A QUASI-PERIODIC POINCARÉ'S THEOREM

YONG LI AND YINGFEI YI

ABSTRACT. We study the persistence of invariant tori on resonant surfaces of a nearly integrable Hamiltonian system under the usual Kolmogorov non-degenerate condition. By introducing a quasi-linear iterative scheme to deal with small divisors, we generalize the Poincaré theorem on the maximal resonance case (i.e., the periodic case) to the general resonance case (i.e., the quasi-periodic case) by showing the persistence of majority of invariant tori associated to non-degenerate relative equilibria on any resonant surface.

1. INTRODUCTION

The present work concerns the study of the persistence of invariant tori in the resonance zone of a nearly integrable, real analytic Hamiltonian system of the following action-angle form

$$H(x, y, \varepsilon) = H_0(y) + \varepsilon P(x, y, \varepsilon), \tag{1.1}$$

where $y \in G$, $G \subset R^d$ is a bounded closed region (closure of a bounded, non-empty open set), $x \in T^d (= \frac{R^d}{Z^d})$, d is the degree of freedom, and $\varepsilon > 0$ is a small parameter.

With the symplectic form

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

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 $\omega(y) = \frac{\partial H_0}{\partial y}(y)$. Thus, for $\varepsilon = 0$, the phase space $R^d \times T^d$ is foliated into invariant tori $T_y = \{y\} \times T^d$ with the frequency vectors $\omega(y)$.

We first assume the usual Kolmogorov non-degenerate condition:

A1) *The Hessian* $\frac{\partial^2 H_0}{\partial y^2}(y)$ *is non-singular for all* $y \in G$.

Under this condition, the map $\omega : G \to \omega(G), y \mapsto \omega(y)$ defines a local diffeomorphism. Without loss of generality, we further assume that the diffeomorphism is global on G. Thus, points of G have one to one correspondence with the frequency vectors in the region $\omega(G)$, among which the non-resonant ones form

---

1991 *Mathematics Subject Classification.* Primary 58F05, 58F27, 58F30.

Key words and phrases. Hamiltonian systems, KAM theory, normal form, persistence, invariant tori, Poincaré theorem, quasi-linear scheme, resonant surface.

The first author was partially supported by NSFC grant 19971042, the National 973 Project of China: Nonlinearity, and the outstanding youth project of the Ministry of Education of China.

The second author was partially supported by NSF grant DMS9803581.

a full (Lebesgue) measure set. On one hand, by the celebrated KAM theorem (Kolmogorov [8], Arnold [1], Moser [9]), majority of the non-resonant tori among the family $\{T_y : y \in G\}$ will persist as ε small. On the other hand, the family $\{T_y : y \in G\}$ also contains resonant tori of all type of resonances, and, it is well known that these resonant tori tend to be destroyed under arbitrary generic perturbations and give rise to a resonance zone containing both stochastic trajectories and regular orbits (see [3], [13] and references therein). To characterize regular orbits in the resonance zone, an essential problem is to analyze mechanisms of destruction of the resonant tori and the persistence of certain lower dimensional tori which are split from the resonant ones.

Such persistence problem was first considered by Poincaré ([10]) within the class of maximal resonances, i.e., a resonant torus in this class is foliated into periodic orbits. With respect to (1.1), the Poincaré theorem states that any periodic orbit associated to a non-degenerate relative equilibrium will persist.

To characterize general resonance types, we consider a proper, non-trivial subgroup g of Z^d of rank m, where $0 < m < d$. Then

$$O(g, G) = \{y \in G : \langle k, \omega(y)\rangle = 0, k \in g\}$$

is a $n = d - m$ dimensional sub-manifold of G – the so called *g-resonant surface with multiplicity m*. This manifold characterizes a unique class of resonant tori $\{T_y : y \in O(g, G)\}$ associated to the resonance type determined by g.

The group g also determines the splitting of the resonant tori into lower dimensional ones. Let $\{\tau_1, \cdots, \tau_m\}$ and $\{\tau_1', \cdots, \tau_n'\}$ be bases of g and the quotient group $\frac{Z^d}{g}$, respectively, such that

$$K_0 = (K_1, K_2)$$

is unimodular (i.e., $\det K_0 = 1$), where

$$K_1 = (\tau_1', \cdots, \tau_n'), \quad K_2 = (\tau_1, \cdots, \tau_m).$$

Then the toral automorphism K_0 defines a new symplectic coordinate $\binom{\psi}{\varphi} = K_0^\top x$ on T^d, where $\varphi = K_1^\top x \in T^m, \psi = K_2^\top x \in T^n$, under which the g-resonant surface becomes

$$O(g, G) = \{y \in G : K_2^\top \omega(y) = 0\}.$$

Moreover, it is easy to see that for each $y \in O(g, G)$, the resonant torus T_y is foliated into invariant n-tori

$$T_y(\varphi) = \{y\} \times T^n \times \{\varphi\}, \quad \varphi \in T^m.$$

Under the new coordinate, the unperturbed motion of (1.1) becomes

$$\begin{cases} \dot\psi &= K_1^\top \omega(y), \\ \dot\varphi &= 0, \\ \dot y &= 0. \end{cases}$$

Thus, each (φ, y) is a relative equilibrium of the unperturbed system, and, the flow on a n-torus $T_y(\varphi)$ is parallel with the frequency vector $K_1^\top \omega(y)$.

Define

$$h_0(\varphi, y) = \int_{T^n} \bar P(\psi, \varphi, y) d\psi,$$

where,

$$\bar P(\psi, \varphi, y) = P((K_0^\top)^{-1} \begin{pmatrix} \psi \\ \varphi \end{pmatrix}, y, 0).$$

A n-torus $T^n_y(\varphi)$ is said to be *associated to a non-degenerate relative equilibrium* (φ, y) if φ is a non-degenerate critical point of $h_0(\cdot, y)$, i.e., $\frac{\partial h_0}{\partial \varphi}(\varphi, y) = 0$ and $\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi, y)$ is non-singular.

Motivated by the condition of Poincaré , we consider the set

$$O_0(g, G) = \{y \in O(g, G) : h_0(\cdot, y) \text{ admits a non-degenerate critical point}\}$$

and assume that

A2) *$O_0(g, G)$ is non-empty.*

Clearly, the set $O_0(g, G)$ depends only on g, G but not on a particular choice of K_1, K_2. Moreover, since the non-degeneracy described above is an open property, $O_0(g, G)$ also admits positive Lebesgue measure under the condition A2). We note that A2) is a generic condition in the sense that for any given g, G, H_0, perturbations $P|_{\varepsilon=0}$ satisfying A2) form an open dense subset of the set of real analytic functions on $T^d \times G$.

Due to the existence of resonant frequency vectors among the n-tori which are associated to non-degenerate relative equilibria, the persistence problem in the nature of the Poincaré theorem for general multiplicity resonances should be considered with respect to the Lebesgue measure on $O(g, G)$, similarly to the classical KAM case.

For any small $\rho > 0$, let

$$O_\rho(g, G) = \{y \in O_0(g, G) : \text{dist}(y, \partial O_0(g, G)) \geq \rho\},$$

where ∂O_0 denotes the boundary of O_0. Our main result states as follows.

Theorem 1. *Assume* A1), A2). *Then for any fixed sufficiently small $\rho > 0$ and any positive integer l_0, there is an $\varepsilon_0 = \varepsilon_0(g, G, \rho, l_0) > 0$ and a family of Cantor sets $\Lambda_\varepsilon \subset O_\rho(g, G)$, $0 < \varepsilon \leq \varepsilon_0$, such that the following holds.*

1) *Each n-torus $T_y(\varphi)$, $y \in \Lambda_\varepsilon$, associated to a non-degenerate relative equilibrium (φ, y) will persist and gives rise to an analytic, quasi-periodic, invariant n-torus $T_{\varepsilon,y}(\varphi)$ of the perturbed system* (1.1). *Moreover, all such perturbed tori corresponding to a same $y \in \Lambda_\varepsilon$ are symplectically conjugated to the standard quasi-periodic n-torus T^n with the Diophantine frequency vector $\omega^* = K_1^\top \omega(y)$;*
2) *Within each connected component of $O_\rho(g, G)$, the set of perturbed tori divides into a finite number of C^{l_0} Whitney smooth families over Λ_ε, varying analytically in ε.*
3) *The relative Lebesgue measure $|O_\rho(g, G) \setminus \Lambda_\varepsilon| \to 0$ as $\varepsilon \to 0$.*

In applications of the theorem, one can also replace the global conditions A1), A2) by similar non-degenerate conditions at a particular resonant point $y_0 \in R^d$ (i.e., $\omega(y_0)$ is a resonant vector). Indeed, let the Hessian $\frac{\partial^2 H_0}{\partial y^2}(y_0)$ be non-singular and g be a subgroup of Z^d such that $\langle k, \omega(y_0)\rangle = 0$ for all $k \in g$ (such g is uniquely defined). If φ_0 is a non-degenerate critical point of $h_0(\cdot, y_0)$, then the implicit function theorem implies that there is a neighborhood G of y_0 on which both A1) and A2) are valid, i.e., the above theorem holds on G.

Poincaré's theorem has been recently explored in various cases of lower multiplicity resonances in Hamiltonian systems like (1.1). In particular, persistence problems for multiplicity 1 (i.e. $n = d-1$) resonant tori have been studied extensively in works of Eliasson [7], Cheng [4], Chierchia and Gallavotti [5], Rudnev and Wiggins [12], etc. The general multiplicity resonance cases were considered in [15] and [6]. In [15], Treshchev proved the following result: If $y_0 \in O(g,G)$ is such that $\omega^* = K_1^\top \omega(y_0)$ is Diophantine and if no eigenvalue of $\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_0, y_0) K_2^\top \frac{\partial^2 H_0}{\partial y^2}(y_0) K_2$ is positive or zero, then the torus $T_{y_0}(\varphi_0)$ persists (see also [2] for a variational approach of the Treshchev's result). Recently, Cong *et al* ([6]) showed a general multiplicity result under the g non-degenerate condition, i.e., $K_2^\top \frac{\partial^2 H_0}{\partial y^2}(y) K_2$ is non-singular over G.

Treshchev's result restricts to the case that not only φ_0 is a hyperbolic critical point of $h_0(\cdot, y_0)$, but also $T_{y_0}(\varphi_0)$ is a normally hyperbolic torus. Our result works for any type of non-degenerate critical points of $h_0(\cdot, y_0)$, and, the associated unperturbed tori can certainly be normally degenerate (see the example in Section 2). Our generalization to the result of Cong *et al* is also significant because of the failure of the g non-degeneracy in general (see also Section 2). In a case that the g non-degenerate condition fails, one could try to apply the result of Cong *et al* by introducing a new symplectic transformation $y = NY$, $N = (N_1, N_2)$ which makes $N_2^\top K_0^\top \frac{\partial^2 H_0}{\partial y^2}(y) K_0 N_2$ non-singular. But this would change the group g and therefore change the resonance type originally considered.

The above theorem will be proved based on the normal form theory and a quasi-linear iterative scheme. It follows from arguments in Treshchev [15] that, near each non-degenerate relative equilibrium (φ_0, y_0) such that $\omega = K_1^\top \omega(y_0)$ is Diophantine, one can separate the first order resonant terms according to the group g and reduce the Hamiltonian (1.1) into a canonical form

$$H = e(\omega) + \langle \omega, y \rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M(\omega) \begin{pmatrix} y \\ z \end{pmatrix} \rangle + P(x,y,z,\omega) \tag{1.2}$$

which carries the standard symplectic structure, where $(x,y,z) \in T^n \times R^n \times R^{2m}$, $\delta > 0$ is a constant, M is real symmetric and non-singular over a bounded closed region $\mathcal{O} \subset R^n$, e, M are C^∞ on $\mathcal{O}$, and, P 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^∞ in $\omega \in \mathcal{O}$ (see Section 5).

The following theorem from which Theorem 1 will follow should be of importance on its own right.

