$\begin{aligned} |\partial_\lambda^l D(Q\circ\phi^{-1})|_{D_3} &\le \frac{c}{(r-r_+)(s-3s_+)}|\partial_\lambda^l (Q\circ\phi^{-1})|_{D(s)} \\ &\le \frac{c}{r-r_+}(s^2\mu + \gamma^{n+1}s\mu^2)\zeta^2. \end{aligned} \tag{3.58}$$

Denote

$$\begin{aligned} W_0 &= \int_0^1 \{R_t, F\}\circ\phi_F^t \mathrm{d}t, \\ W_1 &= \frac{1}{2}\langle y_*, (A - 2\mathrm{diag}(U,O))y_*\rangle \\ &+ \sum_{|k|\le K_+}(\langle y_*, P_{k20}y_*\rangle + \langle y_*, 2P_{k20}y\rangle + \langle y_*, P_{k11}z\rangle)e^{\sqrt{-1}\langle k,x\rangle}. \end{aligned}$$

Then by (3.14),

$$P_+ = W_0\circ\phi + W_1 + Q + q + (P-R)\circ\Phi_+. \tag{3.59}$$

Since by (3.6) and (3.7)

$$\begin{aligned} R_t &= (1-t)\{N,F\} + R \\ &= (1-t)((Q - \langle y_*, (A-[A])y\rangle - \langle y_*, Bz\rangle)\circ\phi^{-1} - (R-R')) + R, \end{aligned}$$

we have

$$W_0 = \int_0^1 \{(1-t)((Q - \langle y_*, (A-[A])y\rangle - \langle y_*, Bz\rangle)\circ\phi^{-1} - R + R') + R, F\}\circ\phi_F^t \mathrm{d}t.$$

Using Lemma 3.3 1), 2) and Lemma 3.4 3), we have by (3.58) that

$$|\partial_\lambda^l \int_0^1 \{Q\circ\phi^{-1}, F\}\circ\phi_F^t \mathrm{d}t\circ\phi|_{D_+} \le \frac{c}{r-r_+}(s^3\mu^2 + \gamma^{n+1}s^2\mu^3)\zeta^3, \tag{3.60}$$

and, by (3.53) and (3.54) that

$$(|\partial_\lambda^l \int_0^1 \{R, F\}\circ\phi_F^t \mathrm{d}t\circ\phi| + |\partial_\lambda^l \int_0^1 \{R', F\}\circ\phi_F^t \mathrm{d}t\circ\phi|)_{D_+} \le c\gamma^{n+1}s^2\mu^2\zeta^3. \tag{3.61}$$

By Lemma 3.3, it is also clear that

$$|\partial_\lambda^l \int_0^1 \{(\langle y_*, (A-[A])y\rangle + \langle y_*, Bz\rangle)\circ\phi^{-1}, F\}\circ\phi_F^t \mathrm{d}t\circ\phi|_{D_+} \le c\gamma^{n+1}s^2\mu^2\zeta^3.$$

This together with (3.60) and (3.61) yields that

$$|\partial_\lambda^l W_0\circ\phi|_{D_+} \le c\frac{\mu^2}{r-r_+}(s^3 + \gamma^{n+1}s^2)\zeta^3. \tag{3.62}$$

Using the above arguments along with (3.42) and Lemma 3.3, we also have

$$|\partial_\lambda^l W_1|_{D_+} \le c\gamma^{n+1}(s_+ + s)s\mu^2\zeta^2. \tag{3.63}$$

Above all, it follows from (3.52), (3.55), (3.56), (3.59), (3.62), (3.63) that

$$|\partial_\lambda^l P_+|_{D_+} \le c\frac{\zeta^3}{r-r_+}(s_+ s^2\mu + \gamma^{n+1}s^2\mu^2 + s^3\mu^2 + \gamma^{n+1}\frac{s_+^3}{s}\mu) \le c\Delta.$$