Theorem 2. *Consider* (1.2) *and let* $\hat{\mathcal{O}}_\gamma$, $\gamma > 0$, *denote the Diophantine set*

$$\{\omega \in \mathcal{O} : \ |\langle k, \omega \rangle| > \frac{\gamma}{|k|^\tau}, \ \textit{for all } k \in Z^n \setminus \{0\}\},$$

where $|k| = \sum_{i=1}^n |k_i|$ *and* $\tau > n-1$ *is fixed. Given an integer* $l_0 \geq (n+2m)^2$. *If* δ *is sufficiently small and if there exists a sufficiently small* $\mu = \mu(r,s,l_0) > 0$ *such that*

$$|\partial_\omega^l \partial_x^i \partial P|, \ s|\partial_\omega^l \partial_x^i \partial_{(y,z)} P|, \ s^2 |\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P| \leq \delta \gamma^{3b} s^2 \mu \tag{1.3}$$

for all $(x,y,z) \in D(r,s)$, $\omega \in \mathcal{O}$, $(l,i,j) \in Z_+^n \times Z_+^n \times Z_+^{n+2m}$ *with* $|l| + |i| \leq l_0, 2 \leq |j| \leq l_0$, *where* $b = (2{l_0}^2 + 3)(n+2m)^2$, *then the following holds.*

1) *There exists a Cantor set* $\mathcal{O}_\gamma \subset \hat{\mathcal{O}}_\gamma$ *and a* C^{l_0-1} *Whitney smooth family of* C^{l_0} *symplectic transformations*

$$\Psi_\omega : D(\frac{r}{2}, \frac{s}{4}) \to D(r,s), \qquad \omega \in \mathcal{O}_\gamma,$$

which is C^{l_0} *uniformly close to the identity and is real analytic in* x *when* $y = 0, z = 0$, *such that*

$$H \circ \Psi_\omega = e_* + \langle \omega, y \rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M_* \begin{pmatrix} y \\ z \end{pmatrix} \rangle + P_*,$$

where $e_* = e_*(\omega), M_* = M_*(\omega)$ *are* C^{l_0-1} *Whitney smooth on* $\omega \in \mathcal{O}_\gamma$, $P_* = P_*(x,y,z,\omega)$ *is* C^{l_0} *in* (x,y,z) *and* C^{l_0-1} *Whitney smooth in* ω, *and moreover, for arbitrary* $0 < \epsilon < 1$,

$$|\partial_\omega^l e_* - \partial_\omega^l e|_{\mathcal{O}_\gamma} = o(\gamma^b \mu^{1-\epsilon}), \; |l| \le l_0 - 1,$$

$$|\partial_\omega^l M_* - \partial_\omega^l M|_{\mathcal{O}_\gamma} = o(\gamma^b \mu^{1-\epsilon}), \; |l| \le l_0 - 1,$$

$$\partial_y^j \partial_z^k P_*|_{(y,z)=(0,0)} = 0, \; x \in T^n, \; \omega \in \mathcal{O}_\gamma, \; j \in Z_+^n, \; k \in Z_+^{2m}, \; |j| + |k| \le 2.$$

Thus, for each $\omega \in \mathcal{O}_\gamma$, *the perturbed system* (1.2) *admits an analytic, quasi-periodic,* (*Floquet*) *invariant* n*-torus with the Diophantine frequency vector* ω, *which is slightly deformed from the unperturbed torus with the same frequency vector* ω. *Moreover, these perturbed tori form a* C^{l_0-1} *Whitney smooth family.*

2) *The Lebesgue measure* $|\mathcal{O} \setminus \mathcal{O}_\gamma| = O(\gamma)$ *as* $\gamma \to 0$.

In the above, derivatives of e_*, M_* with respect to $\omega \in \mathcal{O}_*$ are in the sense of Whitney. For the notion of Whitney smoothness on a Cantor set, we refer the readers to [3], [11].

In the proof of Theorem 2, we shall use a quasi-linear iterative scheme instead of the usual KAM linear scheme to overcome technical difficulties due to the generality of the Hamiltonian (1.2). We note that the matrix $M(\omega)$ is not necessarily of a diagonal form, and, in general, it cannot be made diagonal via a smooth family of symplectic change of variables. Moreover, associated to the possible g degeneracy of (1.1), the z direction in (1.2) can well be degenerate, i.e., the right lower $2m \times 2m$ minor of M can be singular over $\mathcal{O}$ (see the example in Section 2).

With almost the same proofs using the quasi-linear iterative scheme, Theorem 2 can be made slightly general to apply to a Hamiltonian system like (1.2) with a higher order term $h(y,z,\omega) = O(y^\imath z^\jmath)$, $|\imath| + |\jmath| \ge 3$, adding to the integrable part. We prefer not to do so in the present paper for the sake of briefness.

The paper is organized as follows. In Section 2, we give an example to illustrate some significance of our result. The quasi-linear iterative scheme is described in detail in Section 3 for Hamiltonian (1.2) for one KAM cycle. In Section 4, we prove Theorem 2 by deriving an iteration lemma which ensures the validity of all KAM steps. The proof of Theorem 1 will be completed in Section 4 by using Treshchev's reduction and Theorem 2.

Throughout the paper, we shall use the same symbol $|\cdot|$ to denote norm of vectors, matrices, absolute value of functions, and measure of sets etc., and use $|\cdot|_D$ to denote the supremum norm of functions on a domain D. They will have obvious meanings unless specified otherwise. Also, for any two complex column vectors ξ, ζ of same dimension, $\langle \xi, \zeta \rangle$ always stands for $\xi^\top \zeta$, i.e., the transpose of ξ times ζ. For simplicity, we shall not specify smoothness orders for functions which are either sufficiently smooth or have obvious orders of smoothness indicated by their derivatives taken.

Aknowledgement. This work is partially done when the first author was visiting the Center for Dynamical Systems and Nonlinear Studies, Georgia Institute of Technology, and the second author was visiting the National University of Singapore. We would like to thank both Institutions for support. We are grateful to the referee for careful reading and for valuable comments which lead to improvements of the paper.

2. An example

Consider the following Hamiltonian on $T^5 \times R^5$:

$$H = H_0(y) + \varepsilon P_0(x) + \varepsilon^2 P_1(x, y, \varepsilon), \tag{2.1}$$

where $y = (y_1, \cdots, y_5)^\top \in R^5, x = (x_1, \cdots, x_5)^\top \in T^5$, ε is a small parameter,

$$H_0(y) = y_1 y_5 + y_2 y_4 + \frac{1}{2} y_3^2, \tag{2.2}$$

$$P_0(x) = \sin 2\pi x_4 + \sin 2\pi x_5,$$

and P_1 is a real analytic function.

We note that

$$\omega(y) = \frac{\partial H_0}{\partial y}(y) = \begin{pmatrix} y_5 \\ y_4 \\ y_3 \\ y_2 \\ y_1 \end{pmatrix},$$

$$\frac{\partial^2 H_0}{\partial y^2}(y) = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Clearly, the condition A1) is satisfied on any bounded close region $G \subset R^5$. We now consider the rank 2 subgroup $g = \{z = (z_1, \cdots, z_5)^\top \in Z^5 : z_1 = z_2 = z_3 = 0\}$. Such choice of g yields the g-resonant surface

$$O(g, G) = \{y \in G : \langle k, \omega(y)\rangle = 0, k \in g\} = \{y \in G : y_1 = y_2 = 0\},$$

which is just the intersection of G with the $y_3y_4y_5$-plane.

Since the condition A2) only depends on g, G, we take the liberty to choose K_1, K_2, K_0 which are easy to work with. Let $K_1 = (\tau_1', \tau_2', \tau_3'), K_2 = (\tau_1, \tau_2)$, where $\tau_1' = (1, 0, 0, 0, 0)^\top$, $\tau_2' = (0, 1, 0, 0, 0, 0)^\top, \tau_3' = (0, , 0, 1, 0, 0)^\top$, $\tau_1 = (0, 0, 0, 1, 0)^\top$, $\tau_2 = (0, 0, 0, 0, 1)^\top$. It turns out that $K_0 = (K_1, K_2)$ is the identity matrix I_5. Therefore, $\psi = (x_1, x_2, x_3)^\top \in T^3, \varphi = (x_4, x_5)^\top \in T^2$, and $h_0(\varphi, y) \equiv h_0(\varphi) = \sin 2\pi x_4 + \sin 2\pi x_5$.

The function h_0 has 4 critical points $\varphi_1 = (\frac{1}{4}, \frac{1}{4})^\top, \varphi_2 = (\frac{1}{4}, \frac{3}{4})^\top, \varphi_3 = (\frac{3}{4}, \frac{1}{4})^\top$, $\varphi_4 = (\frac{3}{4}, \frac{3}{4})^\top$ which are all non-degenerate. Indeed,

$$\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_1) = -\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_4) = -(2\pi)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_2) = -\frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_3) = -(2\pi)^2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Thus, $O_0(g,G) = O(g,G)$, and, (2.2) admits four families of non-degenerate relative equilibria $\{\varphi_i\} \times O(g,G)$, $i = 1,2,3,4$. Applying Theorem 1 to a larger region containing G, we conclude that for sufficiently small $\varepsilon > 0$, there is a Cantor set $\Lambda_\varepsilon \subset O(g,G)$, with $|O(g,G) \setminus \Lambda_\varepsilon| \to 0$ as $\varepsilon \to 0$, such that the Hamiltonian (2.1) admits four Whitney smooth families of quasi-periodic 3-tori $\{T^i_{\varepsilon,y} : y \in \Lambda_\varepsilon\}$, $i = 1,2,3,4$, which, for each fixed $y \in O_\varepsilon(g,G)$, are all symplectically conjugated to the quasi-periodic 3-torus T^3 with the frequency vector $\omega^* = K_1^\top \omega(y) = (y_5, y_4, y_3)^\top$.

Remark:

1) Among the four families of non-degenerate relative equilibria of (2.2) associated to non-degenerate critical points of $h_0(\varphi)$, there are one family of hyperbolic type, one family of elliptic type, and two families of mixed type. Thus, for this example, the persistence of invariant tori associated to non-degenerate relative equilibria of all type is guaranteed by our theorem.
2) Since $K_2^\top \frac{\partial^2 H_0}{\partial y^2}(y) K_2 \equiv 0$ for any choice of K_2, all the unperturbed tori associated to $\{\varphi_i\} \times O(g,G)$, $i = 1,2,3,4$, are g-degenerate (i.e., normally degenerate). Therefore, neither the result of Treshchev ([15]) nor the result of Cong *et al* ([6]) is applicable to this example.
3) The application of Theorem 1 depends on a careful splitting of resonances, i.e., on an appropriate choice of g. For instance, if we chose a different rank 2 subgroup $g = \{z = (z_1, \cdots, z_5)^\top \in Z^5 : z_1 = z_4 = z_5 = 0\}$, then $\varphi = (x_2, x_3)^\top \in T^2$ and $h_0(\varphi, y) \equiv$constant, which makes $O_0(g,G)$ empty. To study the persistence problem for this type of resonance, one perhaps need to exam further non-degeneracy among higher order perturbations.

3. KAM Step

The next two sections are devoted to the proof of Theorem 2. Following the framework of the classical KAM theory, we shall find a symplectic transformation, involving infinitely many successive steps (referred to as *KAM steps*) of iterations, to the Hamiltonian (1.2), so that after each iteration the x-dependent terms are pushed into a new perturbation consisting of terms of either smaller scales or at least cubic order in the action variable. As usual, we shall carry out the KAM steps by induction.