□

4. Iteration Lemma

Let $r_0, s_0, \mu_0, \gamma_0, \mathcal{O}_0, H_0, N_0, e_0, \Omega_0, \mathcal{M}^0, A^0, B^0, M^0, \mathcal{A}^0, h_0, P_0$ be defined as in Section 3 and let $\tilde{D}_0 = D(r_0, \beta_0)$, $D_0 = D(r_0, s_0)$, $\mathcal{D}_0 = \{x : |\mathrm{Im}x| < r_0\}$, $K_0 = 0$, $\Phi_0 = id$. For $\nu = 1, 2\cdots$, we index all index-free quantities in Section 3 by ν and index all "$+$"-indexed quantities in Section 3 by $\nu+1$. This yields the following sequences with the properties stated in Section 3:

$$\begin{aligned}
H_\nu &= H_\nu(x,y,z,\lambda) = N_\nu + P_\nu,\\
N_\nu &= e_\nu + \langle \Omega_\nu, y\rangle + \frac{1}{2}\langle \begin{pmatrix} y\\ z\end{pmatrix}, \mathcal{M}^\nu \begin{pmatrix} y\\ z\end{pmatrix}\rangle + h_0(x,y,z,\lambda),\\
\mathcal{M}^\nu &= \begin{pmatrix} A^\nu & B^\nu\\ (B^\nu)^\top & M^\nu\end{pmatrix},\\
r_\nu &= r_0(1 - \frac{1}{2}(1-\delta)\sum_{i=1}^{\nu}\delta^{i+1}),\\
s_\nu &= s_{\nu-1}^{1+b+\sigma},\\
\beta_\nu &= \beta_0(1 - \sum_{i=1}^{\nu}\frac{1}{2^{i+1}}),\\
\gamma_\nu &= \gamma_0(1 - \sum_{i=0}^{\nu-1}\frac{1}{2^{i+2}}),\\
\mu_\nu &= c_0 s_{\nu-1}^{\sigma}\mu_{\nu-1},\ c_0 = \max\{1, c_1, \cdots, c_7\},\\
K_\nu &= ([\log\frac{1}{s_{\nu-1}}] + 1)^3,\ \nu \ge 1,\\
\mathcal{O}_\nu &= \{\lambda \in \mathcal{O}_{\nu-1} : |\langle k, \Omega_{\nu-1}(\lambda)\rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau}, 0 < |k| \le K_\nu\},\ \nu \ge 1,\\
D_\nu &= D(r_\nu, s_\nu),\\
\tilde{D}_\nu &= D(r_\nu + \frac{7}{8}(r_{\nu-1} - r_\nu), \beta_\nu),\\
\mathcal{D}_\nu &= \{x : |\mathrm{Im}x| < r_\nu\},
\end{aligned}$$

$\nu = 0, 1, \cdots$.

We now prove an iteration lemma which checks the validity of all KAM steps.

Lemma 4.1. (**Iteration Lemma**) *If $\mu_* = \mu_*(r_*, \beta_0, \sigma_0, U^0)$ is sufficiently small, then the following holds for all $|l| \le n; \nu = 1, 2, \cdots$.*

1)

$$|\partial_\lambda^l (e_\nu - e_0)|_{\mathcal{O}_\nu} \le \gamma_0^{2(n+1)} \mu_*^{\frac{1}{2}}, \tag{4.1}$$

$$|\partial_\lambda^l (e_\nu - e_{\nu-1})|_{\mathcal{O}_\nu} \le \frac{\gamma_0^{2(n+1)} \mu_*^{\frac{1}{2}}}{2^\nu}, \tag{4.2}$$

$$|\partial_\lambda^l (\Omega_\nu - \Omega_0)|_{\mathcal{O}_\nu} \le \gamma_0^{2(n+1)} \mu_*^{\frac{1}{2}}, \tag{4.3}$$

$$|\partial_\lambda^l (\Omega_\nu - \Omega_{\nu-1})|_{\mathcal{O}_\nu} \le \frac{\gamma_0^{2(n+1)} \mu_*^{\frac{1}{2}}}{2^\nu}, \tag{4.4}$$

$$|\partial_\lambda^l (\mathcal{M}^\nu - \mathcal{M}^0)|_{\mathcal{D}_\nu \times \mathcal{O}_\nu} \le \gamma_0^{n+1} \mu_*^{\frac{1}{4}}, \tag{4.5}$$

$$|\partial_\lambda^l (\mathcal{M}^\nu - \mathcal{M}^{\nu-1})|_{\mathcal{D}_\nu \times \mathcal{O}_\nu} \le \frac{\gamma_0^{n+1} \mu_*^{\frac{1}{4}}}{2^\nu}, \tag{4.6}$$

$$|\partial_\lambda^l P_\nu|_{D_\nu \times \mathcal{O}_\nu} \le \gamma_\nu^{n+1} s_\nu^2 \mu_\nu. \tag{4.7}$$

2) $(\Omega_\nu(\lambda))_i = \omega_i(\lambda),\ i = 1, 2, \cdots, n_0$.

3) *There is a transformation* $\Phi_\nu : \tilde{D}_\nu \times \mathcal{O}_\nu \longrightarrow \tilde{D}_{\nu-1},\ D_\nu \times \mathcal{O}_\nu \longrightarrow D_{\nu-1}$, *which is symplectic and analytic in* $(x, y, z) \in \tilde{D}_{\nu+1}$, *and smooth in* $\lambda \in \mathcal{O}_{\nu+1}$, *such that*

$$H_\nu = H_{\nu-1} \circ \Phi_\nu$$

and

$$|\partial_\lambda^l D^i (\Phi_\nu - id)|_{\tilde{D}_\nu \times \mathcal{O}_\nu} \le \frac{\mu_*^{\frac{1}{4}}}{2^\nu},\ 0 \le i \le 3. \tag{4.8}$$

4)

$$\mathcal{O}_{\nu+1} = \{\lambda \in \mathcal{O}_\nu : |\langle k, \Omega_\nu(\lambda)\rangle| > \frac{\gamma_\nu}{|k|^\tau},\ K_\nu < |k| \le K_{\nu+1}\}.$$

Proof. The lemma will be proved by induction. We first verify the conditions H1), H2), H5)-H10) in Section 3 for all $\nu = 0, 1, \cdots$.

By definitions of μ_ν, s_ν, we clearly have

$$\mu_\nu = c_0^\nu \mu_0 s_0^{\frac{\sigma}{b+\sigma}((1+b+\sigma)^\nu - 1)}, \tag{4.9}$$

$$s_\nu = s_0^{(1+b+\sigma)^\nu}, \tag{4.10}$$

$$s_0 = \gamma^{n+1+\frac{a_0}{2}} \mu_*^{\frac{1}{4}}. \tag{4.11}$$

Define

$$\begin{aligned} E_\nu &= \frac{r_\nu - r_{\nu+1}}{8} = \frac{1}{16} r_0 (1-\delta)\delta^{\nu+2}, \\ \chi &= n + 6 + (n+1)[\tau] \end{aligned}$$

and let $\varepsilon > 0$ be fixed such that

$$\begin{aligned} &(n + 1 + a_0)(1 - 2b - 3\sigma - 2\varepsilon) > (n + 1), \\ &\sigma > (b + \sigma)(2b + 3\sigma + 2\varepsilon) + \varepsilon, \\ &3\varepsilon < b. \end{aligned}$$

By (4.9)-(4.11), it is clear that if μ_* (hence s_0) is small, then

$$\begin{aligned} c_0 \frac{\mu_\nu}{s_\nu^{2b+3\sigma+2\varepsilon}} &= \frac{c_0^{\nu+1}\mu_0 s_0^{\frac{\sigma}{b+\sigma}((1+b+\sigma)^\nu-1)}}{s_0^{(1+b+\sigma)^\nu(2b+3\sigma+2\varepsilon)}} \\ &= c_0 s_0^{-\frac{\sigma}{b+\sigma}}\mu_0 c_0^\nu s_0^{(\frac{\sigma}{b+\sigma}-(2b+3\sigma+2\varepsilon))(1+b+\sigma)^\nu} \\ &\le s_0^{-\frac{\sigma}{b+\sigma}}\mu_0(c_0 s_0^\varepsilon)^\nu \le s_0^{-\frac{\sigma}{b+\sigma}}\mu_0 \le \mu_*^{\frac{1}{4}}. \end{aligned} \tag{4.12}$$

Since

$$\begin{aligned} \Gamma_\nu &= \Gamma(r_\nu - r_{\nu+1}) = e^{E_0}\sum_{0<|k|\le K_{nu+1}} |k|^{n+(n+1)\tau+6} e^{-|k|E_\nu} \\ &\le e^{E_0}\int_1^\infty \lambda^\chi e^{-\lambda\frac{E_\nu}{2}}\mathrm{d}\lambda \le \frac{e^{E_0}2^\chi(\chi+1)!}{E_\nu^\chi}, \end{aligned}$$

we can make μ_* (hence s_0) small such that

$$c_0\frac{s_\nu^\varepsilon\Gamma_\nu^3}{E_\nu} \le c_0(e^{E_0}2^\chi(\chi+1)!)^3\frac{s_\nu^\varepsilon}{E_\nu^{3\chi+1}} \le \frac{c_0 e^{3E_0}2^{3\chi}(\chi+1)!)^3}{E_0^{3\chi+1}}\left(\frac{s_0^{(b+\sigma)\varepsilon}}{\delta}\right)^\nu s_0 \le 1.$$