Let $s, r, \gamma, \mu, \delta$ be as in Theorem 2. Without loss of generality, we assume that $0 < \delta, r, s \le 1$, $0 < \gamma \le \frac{1}{4}$. For the sake of induction, we initially set $e_0 = e, r_0 = r, \beta_0 = s, \gamma_0 = 4\gamma, M_0 = M, \mathcal{O}_0 = \mathcal{O}, P_0 = P$. Denote

$$c^* = \sup\{|\omega| : \omega \in \mathcal{O}_0\}$$

and let M^*, M_* be fixed such that

$$\max_{l \in Z_+^n, |l| \le l_0} |\partial_\omega^l M_0|_{\mathcal{O}_0} \le M^*, \ \ |M_0^{-1}|_{\mathcal{O}_0} \le M_*.$$

By monotonicity, we define $0 < \mu_0 \le 1$ implicitly through the following equation:

$$\mu = \frac{4^{3b-1}\mu_0}{C_0^2([\log \frac{1}{\mu_0}] + 1)^{6\eta((n+2m)^2\tau + (n+2m)^2+1)}}, \tag{3.1}$$

where $[\,\cdot\,]$ denotes the integral part of a real number,

$$C_0 = (n+2m)^4 (c^*)^{(n+2m)^2} (M^*+1)^{(n+2m)^2},$$

and η is a fixed positive integer such that $(1+\sigma)^\eta > 2$ with $\sigma = \frac{1}{12}$. It is easy to see that $\mu_0 \to 0$ iff $\mu \to 0$, and, for any fixed $0 < \epsilon < 1$,

$$\mu_0 = o(\mu^{1-\epsilon}) \quad \text{as} \quad \mu \to 0. \tag{3.2}$$

Let

$$s_0 = \frac{\beta_0 \gamma_0^b}{2C_0 K_1^{(n+2m)^2\tau + (n+2m)^2+1}},$$

where,

$$K_1 = ([\log \frac{1}{\mu_0}] + 1)^{3\eta}.$$

With μ being sufficiently small, we can assume without loss of generality that

$$2C_0 K_1^{(n+2m)^2\tau + (n+2m)^2+1} > 1. \tag{3.3}$$

Hence,

$$0 < s_0 \le \min\{\beta_0, \frac{\gamma_0^b}{2C_0 K_1^{(n+2m)^2\tau+(n+2m)^2+1}}\}. \tag{3.4}$$

For $j \in Z_+^{n+2m}$, define

$$\begin{aligned}
a_j &= 1 - \mathrm{sgn}(|j|-1) = \begin{cases} 2, & |j| = 0, \\ 1, & |j| = 1, \\ 0, & |j| \ge 2, \end{cases} \\
b_j &= b(1 - \mathrm{sgn}(|j|)\mathrm{sgn}(|j|-1)\mathrm{sgn}(|j|-2)) = \begin{cases} b, & |j| = 0,1,2, \\ 0, & |j| \ge 3, \end{cases} \\
d_j &= 1 - \lambda_0 \mathrm{sgn}(|j|)\mathrm{sgn}(|j|-1)\mathrm{sgn}(|j|-2) = \begin{cases} 1, & |j| = 0,1,2, \\ 1-\lambda_0, & |j| \ge 3, \end{cases}
\end{aligned}$$

where $\frac{2}{13} < \lambda_0 < 1$ is fixed. Then

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P_0|_{D(r_0,s_0)\times\mathcal{O}_0} \le \delta\gamma_0^b s_0^{1-\mathrm{sgn}(|j|-1)} \mu_0 \le \delta \gamma_0^{b_j} s_0^{a_j} \mu_0^{d_j} \tag{3.5}$$

for all $(l,i,j) \in Z_+^n \times Z_+^n \times Z_+^{n+2m}$, $|l| + |i| + |j| \le l_0$.

As an induction hypothesis, we assume that, after a νth step, we have arrived at a Hamiltonian

$$\begin{aligned}
H &= H_\nu = N + P, \\
N &= N_\nu = e + \langle \omega, y\rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M \begin{pmatrix} y \\ z\end{pmatrix}\rangle
\end{aligned} \tag{3.6}$$

which is defined on a domain $D(r,s) = D(r_\nu, s_\nu)$ with $0 < s = s_\nu \le s_0$, $0 < r = r_\nu \le r_0$, where $M = M_\nu(\omega)$ is real symmetric, non-singular and smooth on a frequency domain $\mathcal{O} = \mathcal{O}_\nu \subset \mathcal{O}_0$, $e = e_\nu(\omega)$ is smooth on $\mathcal{O}$, and, $P = P_\nu(x,y,z,\omega)$ is analytic in $(x,y,z) \in D(r,s)$, smooth in $\omega \in \mathcal{O}$, and satisfies

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P|_{D(r,s)\times\mathcal{O}} \le \delta \gamma^{b_j} s^{a_j} \mu^{d_j} \tag{3.7}$$

for all $(l,i,j) \in Z_+^n \times Z_+^n \times Z_+^{n+2m}$, $|l|+|i|+|j| \le l_0$, with some $0 < \mu = \mu_\nu \le \mu_0$, $0 < \gamma = \gamma_\nu \le \gamma_0$.

The purpose of this section is to carry out the next cycle of KAM iterations, i.e., the $(\nu+1)$th step. More precisely, we shall seek for a Hamiltonian $F = F_\nu$ such that the time-1 map ϕ_F^1 of the vector field generated by F, as a symplectic transformation, will essentially transform $H = H_\nu$, on a smaller domain $D(r_+, s_+) \times \mathcal{O}_+ = D(r_{\nu+1}, s_{\nu+1}) \times \mathcal{O}_{\nu+1}$, to a new Hamiltonian

$$\begin{aligned} H_+ &= H_{\nu+1} = N_+ + P_+, \\ N_+ &= N_{\nu+1} = e_+ + \langle \omega, y\rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M_+ \begin{pmatrix} y \\ z \end{pmatrix}\rangle, \end{aligned}$$

where $e_+ = e_{\nu+1}(\omega), M_+ = M_{\nu+1}(\omega), P_+ = P_{\nu+1}(x, y, z, \omega)$ enjoy similar properties as $e = e_\nu, M = M_\nu, P = P_\nu$, respectively, and moreover,

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P_+|_{D(r_+,s_+)\times\mathcal{O}_+} \le \delta \gamma_+^{b_j} s_+^{a_j} \mu_+^{d_j}$$

for all $(l, i, j) \in Z_+^n \times Z_+^n \times Z_+^{n+2m}$, $|l| + |i| + |j| \le l_0$, with some smaller $0 < \mu_+ = \mu_{\nu+1} \le \mu_0$, $0 < \gamma_+ = \gamma_{\nu+1} \le \gamma_0$.

We shall give detailed estimates on the transformation, new normal form and perturbation etc for this cycle of KAM steps. For simplicity, in the rest of this section, we omit the index for all quantities at the νth step and use "+" to index all quantities at the $(\nu + 1)$th step.

Below, all constants c_1–c_{12} are positive and independent of the iteration process. We shall also use the same symbol c to denote any intermediate positive constant which is independent of the iteration process.

3.1. **Truncation of the perturbation.** As usual, the symplectic transformation at each KAM step should add the average of quadratic terms of P into the new normal form. Therefore, we need to separate the quadratic terms from P, which leads to a truncation process. Moreover, for the reminder of the truncation to qualify as part of the new perturbation, we also need to estimate the reminder term on a smaller domain.

Consider the Taylor-Fourier series of P:

$$P = \sum_{\imath \in Z_+^n, \jmath \in Z_+^{2m}, k \in Z^n} p_{k\imath\jmath} y^\imath z^\jmath e^{\sqrt{-1}\langle k, x\rangle},$$

and, let R be the truncation of P of form

$$\begin{aligned} R &= \sum_{|k|\le K_+, |\imath|+|\jmath|<3} p_{k\imath\jmath} y^\imath z^\jmath e^{\sqrt{-1}\langle k,x\rangle} \\ &= \sum_{|k|\le K_+} (P_{k00} + \langle P_{k10}, y\rangle + \langle P_{k01}, z\rangle \\ &+ \langle y, P_{k20} y\rangle + \langle z, P_{k11} y\rangle + \langle z, P_{k02} z\rangle) e^{\sqrt{-1}\langle k,x\rangle}, \end{aligned} \tag{3.8}$$

where P_{k20}, P_{k02} are symmetric matrices, and,

$$K_+ = ([\log \frac{1}{\mu}] + 1)^{3\eta}$$

with η being defined at the beginning of the section.

Denote

$$\begin{aligned} D_\alpha &= D(r_+ + \frac{7}{8}(r - r_+), \alpha s), \\ \hat{D}(\lambda) &= D(r_+ + \frac{7}{8}(r - r_+), \lambda), \\ D(\lambda) &= \{(y, z) \in C^n \times C^{2m} : |y| < \lambda, |z| < \lambda\}, \end{aligned}$$

where $\alpha = \mu^{\frac{1}{3}}$, $\lambda > 0$, and,

$$r_+ = \frac{r}{2} + \frac{r_0}{4}. \tag{3.9}$$

Lemma 3.1. *Assume that*

H1) $K_+ \geq \frac{8(n+l_0)}{r-r_+}$,

H2) $\int_{K_+}^{\infty} \lambda^{n+l_0} e^{-\lambda\frac{r-r_+}{8}} \mathrm{d}\lambda \leq \mu$.

Then there is a constant c_1 such that for all $|l|+|i|+|j| \leq l_0$, $\omega \in \mathcal{O}$,

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j (P-R)|_{D_\alpha} \leq c_1 \delta \gamma^{b_j} s^{a_j} \mu^{d_j+1}.$$

Proof. Without loss of generality, we let $\mu_0 \leq \frac{1}{8}$. Hence $\alpha \leq \frac{1}{2}$, and, $D_\alpha \subset \hat{D}(s) \subset D(r,s)$.

Define

$$\begin{aligned} p_k &= p_k(y,z) = \sum_{\imath \in Z_+^n, \jmath \in Z_+^{2m}} p_{k\imath\jmath} y^\imath z^\jmath, \\ I &= \sum_{|k|>K_+, \imath \in Z_+^n, \jmath \in Z_+^{2m}} p_{k\imath\jmath} y^\imath z^\jmath e^{\sqrt{-1}\langle k,x\rangle}, \\ II &= \sum_{|k|\leq K_+, |\imath|+|\jmath|\geq 3} p_{k\imath\jmath} y^\imath z^\jmath e^{\sqrt{-1}\langle k,x\rangle}. \end{aligned}$$

Then

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

and, it follows from the standard Cauchy estimate that

$$|\partial_\omega^l \partial_{(y,z)}^j p_k|_{D(s)} \leq |\partial_\omega^l \partial_{(y,z)}^j P|_{D(r,s)} e^{-|k|r} \leq \delta \gamma^{b_j} s^{a_j} \mu^{d_j} e^{-|k|r}$$

for all k.

Denote $\partial^{l,i,j} = \partial_\omega^l \partial_x^i \partial_{(y,z)}^j$. Since, by H1), the function $t^{n+l_0} e^{-t\frac{r-r_+}{8}}$ is strictly decreasing as $t \geq \frac{8(n+l_0)}{r-r_+}$, we have

$$\begin{aligned} |\partial^{l,i,j} I|_{\hat{D}(s)} &\leq c \sum_{|k|>K_+} |k|^{|i|} |\partial_\omega^l \partial_{(y,z)}^j p_k|_{D(s)} e^{|k|(r_+ + \frac{7}{8}(r-r_+))} \\ &\leq c\delta\gamma^{b_j} s^{a_j} \mu^{d_j} \sum_{|k|>K_+} |k|^{l_0} e^{-|k|\frac{r-r_+}{8}} \\ &\leq c\delta\gamma^{b_j} s^{a_j} \mu^{d_j} \sum_{\kappa=K_+}^{\infty} \kappa^{n+l_0} e^{-\kappa\frac{r-r_+}{8}} \\ &\leq c\delta\gamma^{b_j} s^{a_j} \mu^{d_j} \int_{K_+}^{\infty} \lambda^{n+l_0} e^{-\lambda\frac{r-r_+}{8}} \mathrm{d}\lambda \leq c\delta\gamma^{b_j} s^{a_j} \mu^{d_j+1}. \end{aligned} \tag{3.10}$$

This together with (3.7) implies that

$$|\partial^{l,i,j}(P-I)|_{\hat{D}(s)} \leq |\partial^{l,i,j} P|_{\hat{D}(s)} + |\partial^{l,i,j} I|_{\hat{D}(s)} \leq c\delta\gamma^{b_j} s^{a_j} \mu^{d_j}.$$

Next, for $p = (p_1, \cdots, p_n)^\top$, $q = (q_1, \cdots, q_{2m})^\top$ with $|p|+|q| = 3$, write

$$II = \int \frac{\partial^{(p,q)}}{\partial(y^p, z^q)} \sum_{|k|\leq K_+, |\imath|+|\jmath|\geq 3} p_{k\imath\jmath} e^{\sqrt{-1}\langle k,x\rangle} y^\imath z^\jmath \mathrm{d}y\mathrm{d}z,$$

where $\int = \int_0^{y_1^{p_1}} \cdots \int_0^{y_n^{p_n}} \int_0^{z_1^{q_1}} \cdots \int_0^{z_{2m}^{q_{2m}}}$.