On the other hand, we can also make μ_* (hence s_ν) small such that

$$s_\nu^\varepsilon K_{\nu+1}^3 = s_\nu^\varepsilon([\log\frac{1}{s_\nu}]+1)^9 \le 1.$$

Therefore,

$$c_0\frac{s_\nu^{2\varepsilon}\zeta_\nu^3}{E_\nu} = \left(\frac{c_0 s_\nu^\varepsilon\Gamma_\nu^3}{E_\nu}\right)(s_\nu^\varepsilon K_{\nu+1}^3) \le 1. \tag{4.13}$$

By (4.9) and (4.13), we can again make μ_* (hence s_0) small such that

$$\begin{aligned} c_0\mu_\nu\zeta_\nu^3 &= c_0\frac{\mu_\nu}{s_\nu^{2\varepsilon}}(s_\nu^{2\varepsilon}\zeta_\nu^3) \le c_0\frac{\mu_\nu}{s_\nu^{2\varepsilon}} \le \mu_*^{\frac{1}{4}} s_\nu^{2b+3\sigma} \\ &= \mu_*^{\frac{1}{4}} s_0^{(2b+3\sigma)(1+b+\sigma)^\nu} \le \mu_*^{\frac{1}{4}}\left(s_0^{(2b+3\sigma)(b+\sigma)}\right)^\nu \le \frac{\mu_*^{\frac{1}{4}}}{2^{\nu+1}}. \end{aligned} \tag{4.14}$$

Using the definition of s_ν, we clearly have

$$s_{\nu+1} \le s_0^{b+\sigma}s_\nu \le \frac{s_\nu}{16},$$

i.e., H1) holds.

By making μ_* small, we can apply (4.10) and (4.11) to make s_ν small such that

$$\begin{aligned} &\log(n+1)! + 3n\log([\log\frac{1}{s_\nu}]+1) - \frac{E_\nu}{2}([\log\frac{1}{s_\nu}]+1)^3 + (n+1)(\log 2 + \log E_\nu) \\ &\le \log(n+1)! + 3n\log(\log\frac{1}{s_\nu}+2) - (\log\frac{1}{s_\nu})^2 \\ &+(n+1)(\log 2 + \log(2^{-4}r_0(1-\delta)\delta^2) + \nu\log\delta) \\ &\le -\log\frac{1}{s_\nu}. \end{aligned}$$

Hence,

$$\int_{K_{\nu+1}}^\infty \lambda^n e^{-\lambda\frac{E_\nu}{2}}\mathrm{d}\lambda \le 2^n(n+1)!\left(\frac{K_{\nu+1}}{E_\nu}\right)^n e^{-K_{\nu+1}\frac{E_\nu}{2}} \le s_\nu.$$

This verifies H2).

Applying (4.12) and (4.13), we have that

$$\begin{aligned}\frac{c_0\mu_\nu\zeta_\nu}{E_\nu} &= \frac{\mu_\nu}{s_\nu^{2\varepsilon}}\frac{c_0 s_\nu^{2\varepsilon}\zeta_\nu}{E_\nu} \le \frac{\mu_\nu}{s_\nu^{2\varepsilon}} \le s_\nu^{2b+3\sigma} < 1,\\ \frac{c_0 s_\nu\mu_\nu\zeta_\nu}{s_{\nu+1}} &= \frac{\mu_\nu}{s_\nu^{b+\sigma+2\varepsilon}}(c_0 s_\nu^{2\varepsilon}\zeta_\nu) \le \frac{\mu_\nu}{s_\nu^{b+\sigma+2\varepsilon}} < 1,\end{aligned}$$

which verify H6) and H7) respectively. H9) can be verified similarly.

Since

$$\beta_\nu - \beta_{\nu+1} = \frac{\beta_0}{2^{\nu+2}},$$

it follows from (4.14) that

$$\frac{c_0\mu_\nu\zeta_\nu}{\beta_\nu - \beta_{\nu+1}} \le 2\frac{\mu_*^{\frac{1}{4}}}{\beta_0} < \frac{1}{2}$$

as $\mu_* < \frac{\beta_0^4}{2^8}$. This verifies H8).

Applying (4.12) and (4.13) again, we have that

$$\begin{aligned}\frac{c_0\Delta_\nu}{\gamma_{\nu+1}^{n+1}s_{\nu+1}^2\mu_{\nu+1}} &= c_0\frac{\zeta_\nu^3}{4E_\nu}\Big(\frac{s_\nu^3\mu_\nu}{\gamma_\nu^{n+1}s_{\nu+1}^2\mu_{\nu+1}} + \frac{s_\nu^2\mu_\nu^2}{s_{\nu+1}^2\mu_{\nu+1}} + \frac{s_{\nu+1}\mu_\nu}{s_\nu\mu_{\nu+1}}\Big)\\ &= \frac{1}{4}\Big(\frac{c_0 s_\nu^{2\varepsilon}\zeta_\nu^3}{E_\nu}\Big)\Big(\frac{s_\nu^{1-2b-3\sigma-2\varepsilon}}{c_0\gamma_\nu^{n+1}} + \frac{\mu_\nu}{c_0 s_\nu^{2b+3\sigma+2\varepsilon}} + \frac{s_\nu^{b-2\varepsilon}}{c_0}\Big)\\ &\le \frac{1}{4}\Big(\frac{c_0 s_\nu^{2\varepsilon}\zeta_\nu^3}{E_\nu}\Big)\Big(\frac{s_0^{1-2b-3\sigma-2\varepsilon}}{\gamma_\nu^{n+1}} + 2\Big) < 1,\end{aligned}$$

i.e., the first part of H10) holds. The second part of H10) is obvious.

Next, we verify H3)-H5) by induction. For each $\nu = 0, 1, \cdots$, we define M_{kj}^ν, $|k|, |j| > 0$, as in (3.28) with $M_k^\nu, B_k^\nu, U^\nu, Y^\nu$ in place of M_k, B_k, U, Y, respectively, for all $|k| > 0$. Using the same ordering $\mathcal{T}$ as in Section 3, we also have the matrices

$$\mathcal{A}^\nu = (e^{(|k|-|j|)\xi(r_\nu)}M_{kj}^\nu),\ \mathcal{A}^0 = (e^{(|k|-|j|)\xi(r_\nu)}M_{kj}^0),\ 0 < |k|,\ |j| \le K_{\nu+1},$$

where ξ is as in (3.38).

Clearly, H3)-H5) trivially hold for $\nu = 0$. We now assume that for some positive integer ν_* H3)-H5) hold for all $\nu = 1, \cdots, \nu_*$. Then the KAM step described in Section 3 is valid for all $\nu = 1, \cdots, \nu_*$. In particular, Lemmas 3.5, 3.6 hold for all

$\nu = 1, \cdots, \nu_*$. It follows from (4.14) that

$$
\begin{aligned}
&|\partial_\lambda^l(\Omega_{\nu_*+1} - \Omega_0)| \le \sum_{\nu=0}^{\nu_*} |\partial_\lambda^l(\Omega_{\nu+1} - \Omega_\nu)|_{\mathcal{O}_{\nu+1}} \le \sum_{\nu=0}^{\nu_*} c_0\gamma_0^{n+1} s_\nu \mu_\nu \zeta_\nu \\
&\le \gamma_0^{n+1} s_0 \sum_{\nu=0}^{\nu_*} \frac{\mu_*^{\frac14}}{2^{\nu+1}} \le \gamma_0^{2(n+1)} \mu_*^{\frac12} < \mu_*^{\frac14}, \\
&|\partial_\lambda^l(\mathcal{M}^{\nu_*+1} - \mathcal{M}^0)|_{D_{\nu_*+1}\times\mathcal{O}_{\nu_*+1}} \le \sum_{\nu=0}^{\nu_*} |\partial_\lambda^l(\mathcal{M}_{\nu+1} - \mathcal{M}_\nu)|_{D_{\nu+1}\times\mathcal{O}_{\nu+1}} \\
&\le \sum_{\nu=0}^{\nu_*} c_0\gamma_0^{n+1}\mu_\nu\zeta_\nu \le \gamma_0^{n+1} \sum_{\nu=0}^{\nu_*} \frac{\mu_*^{\frac14}}{2^{\nu+1}} \le \gamma_0^{n+1}\mu_*^{\frac14} < \mu_*^{\frac14}, \\
&\max_{|k|>0} \sum_{|j|>0} e^{(|k|-|j|)\xi(r_{\nu_*+1})} |M_{kj}^{\nu_*+1} - M_{kj}^0|_{\mathcal{O}_{\nu_*+1}} \\
&\le \sum_{\nu=0}^{\nu_*} \max_{|k|>0} \sum_{|j|>0} e^{(|k|-|j|)\xi(r_{\nu+1})} |M_{kj}^{\nu+1} - M_{kj}^{\nu}|_{\mathcal{O}_{\nu+1}} \\
&\le \sum_{\nu=0}^{\nu_*} c_0\gamma_0^{n+1}\mu_\nu\zeta_\nu^2 \le \gamma_0^{n+1} \sum_{\nu=0}^{\nu_*} \frac{\mu_*^{\frac14}}{2^{\nu+1}} < \mu_*^{\frac14}.
\end{aligned}
$$

Thus, H3)-H5) also hold for $\nu = \nu_* + 1$.