Using the Cauchy estimate for $\partial^{l,i,j}(P-I)$ on $\hat{D}(s)$, we have

$$
\begin{aligned}
|\partial^{l,i,j} II|_{D_\alpha} &= |\partial^{l,i,j} \int \frac{\partial^{(p,q)}}{\partial(y^p, z^q)} \sum_{|k| \le K_+, |\imath|+|\jmath| \ge 3} p_{k\imath\jmath} e^{\sqrt{-1}\langle k,x\rangle} y^\imath z^\jmath \mathrm{d}y\mathrm{d}z|_{D_\alpha} \\
&\le |\int |\frac{\partial^{(p,q)}}{\partial y^p \partial z^q} \partial^{l,i,j}(P-I)|_{\hat{D}(s)} \mathrm{d}y\mathrm{d}z|_{D_\alpha} \\
&\le c\left(\frac{1}{(1-\alpha)s}\right)^3 \delta\gamma^{b_j} s^{a_j} \mu^{d_j} |\int \mathrm{d}y\mathrm{d}z|_{D_\alpha} \\
&\le 2^3 c\delta\gamma^{b_j} s^{a_j} \frac{(\alpha s)^3}{s^3} \mu^{d_j} = c\delta\gamma^{b_j} s^{a_j} \mu^{d_j+1}. \qquad (3.11)
\end{aligned}
$$

The lemma now follows from (3.10) and (3.11). □

3.2. **Quasi-linear equations.** We shall look for a Hamiltonian F which averages out the truncation R. The order of R in $y^\imath z^\jmath$ suggests that F should have the following form

$$
\begin{aligned}
F &= \sum_{0<|k|\le K_+, |\imath|+|\jmath|<3} f_{k\imath\jmath} y^\imath z^\jmath e^{\sqrt{-1}\langle k,x\rangle} \\
&= \sum_{0<|k|\le K_+} (F_{k00} + \langle F_{k10}, y\rangle + \langle F_{k01}, z\rangle \\
&+ \langle y, F_{k20} y\rangle + \langle z, F_{k11} y\rangle + \langle z, F_{k02} z\rangle) e^{\sqrt{-1}\langle k,x\rangle}, \qquad (3.12)
\end{aligned}
$$

where $F_{kij} = F_{kij}(y,z)$, $0 \le i+j \le 2$, are matrices or vectors of obvious dimensions. Note that, differing from usual KAM linear schemes, we have allowed the coefficients (matrices) in the above to depend on y and z.

Let ϕ_F^t be the flow generated by F. Then the time-1 map ϕ_F^1 will transform (3.6) into the following Hamiltonian

$$
\begin{aligned}
H \circ \phi_F^1 &= (N+R)\circ\phi_F^1 + (P-R)\circ\phi_F^1 \\
&= N + R + \{N, F\} + \int_0^1 \{R_t, F\} \circ \phi_F^t \mathrm{d}t + (P-R)\circ\phi_F^1, \qquad (3.13)
\end{aligned}
$$

where

$$R_t = \{(1-t)N, F\} + R.$$

The integral in (3.13) is of quadratic order in R, F, and, the term $(P-R)\circ\phi_F^1$ is of order $O(y^\imath z^\jmath)$, $|\imath| + |\jmath| \ge 3$, in y, z. These two terms will be treated as part of the new perturbation.

Write $M = M(\omega)$ into blocks

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where $M_{11}, M_{12}, M_{21}, M_{22}$ are $n\times n, n\times 2m, 2m\times n, 2m\times 2m$ minors of M respectively, and define

$$\begin{aligned} h(y,z) &= \frac{1}{2}\langle \begin{pmatrix} y\\ z\end{pmatrix}, M\begin{pmatrix} y\\ z\end{pmatrix}\rangle,\\ \Delta &= \Delta(y,z) = \delta\frac{\partial h}{\partial y}(y,z) = \delta(M_{11}y + M_{12}z). \end{aligned}$$

It is clear that

$$\{N,F\} = -\langle N_y, F_x\rangle + \delta\langle \frac{\partial h}{\partial z}, J\frac{\partial F}{\partial z}\rangle = -\sqrt{-1}\langle k, \omega+\Delta\rangle F + \delta\langle \frac{\partial h}{\partial z}, J\frac{\partial F}{\partial z}\rangle. \tag{3.14}$$

In the above and also below, J denotes symplectic matrices of appropriate dimensions which match the symplectic structure of the Hamiltonian.

Let

$$\begin{aligned} R' &= \delta\langle \frac{\partial h}{\partial z}, J\frac{\partial F}{\partial z}\rangle\\ &\quad -\delta\langle M_{21}y + M_{22}z, J\sum_{0<|k|\le K_+}(F_{k01} + F_{k11}y + F_{k02}z + F_{k02}^{\top}z\\ &\quad +F_{k00z} + F_{k10z}y + F_{k01z}z)e^{\sqrt{-1}\langle k,x\rangle}\rangle. \end{aligned} \tag{3.15}$$

Then R' is of order $O(y^{\imath}z^{\jmath})$, $|\imath| + |\jmath| \ge 3$, in y, z, which can be also included in the new perturbation. Thus, the main idea of the quasi-linear iterative scheme is to solve F through the following equation:

$$\{N,F\} + R - [R] - R' = 0, \tag{3.16}$$

where

$$[R] = \int_{T^n} R(x,\cdot)\mathrm{d}x.$$

If (3.16) is solvable, then (3.13) becomes

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

where

$$\bar{N}_+ = N + [R] \tag{3.17}$$

is essentially the new normal form (the resonant terms involved in $[R]$ can be transformed away), and,

$$\bar{P}_+ = \int_0^1 \{R_t, F\}\circ\phi_F^t\mathrm{d}t + (P-R)\circ\phi_F^1 + R' \tag{3.18}$$

essentially serves as the new perturbation. We note that, using (3.16), R_t can be rewritten as

$$R_t = (1-t)(R' - R + [R]) + R. \tag{3.19}$$

We now construct F of form (3.12) which satisfies (3.16). Substituting (3.8), (3.12), (3.14) and (3.15) into (3.16) yields

$$
\begin{aligned}
&-\sum_{0<|k|\le K_+}\sqrt{-1}\langle k,\omega+\Delta\rangle(F_{k00}+\langle F_{k10},y\rangle+\langle F_{k01},z\rangle\\
&+\langle y,F_{k20}y\rangle+\langle z,F_{k11}y\rangle+\langle z,F_{k02}z\rangle)e^{\sqrt{-1}\langle k,x\rangle}\\
&+\sum_{0<|k|\le K_+}\delta\langle M_{21}y+M_{22}z,J(F_{k01}+F_{k11}y+F_{k02}z+F_{k02}^{\top}z\\
&\quad+F_{k00z}+F_{k10z}y+F_{k01z}z)\rangle e^{\sqrt{-1}\langle k,x\rangle}\\
&=-\sum_{0<|k|\le K_+}(P_{k00}+\langle P_{k10},y\rangle+\langle P_{k01},z\rangle\\
&+\langle y,P_{k20}y\rangle+\langle z,P_{k11}y\rangle+\langle z,P_{k02}z\rangle)e^{\sqrt{-1}\langle k,x\rangle}.
\end{aligned}
\tag{3.20}
$$

By equating the coefficients in the above, we obtain, for each $0<|k|\le K_+$, the following family of *quasi-linear equations*:

$$(\sqrt{-1}\langle k,\omega+\Delta\rangle)F_{k00}=P_{k00}, \tag{3.21}$$
$$(\sqrt{-1}\langle k,\omega+\Delta\rangle I_{2m}-\delta M_{22}J)F_{k01}+\delta M_{22}JF_{k00z}=P_{k01}, \tag{3.22}$$
$$(\sqrt{-1}\langle k,\omega+\Delta\rangle)F_{k10}+\delta M_{12}J(F_{k00z}+F_{k01})=P_{k10}, \tag{3.23}$$
$$(\sqrt{-1}\langle k,\omega+\Delta\rangle)F_{k20}+\delta(F_{k11}+F_{k10z})JM_{21}=P_{k20}, \tag{3.24}$$
$$
\begin{aligned}
&(\sqrt{-1}\langle k,\omega+\Delta\rangle)F_{k11}-\delta M_{22}JF_{k11}-\delta M_{12}J(F_{k02}+F_{k02}^{\top}+F_{k01z}^{\top})\\
&\qquad+\delta F_{k10z}JM_{22}=P_{k11},
\end{aligned}
\tag{3.25}
$$
$$(\sqrt{-1}\langle k,\omega+\Delta\rangle)F_{k02}-\delta M_{22}JF_{k02}+\delta F_{k02}JM_{22}-\delta M_{22}JF_{k01z}^{\top}=P_{k02}. \tag{3.26}$$

Now, if the above quasi-linear equations are solvable, then their solutions F_{kij} will satisfy (3.20) and hence uniquely determine the Hamiltonian F. We note that, when substituting these solutions into the expression of F in (3.12), one can replace the matrices F_{k20}, F_{k02} by $\frac{1}{2}(F_{k20}+F_{k20}^{\top})$, $\frac{1}{2}(F_{k02}+F_{k02}^{\top})$, respectively, so that they become symmetric.

For any $p\times q$ matrix $A=(a_{ij})$, we let $T(A)$ denote the column vector formed by all row vectors of A, i.e.,

$$T(A)=(a_{11}\cdots a_{1q}\cdots a_{p1}\cdots a_{pq})^{\top}.$$

Define

$$F_k=\begin{pmatrix}T(F_{k20})\\T(F_{k11})\\T(F_{k02})\end{pmatrix},\quad P_k=\begin{pmatrix}T(P_{k20})\\T(P_{k11})\\T(P_{k02})\end{pmatrix}.$$

We can rewrite (3.21)-(3.26) equivalently into the following system form

$$
\begin{cases}
L_{0k}F_{k00}=P_{k00},\\
L_{1k}F_{k01}+\delta M_{22}JF_{k00z}=P_{k01},\\
L_{0k}F_{k10}-\delta M_{12}J(F_{k00z}+F_{k01})=P_{k10},\\
L_{2k}F_k=P_k+Q_k,
\end{cases}
\tag{3.27}
$$

where

$$L_{0k} = \sqrt{-1}\langle k, \omega + \Delta\rangle,$$
$$L_{1k} = \sqrt{-1}\langle k, \omega + \Delta\rangle I_{2m} - \delta M_{22} J,$$
$$L_{2k} = \sqrt{-1}\langle k, \omega + \Delta\rangle I_{(n+2m)^2} + \delta \bar{M}(\omega),$$

$\bar{M}(\omega)$ integrates the coefficients of F_k involving δ, and Q_k corresponds to the remaining terms in (3.24)-(3.26) which linearly depend on F_{k01z} and F_{k10z}.

It turns out that the entire system (3.27) can be solved successively on a further restricted frequency domain

$$\begin{aligned} \mathcal{O}_+ = \mathcal{O}(K_+) = \quad & \{\omega \in \mathcal{O} : \ |\langle k, \omega\rangle| > \frac{\gamma}{|k|^\tau}, \ |\det \bar{L}_{1k}| > \frac{\gamma^{2m}}{|k|^{2m\tau}}, \\ & |\det \bar{L}_{2k}| > \frac{\gamma^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}, \quad \text{for all } 0 < |k| \le K_+\}, \end{aligned} \tag{3.28}$$

where

$$\begin{aligned} \bar{L}_{1k} &= \sqrt{-1}\langle k, \omega\rangle I_{2m} - \delta M_{22} J, \\ \bar{L}_{2k} &= \sqrt{-1}\langle k, \omega\rangle I_{(n+2m)^2} + \delta \bar{M}(\omega). \end{aligned} \tag{3.29}$$

Define

$$\gamma_+ = \frac{\gamma_0}{4} + \frac{\gamma}{2}. \tag{3.30}$$

Lemma 3.2. *Assume that*

H3) $\max_{|l| \le l_0} |\partial_\omega^l M - \partial_\omega^l M_0|_{\mathcal{O}} \le \mu_0^{\frac{1}{2}}$;

H4) $2sC_0 K_+^{(n+2m)^2\tau + (n+2m)^2 + 1} < \gamma^b$.

Then the quasi-linear equations (3.21)-(3.26)*, or equivalently, the system* (3.27) *can be solved on* $\mathcal{O}_+$ *successively to obtain functions* $F_{k00}, F_{k01}, F_{k10}, F_k$, $0 < |k| \le K_+$, *which are smooth in* $\omega \in \mathcal{O}_+$ *and analytic in* $(y, z) \in D(s)$. *Moreover,*

$$\bar{F}_{kij}(\bar{y}, \bar{z}) = F_{-kij}(y, z), \tag{3.31}$$

for all $0 \le i + j \le 2, \ 0 < |k| \le K_+, \ (y, z) \in D(s)$.

Proof. Note by H3) that $|M|_{\mathcal{O}} \le M^* + 1$. For all $\omega \in \mathcal{O}_+$, $0 < |k| \le K_+$, we have by (3.28) and H4) that

$$|L_{0k}| = |\sqrt{-1}\langle k, \omega + \sqrt{-1}\langle k, \Delta\rangle| > \frac{\gamma}{|k|^\tau} - \delta C_0 s K_+ \ge \frac{\gamma}{2|k|^\tau}, \tag{3.32}$$

$$\begin{aligned} |\det L_{1k}| &= |\det(\bar{L}_{1k} + \sqrt{-1}\langle k, \Delta\rangle I_{2m})| \\ &\ge \frac{\gamma^{2m}}{|k|^{2m\tau}} - \delta C_0 s K_+^{2m} \ge \frac{\gamma^{2m}}{2|k|^{2m\tau}}, \end{aligned} \tag{3.33}$$

$$\begin{aligned} |\det L_{2k}| &= |\det(\bar{L}_{2k} + \sqrt{-1}\langle k, \Delta\rangle I_{(n+2m)^2})| \\ &> \frac{\gamma^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}} - \delta C_0 s K_+^{(n+2m)^2} \ge \frac{\gamma^{(n+2m)^2}}{2|k|^{(n+2m)^2\tau}}. \end{aligned} \tag{3.34}$$

In particular, L_{0k}, L_{1k}, L_{2k} are non-singular on $\mathcal{O}_+$. Therefore, the quasi-linear equations (3.21)-(3.26), or equivalently, the system (3.27) can be solved successively to obtain functions $F_{k00}, F_{k01}, F_{k10}, F_k$ with the same regularity properties as the coefficients of (3.27).

(3.31) easily follows from the uniqueness of solutions of (3.21)-(3.26) or (3.27). □

3.3. **Translation.** After finding the transformation F which transforms the the normal form N to $\bar{N}_+$, we need to introduce a translation of coordinate to further remove the first order resonant terms in $[R]$.

Consider the translation

$$\phi : x \to x, \ \begin{pmatrix} y \\ z \end{pmatrix} \to \begin{pmatrix} y \\ z \end{pmatrix} + \begin{pmatrix} y_0 \\ z_0 \end{pmatrix},$$

where

$$\begin{pmatrix} y_0 \\ z_0 \end{pmatrix} = -\frac{M^{-1}}{\delta} \begin{pmatrix} P_{010} \\ P_{001} \end{pmatrix}.$$

Then, ϕ transforms $H \circ \phi_F^1$ into the desired form:

$$H_+ = H \circ \phi_F^1 \circ \phi = N_+ + P_+,$$

where

$$\begin{aligned} N_+ &= \bar{N}_+ \circ \phi - 2\langle \begin{pmatrix} y \\ z \end{pmatrix}, \begin{pmatrix} P_{020} & P_{011} \\ P_{011}^\top & P_{002} \end{pmatrix} \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} \rangle \\ &= e_+ + \langle \omega, y \rangle + \frac{\delta}{2} \langle \begin{pmatrix} y \\ z \end{pmatrix}, M_+ \begin{pmatrix} y \\ z \end{pmatrix} \rangle, \end{aligned}$$

$$P_+ = \bar{P}_+ \circ \phi + 2\langle \begin{pmatrix} y \\ z \end{pmatrix}, \begin{pmatrix} P_{020} & P_{011} \\ P_{011}^\top & P_{002} \end{pmatrix} \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} \rangle, \tag{3.35}$$

with

$$\begin{aligned} e_+ &= e + P_{000} + \langle \omega, y_0 \rangle + \frac{1}{2}(\langle P_{010}, y_0 \rangle + \langle P_{001}, z_0 \rangle) \\ &\quad + \langle \begin{pmatrix} y_0 \\ z_0 \end{pmatrix}, \begin{pmatrix} P_{020} & P_{011} \\ P_{011}^\top & P_{002} \end{pmatrix} \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} \rangle, \end{aligned} \tag{3.36}$$

$$M_+ = M + \frac{2}{\delta} \begin{pmatrix} P_{020} & P_{011} \\ P_{011}^\top & P_{002} \end{pmatrix}. \tag{3.37}$$

Having defined M_+, we can define $\bar{M}_+ = \bar{M}_+(\omega)$ similarly to $\bar{M}$. By Lemma 3.2, in its domain of definition, F is well defined, real analytic in (x, y, z), and smooth in ω. Consequently, the same holds for $\Phi_+ = \phi_F^1 \circ \phi$, N_+ and P_+. Moreover, we have the following.

Lemma 3.3. *Assume* H3). *Then there is a constant c_2 such that for all $|l| \le l_0$*

$$\begin{aligned} |\partial_\omega^l e_+ - \partial_\omega^l e|_{\mathcal{O}} &\le c_2 \gamma^b s \mu, \\ |\partial_\omega^l M_+ - \partial_\omega^l M|_{\mathcal{O}} &\le c_2 \gamma^b \mu, \\ |\partial_\omega^l \bar{M}_+ - \partial_\omega^l \bar{M}|_{\mathcal{O}} &\le c_2 \gamma^b \mu, \\ |\begin{pmatrix} \partial_\omega^l y_0 \\ \partial_\omega^l z_0 \end{pmatrix}|_{\mathcal{O}} &\le c_2 \gamma^b s \mu. \end{aligned}$$

Proof. By (3.7), we immediately have

$$\begin{aligned} |\partial_\omega^l P_{000}|_{\mathcal{O}} &\le c\delta\gamma^b s^2 \mu, \\ |\partial_\omega^l P_{010}|_{\mathcal{O}} + |\partial_\omega^l P_{001}|_{\mathcal{O}} &\le c\delta\gamma^b s \mu, \\ |\begin{pmatrix} \partial_\omega^l P_{020} & \partial_\omega^l P_{011} \\ (\partial_\omega^l P_{011})^\top & \partial_\omega^l P_{002} \end{pmatrix}|_{\mathcal{O}} &\le c\delta\gamma^b \mu. \end{aligned}$$

Without loss of generality, we let μ_0 be small such that $\mu_0^{\frac{1}{2}} M_* \leq \frac{1}{2}$. Then by H3),

$$|M^{-1}|_{\mathcal{O}} \leq \frac{|M_0^{-1}|_{\mathcal{O}_0}}{1-|M-M_0|_{\mathcal{O}}|M_0^{-1}|_{\mathcal{O}_0}} \leq \frac{M_*}{1-\mu_0^{\frac{1}{2}} M_*} \leq 2M_*.$$

It follows from induction that

$$|\partial_\omega^{l'} M^{-1}|_{\mathcal{O}} \leq c|M^{-1}|_{\mathcal{O}}^{l'+1}(M^*+1) \leq cM_*(M^*+1), \ |l'| \leq l_0.$$

Therefore,

$$|\begin{pmatrix} \partial_\omega^l y_0 \\ \partial_\omega^l z_0 \end{pmatrix}|_{\mathcal{O}} \leq \frac{c}{\delta} \sum_{|l'|=0}^{|l|} \begin{pmatrix} l \\ l' \end{pmatrix} |\partial_\omega^{l-l'} M^{-1}|_{\mathcal{O}} |\begin{pmatrix} \partial_\omega^{l'} P_{010} \\ \partial_\omega^{l'} P_{001} \end{pmatrix}|_{\mathcal{O}} \leq c\gamma^b s\mu,$$

which, together with (3.36), (3.37), also implies the first two inequalities of the lemma. □

3.4. **Estimate on the new frequency domain.** For $k \in Z^n \setminus \{0\}$, $\omega \in \mathcal{O}_+$, we define

$$\begin{aligned} \bar{L}_{1k}^+ &= \sqrt{-1}\langle k, \omega \rangle I_{2m} - \delta M_{22}^+ J, \\ \bar{L}_{2k}^+ &= \sqrt{-1}\langle k, \omega \rangle I_{(n+2m)^2} + \delta \bar{M}_+(\omega), \end{aligned} \tag{3.38}$$

where M_{22}^+ is the right lower $2m \times 2m$ minor of M_+.

Lemma 3.4. *Assume that*

H5) $c_2 \mu K_+^{(n+2m)^2\tau+(n+2m)^2+1} < \min\{\frac{\gamma-\gamma_+}{\gamma_0}, \frac{\gamma^{2m}-\gamma_+^{2m}}{\gamma_0^{2m}}, \frac{\gamma^{(n+2m)^2}-\gamma_+^{(n+2m)^2}}{\gamma_0^{(n+2m)^2}}\}$.

Then for all $0 < |k| \leq K_+$, $\omega \in \mathcal{O}_+$,

$$|\langle k, \omega \rangle| > \frac{\gamma_+}{|k|^\tau}, \quad |\det \bar{L}_{1k}^+| > \frac{\gamma_+^{2m}}{|k|^{2m\tau}}, \quad |\det \bar{L}_{2k}^+| > \frac{\gamma_+^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}.$$

Proof. Let $0 < |k| \leq K_+$, $\omega \in \mathcal{O}_+$. By (3.28),

$$|\langle k, \omega \rangle| > \frac{\gamma}{|k|^\tau}, \ |\det \bar{L}_{1k}| > \frac{\gamma^{2m}}{|k|^{2m\tau}}, \ |\det \bar{L}_{2k}| > \frac{\gamma^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}},$$

and by H5),

$$\begin{aligned} &\delta c_2 \gamma^b \mu K_+^{\tau+1} < c_2 \gamma_0 \mu K_+^{\tau+1} < \gamma - \gamma_+, \\ &\delta c_2 \gamma^b \mu K_+^{2m\tau+2m} < c_2 \gamma_0^{2m} \mu K_+^{2m\tau+2m} < \gamma^{2m} - \gamma_+^{2m}, \\ &\delta c_2 \gamma^b \mu K_+^{(n+2m)^2\tau+(n+2m)^2} < c_2 \gamma_0^{(n+2m)^2} \mu K_+^{(n+2m)^2\tau+(n+2m)^2} \\ &\qquad < \gamma^{(n+2m)^2} - \gamma_+^{(n+2m)^2}. \end{aligned}$$

Since

$$\bar{L}_{1k}^+ - \bar{L}_{1k} = \delta(M_{22} - M_{22}^+)J, \ \bar{L}_{2k}^+ - \bar{L}_{2k} = \delta(\bar{M}_+ - \bar{M}_+),$$

the lemma follows from Lemma 3.3 and a similar argument as that for (3.32)-(3.34). □

3.5. **Estimate on F.** We now give some estimates on F and its derivatives, which are vital later in proving the convergence of the transformation sequence and in estimating the new perturbation. Define

$$\Gamma(r-r_+) = \sum_{0<|k|\le K_+} |k|^{\chi} e^{-|k|\frac{r-r_+}{8}},$$

where $\chi = b\tau + 5l_0 + 2\tau + 10$. We note that $\Gamma(r-r_+) \ge e^{-\frac{r_0}{8}}$.

Lemma 3.5. *Assume* H3), H4). *Then the following holds for all* $0 < |k| \le K_+$.