Above all, H1)-H10) hold for all $\nu = 0, 1, \cdots$, i.e., the KAM step described in Section 3 is valid for all $\nu = 0, 1, \cdots$. Now, part 4) of the lemma easily follows from Lemma 3.6. Also, using (4.14), part 3) of the lemma follows from Lemma 3.4; (4.7) follows from Lemma 3.8; (4.2) and (4.4) follow from Lemma 3.5 1); (4.6) follows from Lemma 3.7. Finally, (4.1), (4.3) and (4.5) follow from (4.2), (4.4) and (4.6) respectively, and, part 2) of the lemma follows from an inductive application of (3.10). This completes the proof. □

5. Proof of Theorems 1 and 2

We first consider Theorem 2 1). By making $\mu_* = \mu_*(r, s)$ small in Theorem 2, we obtain the following sequences

$$
\begin{aligned}
&\Psi^\nu = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Phi_\nu : \tilde{D}_{\nu+1} \times \mathcal{O}_{\nu+1} \to \tilde{D}_0, \\
&H \circ \Psi^\nu = H_\nu = N_\nu + P_\nu, \\
&N_\nu = e_\nu + \langle \Omega_\nu, y\rangle + \frac12 \langle \begin{pmatrix} y \\ z \end{pmatrix}, \mathcal{M}^\nu \begin{pmatrix} y \\ z \end{pmatrix} \rangle + h_0(x, y, z, \lambda),
\end{aligned}
$$

$\nu = 0, 1, \cdots$, which satisfy all properties described in Lemma 4.1.

Let

$$
\mathcal{O}_* = \prod_{\nu=0}^{\infty} \mathcal{O}_\nu, \ G_* = D(\frac{r_0}{2}, \frac{\beta_0}{2}) \times \mathcal{O}_*.
$$

Then $\mathcal{O}_*$ is a Cantor-like set consisting of non-resonant frequencies, and moreover, a measure estimate similar to that in [24] (also [4, 13]) yields that $|\mathcal{O}\backslash\mathcal{O}_*| = O(\gamma_0^{\frac{1}{n_*-1}})$.

By Lemma 4.1 1), it is clear that e_ν and Ω_ν converge uniformly on $\mathcal{O}_*$, and, $\mathcal{M}^\nu$ converges uniformly on $\mathcal{D}(\frac{r_0}{2}) \times \mathcal{O}_*$. We denote their limits by e_∞, Ω_∞, $\mathcal{M}^\infty$, respectively. Then $\mathcal{M}^\infty$ is real analytic in x, and, it follows from the Whitney's extension theorem ([17]) that these limits are also Hölder continuous in ω. Moreover, by Lemma 4.1 1), we have that

$$\begin{aligned} |e_\infty - e_0|_{\mathcal{O}_*} &= O(\gamma_0^{2(n+1)}\mu_*^{\frac{1}{2}}), \\ |\Omega_\infty - \Omega_0|_{\mathcal{O}_*} &= O(\gamma_0^{2(n+1)}\mu_*^{\frac{1}{2}}), \\ |\mathcal{M}^\infty - \mathcal{M}^0|_{\mathcal{D}(\frac{r_0}{2})\times\mathcal{O}_*} &= O(\gamma_0^{n+1}\mu_*^{\frac{1}{4}}). \end{aligned}$$

Thus, N_ν converges uniformly on G_* to

$$N_\infty = e_\infty + \langle\Omega_\infty, y\rangle + \frac{1}{2}\langle\begin{pmatrix} y \\ z\end{pmatrix}, \mathcal{M}^\infty\begin{pmatrix} y \\ z\end{pmatrix}\rangle + h_0(x,y,z,\lambda).$$

To show the convergence of Ψ^ν on G_*, we note that

$$\begin{aligned} \Psi^\nu - \Psi^{\nu-1} &= \Phi_0\circ\cdots\circ\Phi_\nu - \Phi_0\circ\cdots\circ\Phi_{\nu-1} \\ &= \int_0^1 D(\Phi_0\circ\cdots\circ\Phi_{\nu-1})(id + \theta(\Phi_\nu - id))\mathrm{d}\theta(\Phi_\nu - id). \end{aligned}$$

It follows from Lemma 4.1 3) that

$$\begin{aligned} & |D(\Phi_1\circ\cdots\circ\Phi_{\nu-1})(id+\theta(\Phi_\nu - id))| \\ \le\ & |D\Phi_1(\Phi_2\circ\cdots\circ\Phi_{\nu-1})(id+\theta(\Phi_\nu - id)))|\cdots|D\Phi_{\nu-1}(id+\theta(\Phi_\nu - id))| \\ \le\ & (1+\frac{\mu_*^{\frac{1}{4}}}{2})\cdots(1+\frac{\mu_*^{\frac{1}{4}}}{2^{\nu-1}}) \le e^{\frac{\mu_*^{\frac{1}{4}}}{2}+\cdots+\frac{\mu_*^{\frac{1}{4}}}{2^{\nu-1}}} \le e^{\mu_*^{\frac{1}{4}}}, \end{aligned}$$

and

$$|\Phi_\nu - id|_{G_*} \le \frac{\mu_*^{\frac{1}{4}}}{2^\nu}.$$

Hence,

$$|\Psi^\nu - \Psi^{\nu-1}|_{G_*} \le e\frac{\mu_*^{\frac{1}{4}}}{2^\nu},$$

which implies the uniform convergence of Ψ^ν. Let Ψ^∞ be the limit of Ψ^ν. Then, Ψ^∞ is uniformly continuous in $\lambda\in\mathcal{O}_*$ and analytic in $(x,y,z)\in D(\frac{r_0}{2},\frac{\beta_0}{2})$, and moreover,

$$|\Psi^\infty - id|_{G_*} = O(\mu_*^{\frac{1}{4}}).$$

Using Lemma 4.1 3) and a similar argument as above, we can further show the uniform convergence of $D\Psi^\nu$, $D^2\Psi^\nu$ to $D\Psi^\infty$, $D^2\Psi^\infty$, respectively, on G_*. Hence, on G_*,

$$P_\nu = H\circ\Psi^\nu - N_\nu,\ DP_\nu,\ D^2P_\nu$$

converge uniformly to

$$P_\infty = H\circ\Psi^\infty - N_\infty,\ DP_\infty,\ D^2P_\infty,$$

respectively. Clearly, these limits above are uniformly continuous in $\lambda\in\mathcal{O}_*$ and analytic in $(x,y,z)\in D(\frac{r_0}{2},\frac{\beta_0}{2})$.

Note that

$$|P_\nu|_{D_\nu} \le \gamma_\nu^{n+1}s_\nu^2\mu_\nu.$$

It follows from Cauchy's estimate that, for any $\lambda\in\mathcal{O}_*$, $j\in Z_+^n$, $k\in Z_+^{2m}$ with $|j|+|k|\le 2$,

$$|\partial_y^j\partial_z^kP_\nu|_{D(r_{\nu+i},\frac{1}{2}s_\nu)} \le \gamma_\nu^{n+1}\mu_\nu.$$

Since, by (4.9), the right hand side of the above converges to 0 as $\nu \to 0$, we have that

$$\partial_y^j \partial_z^k P_\infty|_{(y,z)=0} = 0$$

for all $x \in T^n$, $\lambda \in \mathcal{O}_*$, $j \in Z_+^n$, $k \in Z_+^{2m}$ with $|j| + |k| \le 2$.

Thus, for each $\lambda \in \mathcal{O}_*$, the perturbed system (1.3) possesses an analytic, quasi-periodic, invariant torus with the Diophantine toral frequency $\Omega_\infty(\lambda)$. Since, by Lemma 4.1 2),

$$(\Omega_\nu(\lambda))_i = \omega_i(\lambda),\ \lambda \in \Omega_\nu,\ i = 1, 2, \cdots, n_0,$$

we have that

$$(\Omega_\infty(\lambda))_i = \omega_i(\lambda),\ \lambda \in \mathcal{O}_*,\ i = 1, 2, \cdots, n_0,$$

i.e., the perturbed toral frequencies also preserve the first n_0 components of their corresponding ones. This proves part 1) of Theorem 2.