1) *There is a constant c_3 such that, on $D(s)\times\mathcal{O}_+$,*

$$\begin{aligned}
|\frac{\partial^{(\imath,\jmath,l)} F_{k00}}{\partial(y^\imath, z^\jmath, \omega^l)}| &\le c_3|k|^{(|\imath|+|\jmath|+|l|+1)\tau+|\imath|+|\jmath|+|l|}\delta s^2\mu e^{-|k|r}, \\
&\qquad |\imath|+|\jmath|\le l_0+4,\ |l|\le l_0, \\
|\frac{\partial^{(\imath,\jmath,l)} F_{k01}}{\partial(y^\imath, z^\jmath, \omega^l)}| &\le c_3|k|^{(|\imath|+|\jmath|+|l|+1)2m\tau+2(|\imath|+|\jmath|+|l|)+2\tau+2}\delta s\mu e^{-|k|r}, \\
&\qquad |\imath|+|\jmath|\le l_0+3,\ |l|\le l_0, \\
|\frac{\partial^{(\imath,\jmath,l)} F_{k10}}{\partial(y^\imath, z^\jmath, \omega^l)}| &\le c_3|k|^{(|\imath|+|\jmath|+|l|+1)2m\tau+2(|\imath|+|\jmath|+|l|)+2\tau+2}\delta s\mu e^{-|k|r}, \\
&\qquad |\imath|+|\jmath|\le l_0+3,\ |l|\le l_0, \\
|\frac{\partial^{(\imath,\jmath,l)} F_{k}}{\partial(y^\imath, z^\jmath, \omega^l)}| &\le c_3|k|^{(|\imath|+|\jmath|+|l|+1)(n+2m)^2\tau+2(|\imath|+|\jmath|+|l|)+2\tau+5}\delta\mu e^{-|k|r}, \\
&\qquad |\imath|+|\jmath|\le l_0+2,\ |l|\le l_0.
\end{aligned}$$

2) *There is a constant c_4 such that, on $\hat{D}(s)\times\mathcal{O}_+$,*

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j F| \le c_4\delta s^{a_j}\mu\Gamma(r-r_+),\ |l|\le l_0,\ |i|+|j|\le l_0+2. \tag{3.39}$$

Proof. Let $(y,z)\in D(s)$, $\omega\in\mathcal{O}_+$.

1) Denote $L_k = L_{0k}, L_{1k}, L_{2k}$ and $q=(\imath,\jmath,l)\in Z_+^n\times Z_+^{2m}\times Z_+^n$, where $|\imath|+|\jmath|\le l_0+4$, $|l|\le l_0$. By H3), it is easy to see that

$$|\partial^q L_k| \le c|k|$$

for all $|k|\ge 1$, $|q|\ge 0$. Applying the above and the inequalities

$$|\partial^{q'} L_k^{-1}| \le |L_k^{-1}| \sum_{|q''|=1}^{|q'|} \binom{q'}{q''} |\partial^{q'-q''} L_k^{-1}||\partial^{q''} L_k|,\ |q''|\le|q'|$$

inductively, we deduce that

$$|\partial^q L_k^{-1}| \le c|k|^{|q|}|L_k^{-1}|^{|q|+1}.$$

Using (3.32)-(3.34) and the identities

$$L_{1k}^{-1} = \frac{\mathrm{adj}(L_{1k})}{\det(L_{1k})},\ \ {L_{2k}}^{-1} = \frac{\mathrm{adj}(L_{2k})}{\det(L_{2k})},$$

we then have

$$
\begin{aligned}
|\partial^q L_{0k}^{-1}| &\leq c\frac{|k|^{(|q|+1)\tau+|q|}}{\gamma^{|q|+1}},\\
|\partial^q L_{1k}^{-1}| &\leq c\frac{|k|^{(|q|+1)2m\tau+2|q|+1}}{\gamma^{2m|q|+2m}},\\
|\partial^q L_{2k}{}^{-1}| &\leq c\frac{|k|^{(|q|+1)(n+2m)^2\tau+2|q|+1}}{\gamma^{(n+2m)^2|q|+(n+2m)^2}}.
\end{aligned}
$$

It now follows from (3.21) and the Cauchy estimate that

$$
\begin{aligned}
|\frac{\partial^{(\imath,\jmath,l)}F_{k00}}{\partial(y^\imath,z^\jmath,\omega^l)}| &\leq c\sum_{|l'|=1}^{|l|}\binom{l}{l'}|\partial^{(\imath,\jmath,l-l')}L_{0k}^{-1}||\partial_\omega^{l'}P_{k00}|\\
&\leq c|k|^{(|\imath|+|\jmath|+|l|+1)\tau+|\imath|+|\jmath|+|l|}\delta s^2\mu e^{-|k|r},\\
&\qquad |\imath|+|\jmath|\leq l_0+4,\ |l|\leq l_0.
\end{aligned}
\tag{3.40}
$$

Using (3.22) and (3.40), we also have

$$
\begin{aligned}
|\frac{\partial^{(\imath,\jmath,l)}F_{k01}}{\partial(y^\imath,z^\jmath,\omega^l)}| &\leq c|k|^{(|\imath|+|\jmath|+|l|+1)2m\tau+2(|\imath|+|\jmath|+|l|)+2\tau+2}\delta s\mu e^{-|k|r},\\
&\qquad |\imath|+|\jmath|\leq l_0+3,\ |l|\leq l_0.
\end{aligned}
\tag{3.41}
$$

Similarly, (3.23), (3.40) and (3.41) yield

$$
\begin{aligned}
|\frac{\partial^{(\imath,\jmath,l)}F_{k10}}{\partial(y^\imath,z^\jmath,\omega^l)}| &\leq c|k|^{(|\imath|+|\jmath|+|l|+1)2m\tau+2(|\imath|+|\jmath|+|l|)+2\tau+2}\delta s\mu e^{-|k|r},\\
&\qquad |\imath|+|\jmath|\leq l_0+3,\ |l|\leq l_0.
\end{aligned}
\tag{3.42}
$$

Finally, combining (3.34), (3.41), (3.42) with the last equation in (3.27), we obtain

$$
\begin{aligned}
|\frac{\partial^{(\imath,\jmath,l)}F_{k}}{\partial(y^\imath,z^\jmath,\omega^l)}| &\leq c|k|^{(|\imath|+|\jmath|+|l|+1)(n+2m)^2\tau+2(|\imath|+|\jmath|+|l|)+2\tau+5}\delta\mu e^{-|k|r},\\
&\qquad |\imath|+|\jmath|\leq l_0+2,\ |l|\leq l_0.
\end{aligned}
$$

2) By 1) and direct differentiations to (3.12), we have, on $\hat{D}(s)\times\mathcal{O}_+$, that

$$
\begin{aligned}
|\partial_\omega^l\partial_x^i\partial_{(y,z)}^jF| &\leq c\sum_{|\imath|+|\jmath|\leq|j|,0<|k|\leq K_+}|k|^{|i|}(|\frac{\partial^{(\imath,\jmath,l)}F_{k00}}{\partial(y^\imath,z^\jmath,\omega^l)}|\\
&\quad+|\frac{\partial^{(\imath,\jmath,l)}F_{k01}}{\partial(y^\imath,z^\jmath,\omega^l)}|s^{1-\mathrm{sgn}(|j|)}+|\frac{\partial^{(\imath,\jmath,l)}F_{k10}}{\partial(y^\imath,z^\jmath,\omega^l)}|s^{1-\mathrm{sgn}(|j|)}\\
&\quad+|\frac{\partial^{(\imath,\jmath,l)}F_{k}}{\partial(y^\imath,z^\jmath,\omega^l)}|s^{1-\mathrm{sgn}(|j|-1)})e^{|k|(r_++\frac{7}{8}(r-r_+))}\\
&\leq c\delta s^{a_j}\mu\sum_{0<|k|\leq K_+}|k|^\chi e^{-|k|\frac{r-r_+}{8}}=c\delta s^{a_j}\mu\Gamma(r-r_+).
\end{aligned}
$$

This proves the Lemma. □

To obtain the symplectic transformation stated in Theorem 2, we need to extend the function F smoothly to the domain $\hat{D}(\beta_0)\times\mathcal{O}_0$.

Lemma 3.6. *Assume* H3), H4)*. Then F and $\Theta = \binom{y_0}{z_0}$ can be smoothly extended to functions of Hölder class $C^{l_0+\sigma_0+1,l_0-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)$ and $C^{l_0-1+\sigma_0}(\mathcal{O}_0)$ respectively, where $0<\sigma_0<1$ is fixed. Moreover, there is a constant c_5 such that*

$$\|F\|_{C^{l_0+1+\sigma_0,l_0-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)} \le c_5\delta\mu\Gamma(r-r_+),$$
$$\|\Theta\|_{C^{l_0-1+\sigma_0}(\mathcal{O}_0)} \le c_5\gamma^b\mu.$$

Proof. By the standard Whitney extension theorem (see [11], [14]), F, Θ can be extended smoothly to functions $\tilde{F}$, $\tilde{\Theta}$ of Hölder class $C^{l_0+\sigma_0+1,l_0-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)$, $C^{l_0-1+\sigma_0}(\mathcal{O}_0)$, respectively, such that

$$\|\tilde{F}\|_{C^{l_0+1+\sigma_0,l_0-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)} \le \bar{c}\|F\|_{C^{l_0+2,l_0}(\hat{D}(s)\times\mathcal{O}_+)},$$
$$\|\tilde{\Theta}\|_{C^{l_0-1+\sigma_0}(\mathcal{O}_0)} \le \bar{c}\|\Theta\|_{C^{l_0}(\mathcal{O}_+)}, \tag{3.43}$$

where $\bar{c}$ is a constant depending only on the regularity orders l_0,σ_0 and the dimensions n,m (in particular, not on the iteration process). The lemma now follows from Lemma 3.5 2) and Lemma 3.3. □

3.6. **Estimate on the transformation.** Define

$$D_{\frac{i}{8}\alpha} = D(r_+ + \frac{i-1}{8}(r-r_+), \frac{i}{8}\alpha s),\ i=1,2,\cdots,8.$$

Lemma 3.7. *Assume* H3), H4) *and also that*

H6) $c_4\mu\Gamma(r-r_+) < \frac{1}{8}(r-r_+)$;
H7) $c_4\mu\Gamma(r-r_+) < \frac{1}{8}\alpha$;
H8) $c_2\mu < \frac{1}{8}\alpha$.

Then the following holds.

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

$$\phi_F^t \ :\ D_{\frac{1}{4}\alpha} \longrightarrow D_{\frac{1}{2}\alpha}, \tag{3.44}$$
$$\phi \ :\ D_{\frac{1}{8}\alpha} \to D_{\frac{1}{4}\alpha} \tag{3.45}$$

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

2) *There is a constant c_6 such that for all $0\le t\le 1$, $|l|\le l_0$,*

$$|\partial_\omega^l\partial_x^i\partial_{(y,z)}^j\phi_F^t|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+} \le \begin{cases} c_6 s\mu\Gamma(r-r_+), & |i|+|j|=0,\ |l|\ge 1;\\ c_6\mu\Gamma(r-r_+), & 2\le |l|+|i|+|j|\le l_0+2;\\ c_6, & \text{otherwise.}\end{cases}$$

Proof. 1) (3.45) follows immediately from Lemma 3.3 and H8).

To show (3.44), we write $\phi_F^t = (\phi_1^t,\phi_2^t,\phi_3^t)^\top$, where $\phi_1^t,\phi_2^t,\phi_3^t$ are components of ϕ_F^t in the directions of x,y,z respectively. Let (x,y,z) be any point in $D_{\frac{1}{4}\alpha}$ and let $t_* = \sup\{t\in[0,1] : \phi_F^t(x,y,z)\in D_\alpha\}$. We note that $D_\alpha\subset\hat{D}(s)$. Using Lemma 3.5 2), H6), H7) and the identity

$$\phi_F^t = \mathrm{id} + \int_0^t X_F\circ\phi_F^\lambda \mathrm{d}\lambda, \tag{3.46}$$

where $X_F = (F_y, -F_x, JF_z)^\top$ denotes the vector field generated by F, we have

$$\begin{aligned}
|\phi_1^t(x,y,z) - x| &\leq \int_0^t |F_y \circ \phi_F^\lambda|_{D_\alpha} \mathrm{d}\lambda \leq |F_y|_{\hat{D}(s)} < c_4 s\mu\Gamma(r - r_+) < \frac{1}{8}(r - r_+), \\
|\phi_2^t(x,y,z) - y| &\leq \int_0^t |F_x \circ \phi_F^\lambda|_{D_\alpha} \mathrm{d}\lambda \leq |F_x|_{\hat{D}(s)} < c_4 s^2\mu\Gamma(r - r_+) < \frac{1}{8}\alpha s, \\
|\phi_3^t(x,y,z) - z| &\leq \int_0^t |JF_z \circ \phi_F^\lambda|_{D_\alpha} \mathrm{d}\lambda \leq |F_z|_{\hat{D}(s)} < c_4 s\mu\Gamma(r - r_+) < \frac{1}{8}\alpha s
\end{aligned}$$

for all $\omega \in \mathcal{O}_+$, $0 \leq t \leq t_*$. It follows that

$$\begin{aligned}
&|\phi_1^t(x,y,z)| < r_+ + \frac{1}{4}(r - r_+) < r_+ + \frac{3}{8}(r - r_+), \\
&|\phi_2^t(x,y,z)|,\ |\phi_3^t(x,y,z)| < \frac{3}{8}\alpha s < \frac{1}{2}\alpha s,
\end{aligned}$$

i.e., $\phi_F^t(x,y,z) \in D_{\frac{1}{2}\alpha} \subset D_\alpha$ for all $\omega \in \mathcal{O}_+$, $0 \leq t \leq t_*$. Thus, $t_* = 1$ and (3.44) holds.

The regularity properties of ϕ_F^t, ϕ follow from Lemmas 3.2, 3.3.

By differentiating (3.46) and applying the Gronwall inequality inductively, 2) follows from H7) and Lemma 3.5. □

Define

$$\beta_+ = \frac{\beta}{2} + \frac{\beta_0}{4}.$$

Lemma 3.8. *Assume* H3),H4),H6)-H8) *and also that*

H9) $c_5\mu\Gamma(r - r_+) < \frac{1}{8}(r - r_+)$;

H10) $c_5\mu(\Gamma(r - r_+) + 1) < \beta - \beta_+$.

Let F, $\Theta = \binom{y_0}{z_0}$ *be the extended functions defined in* Lemma 3.6. *Then*

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

is of class $C^{l_0+\sigma_0}$ *and also depends* $C^{l_0-1+\sigma_0}$ *smoothly on* $\omega \in \mathcal{O}_0$, *where* σ_0 *is as in* Lemma 3.6. *Moreover, there is a constant* c_7 *such that*

$$\|\Phi_+ - id\|_{C^{l_0+\sigma_0, l_0-1+\sigma_0}(\hat{D}(\beta_+)\times\mathcal{O}_0)} \leq c_7\mu\Gamma(r - r_+).$$

Proof. Denote

$$\tilde{D}(\beta_+) = \hat{D}(\beta_+ + c_5\delta\mu).$$

Using H9), H10), Lemma 3.6 and a similar argument as in Lemma 3.7, it is easy to see that for all $0 \leq t \leq 1$,

$$\begin{aligned}
&\phi_F^t : \tilde{D}(\beta_+) \to D(r, \beta), \\
&\phi : \hat{D}(\beta_+) \to \tilde{D}(\beta_+)
\end{aligned}$$

are well defined, and, ϕ_F^t is of class $C^{l_0+\sigma_0}$ and also depends $C^{l_0-1+\sigma_0}$ smoothly on $\omega \in \mathcal{O}_0$. Moreover,

$$|\phi_F^t - id|_{\tilde{D}(\beta_+)\times\mathcal{O}_0} \leq \int_0^t |X_F \circ \phi_F^t|_{\hat{D}(\beta_0)\times\mathcal{O}_0} \mathrm{d}\lambda \leq |\partial_\xi F|_{\hat{D}(\beta_0)\times\mathcal{O}_0} \leq c_5\mu\Gamma(r - r_+),$$

where $\xi = (x, y, z)$.

Differentiating (3.46) with respect to ξ yields

$$\begin{aligned}\partial_\xi \phi_F^t &= I_{2n+2m} + \int_0^t DX_F \circ \phi_F^\lambda \partial_\xi \phi_F^\lambda \mathrm{d}\lambda \\ &= I_{2n+2m} + \int_0^t J(\partial_\xi^2 F) \circ \phi_F^\lambda \partial_\xi \phi_F^\lambda \mathrm{d}\lambda. \qquad (3.48)\end{aligned}$$

By applying Lemma 3.6 and the Gronwall inequality to (3.48), we then have

$$\begin{aligned}|\partial_\xi \phi_F^t - I_{2n+2m}|_{\tilde{D}(\beta_+)\times\mathcal{O}_0} &\le \int_0^t |\partial_\xi^2 F|_{\hat{D}(\beta_0)\times\mathcal{O}_0} |\partial_\xi \phi_F^\lambda - I_{2n+2m}|_{\tilde{D}(\beta_+)\times\mathcal{O}_0} \mathrm{d}\lambda \\ &\quad + |\partial_\xi^2 F|_{\hat{D}(\beta_0)\times\mathcal{O}_0} e^{|\partial_\xi^2 F|_{\hat{D}(\beta_0)\times\mathcal{O}_0}} \\ &\le c_5 \mu \Gamma(r-r_+) e^{c_5 \mu \Gamma(r-r_+)} \\ &\le c\mu\Gamma(r-r_+), \ \ 0 \le t \le 1.\end{aligned}$$

Note that

$$\|X_F\|_{C^{|i'|+\sigma_0, |l'|-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)} \le c\|F\|_{C^{|i'|+1+\sigma_0, |l'|-1+\sigma_0}(\hat{D}(\beta_0)\times\mathcal{O}_0)}$$

for all $0 \le |i'| \le l_0, 1 \le |l'| \le l_0$. By using Lemma 3.6, the Gronwall inequality and induction, one can prove similarly to the above that

$$\|\phi_F^t\|_{C^{|i|+\sigma_0, |l|-1+\sigma_0}(\tilde{D}(\beta_+)\times\mathcal{O}_0)} \le c\mu\Gamma(r-r_+) \qquad (3.49)$$

for all $2 \le |i| \le l_0, 1 \le |l| \le l_0, \ 0 \le t \le 1$.

The lemma now follows from H9), Lemma 3.3 and the identity

$$\Phi_+ - id = (\phi_F^1 - id) \circ \phi + \begin{pmatrix} 0 \\ y_0 \\ z_0 \end{pmatrix}.$$

□

3.7. **Estimate on the new perturbation.** It remains to estimate the new perturbation P_+ on the domain $D_+ \times \mathcal{O}_+$, where

$$D_+ = D_{\frac{1}{8}\alpha}.$$

Lemma 3.9. *Assume* H1)-H4) *and* H6)-H8). *Then there is a constant c_8 such that, on $D_+ \times \mathcal{O}_+$,*

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P_+| \le c_8 \delta \gamma^{b_j} s^{a_j} \alpha^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)} \mu \Gamma^3(r-r_+), \ \ |l|+|i|+|j| \le l_0.$$

Proof. Denote $\partial^{l,i,j} = \partial_\omega^l \partial_x^i \partial_{(y,z)}^j$ for $|l|+|i|+|j| \le l_0$. By Lemma 3.5 and the expressions of R, R' in (3.8), (3.15) respectively, we have

$$\begin{aligned}&|\partial^{l,i,j} R|_{\hat{D}(s)\times\mathcal{O}_+} \le c\delta\gamma^b s^{a_j} \mu\Gamma(r-r_+), \\ &|\partial^{l,i,j} R'|_{\hat{D}(s)\times\mathcal{O}_+} \le c\delta s^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)} \mu\Gamma(r-r_+), \\ &|\partial^{l,i,j} R'|_{D_{\frac{1}{2}\alpha}\times\mathcal{O}_+} \le c\delta(\alpha s)^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)} \mu\Gamma(r-r_+). \qquad (3.50)\end{aligned}$$

It follows from Lemma 3.5 2) and Lemma 3.7 2) that, for all $0 \le t \le 1$,

$$
\begin{aligned}
&|\partial^{l,i,j}\{[R],F\}|_{\hat{D}(s)\times\mathcal{O}} + |\partial^{l,i,j}\{R,F\}|_{\hat{D}(s)\times\mathcal{O}} \le c\delta\gamma^{b}s^{a_j}\mu^2\Gamma^2(r-r_+),\\
&|\partial^{l,i,j}\{[R],F\}\circ\phi_F^t|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+} + |\partial^{l,i,j}\{R,F\}\circ\phi_F^t|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+}\\
&\qquad\qquad\qquad \le c\delta\gamma^{b}s^{a_j}\mu^2\Gamma^3(r-r_+),\\
&|\partial^{l,i,j}\{R',F\}|_{\hat{D}(s)\times\mathcal{O}} \le c\delta s^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)}\mu^2\Gamma^2(r-r_+),\\
&|\partial^{l,i,j}\{R',F\}\circ\phi_F^t|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+} \le c\delta s^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)}\mu^2\Gamma^3(r-r_+).
\end{aligned}
$$

Hence, by (3.19) and the fact that $s < \gamma^b$ (H4)), we have

$$
\begin{aligned}
&|\partial^{l,i,j}\int_0^1\{R_t,F\}\circ\phi_F^t\mathrm{d}t|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+}\\
&\le c\delta\gamma^{b_j}s^{a_j}\mu^2\Gamma^3(r-r_+). \qquad (3.51)
\end{aligned}
$$

Recall from Lemma 3.1 that

$$|\partial^{l,i,j}(P-R)|_{D_\alpha\times\mathcal{O}_+} \le \delta\gamma^{b_j}s^{a_j}\mu^{d_j+1}.$$

This together with Lemma 3.7 2) implies that

$$|\partial^{l,i,j}(P-R)\circ\Phi_+|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+} \le c\delta\gamma^{b_j}s^{a_j}\mu^{d_j+1}\Gamma(r-r_+). \qquad (3.52)$$

Using (3.18), (3.50), (3.51) and (3.52), we then have

$$
\begin{aligned}
|\partial^{l,i,j}\bar{P}_+|_{D_{\frac{1}{4}\alpha}\times\mathcal{O}_+} &\le c\delta\gamma^{b_j}s^{a_j}(\mu^{d_j+1}+\alpha^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)}\mu)\Gamma^3(r-r_+)\\
&\le c\delta\gamma^{b_j}s^{a_j}\alpha^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)}\mu\Gamma^3(r-r_+). \qquad (3.53)
\end{aligned}
$$

Since, by Lemma 3.3,

$$|\partial_\omega^l\phi|_{\mathcal{O}_+} \le c_2\gamma^b s\mu,$$

it follows from (3.53) that

$$|\partial^{l,i,j}\bar{P}_+\circ\phi|_{D_{\frac{1}{8}\alpha}\times\mathcal{O}_+} \le c\delta\gamma^{b_j}s^{a_j}\alpha^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)}\mu\Gamma^3(r-r_+),$$

and, from (3.7) and Lemma 3.3 that

$$|\partial^{l,i,j}\langle\begin{pmatrix}y\\z\end{pmatrix},\begin{pmatrix}P_{020} & P_{011}\\ P_{011}^\top & P_{002}\end{pmatrix}\begin{pmatrix}y_0\\z_0\end{pmatrix}\rangle|_{D_{\frac{1}{8}\alpha}\times\mathcal{O}_+} \le c\delta\gamma^b s^{a_j}\mu^2.$$

The lemma now follows from (3.35). □

Finally, let

$$s_+ = \frac{1}{8}\alpha s, \quad \mu_+ = (64c_0)^{\frac{1}{1-\lambda_0}}\mu^{1+\sigma},$$

where $c_0 = \max\{1, c_1, \cdots, c_8\}$ and $\sigma = \frac{1}{12}$. We note that $D(r_+, s_+) = D_+$, and, c_0 only depends on r_0, β_0, l_0. If we further assume that

H11) $\mu^\sigma\Gamma^3(r-r_+) \le \dfrac{\gamma_+^{b_j}}{\gamma^{b_j}}$,

then, on $D_+\times\mathcal{O}_+$,

$$
\begin{aligned}
|\partial_\omega^l\partial_x^i\partial_{(y,z)}^j P_+| &\le 2^{3a_j}c_0\delta\gamma^{b_j}s_+^{a_j}\alpha^{2-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)-a_j}\mu^{1-\sigma}\mu^\sigma\Gamma^2(r-r_+)\\
&\le 64c_0\delta(\gamma^{b_j}\mu^\sigma\Gamma^2(r-r_+))s_+^{a_j}\mu^{(1+\sigma)d_j}\\
&\qquad \alpha^{5-\mathrm{sgn}(|j|)-\mathrm{sgn}(|j|-2)-a_j-3\sigma-3(1+\sigma)d_j}\\
&\le \delta\gamma_+^{b_j}s_+^{a_j}\mu_+^{d_j} \qquad (3.54)
\end{aligned}
$$

for all $|l|+|i|+|j|\le l_0$.