Theorem 1 and part 2) of Theorem 2 are almost immediate consequences of part 1) of Theorem 2 with respect to $n_0 = 0$ and $n_0 = n$ respectively. In the case that $n_0 = 0$, we can modify the arguments in Section 3 by choosing $y_* = 0$, $U^0 = U = \emptyset$, $\mathrm{diag}(I_{n_0}, O) = O$, and, $\mathrm{diag}(O, I_{n-n_0}) = I_n$. In the case that $n_0 = n$, we let $\hat{\mathcal{O}}_\gamma$ be the set of $\lambda \in \mathcal{O}_0$ such that $\Omega_0(\lambda)$ is Diophantine of Diophantine type (γ, τ) for a fixed $\tau > n - 1$. It is clear that $U^0 = [A^0]$, $U = [A]$, $\mathrm{diag}(I_{n_0}, O) = I_n$, and, $\mathrm{diag}(O, I_{n-n_0}) = O$. Hence, $\Omega_\nu \equiv \Omega_0$ and $\mathcal{O}_\nu = \hat{\mathcal{O}}_\gamma$ for all $\nu = 0, 1, \cdots$, i.e., $\Omega_\infty \equiv \Omega_0$ and $\mathcal{O}_* = \hat{\mathcal{O}}_\gamma$.

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

[3] C. -Q. Cheng and Y. S. Sun, Existence of invariant tori in three dimensional measure-preserving mappings, *Celestial Mech. Dyn. Astronom.* **47** (1990), 275-292

[4] S.-N. Chow, Y. Li, and Y. Yi, Persistence of invariant tori on submanifolds in Hamiltonian systems, *J. Nonl. Sci.*, **12** (2002), 585-617

[5] R. de la Llave, A. Gonzalez, A. Jorba, and J. Villanueva, KAM theory without action-angle variables, preprint 2003

[6] L. H. Eliasson, Absolutely convergent series expansions for quasi periodic motions, *MPEJ* **2** (1996), 1-33

[7] L. H. Eliasson, Biasymptotic solutions of perturbed integrable Hamiltonian systems, *Bol. Soc. Mat.* **25**(1994), 57-76

[8] G. Gallavotti and G. Gentile, Hyperbolic low-dimensional invariant tori and summations of divergent series, *Comm. Math. Phys.* **227** (2002), 421–460

[9] S. M. Graff, On the continuation of hyperbolic invariant tori for Hamiltonian systems, *J. Differential Equations* **15** (1974),1-69

[10] J. Gresson, The transfer lemma for Graff tori and Arnold diffusion time, *Discret. Cont. Dyn. Sys.* **7** (2001),787-800

[11] A. Jorba, R. de la Llave, M. Zou, Lindstedt series for lower dimensional tori, In Hamiltonian systems with more than two degrees of freedoms, *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.* **533** (ed. C. Simo), Kluwer Academic Publishers, 1999, 151–167

[12] A. N. Kolmogorov, On the conservation of conditionally periodic motions for a small change in Hamilton's function, *Dokl. Akad. Nauk. SSSR* **98** (1954), 525-530

[13] Y. Li and Y. Yi, Persistence of invariant tori in generalized Hamiltonian systems, *Ergod. Th. Dyn. Sys.* **22** (2002), 1233-1261

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

[15] J. Moser, On invariant curves of area preserving mappings of an annulus, *Nachr. Akad. Wiss. Gott. Math. Phys K1,* **2** (1962), 1-20

[16] P. J. Olver, Applications of Lie groups to differential equations, 2nd ed., Springer-Verlag, New York, 1993

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

[18] M. Rudnev and S. Wiggins, On a partially hyperbolic KAM theorem, *Regul. Chaotic Dyn.* **4** (1999), 39–58

[19] H. Rüssmann, Nondegeneracy in the perturbation theory of integrable dynamical systems, *Number theory and dynamical systems* (York, 1987), 5-18, London Math. Soc. Lecture Note Ser., 134, Cambridge Univ. Press, Cambridge, 1989, *Stochastics, algebra and analysis in classical and quantum dynamics* (Marseille, 1988), 211-223, Math. Appl., 59, Kluwer Acad. Publ., Dordrecht, 1990

[20] H. Rüssmann, Invariant tori in non-degenerate nearly integrable Hamiltonian systems, *Regul. Chaotic Dyn.* **6** no. 2 (2001) 119-204

[21] D. Salamon and E. Zehnder, KAM theory in configuration space, *Comment. Math. Helvetici* **64** (1989), 84-132

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

[23] S. Wiggins, Global Bifurcations and Chaos: Analytic Methods, Springer-Verlag: New York, Heidelberg, Berlin, 1988

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

[25] E. Zehnder, Generalized implicit function theorem with applications to some small divisors I and II, *Comm. Pure Appl. Math.* **28** (1975), 91-140; **29** (1976), 49-111

Y. Li: Department of Mathematics, Jilin University, Changchun 130012, P. R. China
E-mail address: `ylimd@email.jlu.edu.cn`

Y. Yi: School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail address: `yi@math.gatech.edu`