Above all, with the hypotheses H1)-H11), we complete one cycle of KAM steps.

4. Proof of Theorem 2

4.1. **Iteration Lemma.** Let $r_0,\gamma_0,s_0,\beta_0,\mu_0,\mathcal{O}_0,H_0,N_0,e_0,M_0,P_0$ be given at the beginning of Section 3 and let $\hat{D}_0=D(r_0,\beta_0),K_0=0$. For any $\nu=0,1\cdots$, 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:

$$r_\nu,\ \gamma_\nu,\ s_\nu,\ \beta_\nu,\ \mu_\nu,\ \alpha_\nu\ K_\nu,\ \mathcal{O}_\nu,\ \bar{L}^\nu_{1k},\ \bar{L}^\nu_{2k},$$
$$D_\nu,\ \hat{D}_\nu,\ e_\nu,\ M_\nu,\ M^\nu_{22},\ \bar{M}_\nu,\ N_\nu,\ H_\nu,\ P_\nu,\ \Phi_\nu$$

for $\nu=1,2,\cdots$, satisfying

$$\begin{aligned}
r_\nu &= r_0\Big(1-\sum_{i=1}^{\nu}\frac{1}{2^{i+1}}\Big),\\
\gamma_\nu &= \gamma_0\Big(1-\sum_{i=1}^{\nu}\frac{1}{2^{i+1}}\Big),\\
s_\nu &= \frac{1}{8}\alpha_{\nu-1}s_{\nu-1},\\
\beta_\nu &= \beta_0\Big(1-\sum_{i=1}^{\nu}\frac{1}{2^{i+1}}\Big),\\
\mu_\nu &= (64c_0)^{\frac{1}{1-\lambda_0}}\mu_{\nu-1}^{1+\sigma},\\
\alpha_\nu &= \mu_\nu^{\frac{1}{3}},\\
K_\nu &= \Big(\Big[\log\frac{1}{\mu_{\nu-1}}\Big]+1\Big)^{3\eta},\\
\mathcal{O}_\nu &= \Big\{\omega\in\mathcal{O}_{\nu-1}:|\langle k,\omega\rangle|>\frac{\gamma_{\nu-1}}{|k|^\tau},\ |\det\bar{L}^{\nu-1}_{1k}|>\frac{\gamma_{\nu-1}^{2m}}{|k|^{2m\tau}},\\
&\qquad |\det\bar{L}^{\nu-1}_{2k}|>\frac{\gamma_{\nu-1}^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}},\quad \text{for all } 0<|k|\le K_\nu\Big\},\\
\bar{L}^\nu_{1k} &= \sqrt{-1}\langle k,\omega\rangle I_{2m}-\delta M^\nu_{22}J,\\
\bar{L}^\nu_{2k} &= \sqrt{-1}\langle k,\omega\rangle I_{(n+2m)^2}+\delta\bar{M}_\nu,\\
D_\nu &= D(r_\nu,s_\nu),\\
\hat{D}_\nu &= D\Big(r_\nu+\frac{7}{8}(r_{\nu-1}-r_\nu),\beta_\nu\Big),\\
M_\nu &= \begin{pmatrix} M^\nu_{11} & M^\nu_{12}\\ (M^\nu_{12})^\top & M^\nu_{22}\end{pmatrix},\\
H_\nu &= N_\nu+P_\nu,\\
N_\nu &= e_\nu+\langle\omega,y\rangle+\frac{\delta}{2}\Big\langle\begin{pmatrix}y\\z\end{pmatrix},M^\nu\begin{pmatrix}y\\z\end{pmatrix}\Big\rangle.
\end{aligned}$$

The following lemma verifies the validity of all KAM steps.

Lemma 4.1. **(Iteration Lemma)** *If* (3.5) *holds for a sufficiently small* $\mu_0 = \mu_0(r_0, \beta_0, l_0)$*, then the KAM step described in* Section 3 *is valid for all* $\nu = 0, 1, \cdots$*, and the following holds for all* $\nu = 1, 2, \cdots$.

1) $e_\nu = e_\nu(\omega), M_\nu = M_\nu(\omega)$ *are smooth on* $\mathcal{O}_\nu$*,* M_ν *is real symmetric and nonsingular over* $\mathcal{O}_\nu$*,* P_ν *is real analytic in* $(x, y, z) \in D_\nu$*, smooth in* $(x, y, z) \in \hat{D}_\nu$ *and smooth in* $\omega \in \mathcal{O}_\nu$*, and moreover, for all* $|l| \le l_0$*,,*

$$|\partial_\omega^l e_\nu - \partial_\omega^l e_{\nu-1}|_{\mathcal{O}_\nu} \le c_0 \gamma_0^b \frac{\mu_0}{2^{\nu-1}}, \ |l| \le l_0, \tag{4.1}$$

$$|\partial_\omega^l e_\nu - \partial_\omega^l e_0|_{\mathcal{O}_\nu} \le 2c_0 \gamma_0^b \mu_0, \ |l| \le l_0, \tag{4.2}$$

$$|\partial_\omega^l M_\nu - \partial_\omega^l M_{\nu-1}|_{\mathcal{O}_\nu} \le c_0 \gamma_0^b \frac{\mu_0}{2^{\nu-1}}, \ |l| \le l_0, \tag{4.3}$$

$$|\partial_\omega^l M_\nu - \partial_\omega^l M_0|_{\mathcal{O}_\nu} \le 2c_0 \gamma_0^b \mu_0, \ |l| \le l_0, \tag{4.4}$$

$$|\partial_\omega^l \partial_x^i \partial_{(y,z)}^j P_\nu|_{D_\nu \times \mathcal{O}_\nu} \le \delta \gamma_\nu^{b_j} s_\nu^{a_j} \mu_\nu^{d_j}, \ |l| + |i| + |j| \le l_0. \tag{4.5}$$

2) $\Phi_\nu : \hat{D}_\nu \times \mathcal{O}_0 \longrightarrow \hat{D}_{\nu-1}$*,* $D_\nu \times \mathcal{O}_\nu \longrightarrow D_{\nu-1}$*, is symplectic for each* $\omega \in \mathcal{O}_0$*, and is of class* $C^{l_0+1+\sigma_0, l_0-1+\sigma_0}$*,* C^{α, l_0}*, respectively, where* α *stands for real analyticity and* $0 < \sigma_0 < 1$ *is fixed. Moreover,*

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

on $\hat{D}_\nu \times \mathcal{O}_\nu$*, and,*

$$\|\Phi_\nu - id\|_{C^{l_0+1+\sigma_0, l_0-1+\sigma_0}(\hat{D}_\nu \times \mathcal{O}_0)} \le c_0 \gamma^b \frac{\mu_0}{2^\nu}. \tag{4.6}$$

3)

$$\begin{aligned} \mathcal{O}_\nu \quad = \quad & \{\omega \in \mathcal{O}_{\nu-1} : |\langle k, \omega \rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau}, \ |\det \bar{L}_{1k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{2m}}{|k|^{2m\tau}}, \\ & |\det \bar{L}_{2k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}, \quad \text{for all } K_{\nu-1} < |k| \le K_\nu\}. \end{aligned}$$

Proof. We need to verify the conditions H1)-H11) in Section 3 for all $\nu = 0, 1, \cdots$. For simplicity, let $r_0 = 1$. By (3.3), (3.4), it is clear that the conditions H3)-H11) in Section 3 hold for $\nu = 0$ as long as μ_0 is sufficiently small. To verify H1)-H11) for the other cases, we need to estimate μ_ν. For convenience, denote $\theta = \frac{\sigma}{2} = \frac{1}{24}$ and $c_* = (64c_0)^{\frac{1}{\sigma(1-\lambda_0)}}$. By the definition of μ_ν, it is easy to see that

$$\mu_\nu = c_*^{-1} (c_* \mu_0)^{(1+\sigma)^\nu}. \tag{4.7}$$

Since, as μ_0 small,

$$\frac{2^{\frac{1}{1-\theta}}}{\mu_0^{\frac{\theta}{1-\theta}}} < \frac{1}{(c_* \mu_0)^\sigma},$$

there exists a sufficiently large λ such that

$$\frac{2^{\frac{1}{1-\theta}}}{\mu_0^{\frac{\theta}{1-\theta}}} \le \lambda \le \frac{1}{(c_* \mu_0)^\sigma}.$$

It follows that if μ_0 is sufficiently small (hence λ is sufficiently large), then

$$\lambda^{1-\theta} \mu_0^\theta \ge 2, \quad \mu_0 \le \frac{1}{c_* \lambda^{\frac{1}{\sigma}}} \le \frac{\beta_0}{2^6 c_0^2}, \tag{4.8}$$

and consequently,

$$\begin{aligned}
\mu_1 &= c_*^\sigma \mu_0^{1+\sigma} \le \frac{1}{\lambda}\mu_0 \le 2^{-\frac{1}{1-\theta}}\mu_0^{\frac{1}{1-\theta}} \le \mu_0,\\
\mu_2 &= c_*^\sigma \mu_1^{1+\sigma} \le \frac{1}{\lambda}\mu_1 \le \frac{1}{\lambda^2}\mu_0 \le 2^{-\frac{2}{1-\theta}}\mu_0^{\frac{1}{1-\theta}} \le \mu_0,\\
&\cdots\cdots\\
\mu_\nu &= c_*^\sigma \mu_{\nu-1}^{1+\sigma} \le \cdots \le \frac{1}{\lambda^\nu}\mu_0 \le 2^{-\frac{\nu}{1-\theta}}\mu_0^{\frac{1}{1-\theta}} \le \mu_0.
\end{aligned} \tag{4.9}$$

The choice of μ_0 in (4.8) immediately implies H8) for all $\nu \ge 1$. It also follows from (4.7), (4.8) and (4.9) that if μ_0 is sufficiently small (λ is sufficiently large), then

$$\mu_\nu = \mu_\nu^\theta \mu_\nu^{1-\theta} \le \mu_\nu^{1-\theta} \le \frac{\mu_0}{2^\nu}, \tag{4.10}$$

$$\begin{aligned}
&2c_0C_0\mu_{\nu-1}^{\frac{1}{3}}([\log\frac{1}{\mu_\nu}]+1)^{3\eta((n+2m)^2\tau+(n+2m)^2+1)}\\
&\qquad < \frac{1}{4^{\nu b}} < \frac{1}{2^{(\nu+2)(n+2m)^2}} < \frac{1}{2^{(\nu+2)(2m)}} < \frac{1}{2^{\nu+2}},
\end{aligned} \tag{4.11}$$

$$\frac{1}{2^{\nu+5}}([\log\frac{1}{\mu_\nu}]+1)^\eta > n+l_0, \tag{4.12}$$

$$\mu_\nu^{\frac{\theta}{2}}(n+[\chi]+1)!2^{(\nu+5)(n+[\chi]+1)} < 1, \tag{4.13}$$

for all $\nu \ge 1$. Using (3.4) and (4.11), we have

$$\begin{aligned}
&2C_0 s_\nu K_{\nu+1}^{(n+2m)^2\tau+(n+2m)^2+1}\\
&\quad \le 2C_0 s_0 \mu_{\nu-1}^{\frac{1}{3}} K_{\nu+1}^{(n+2m)^2\tau+(n+2m)^2+1}\\
&\quad \le 2C_0 \gamma_0^b \mu_{\nu-1}^{\frac{1}{3}} K_{\nu+1}^{(n+2m)^2\tau+(n+2m)^2+1}\\
&\quad \le \frac{\gamma_0^b}{4^b} < \gamma_\nu^b,\\
&c_0\mu_\nu K_{\nu+1}^{(n+2m)^2\tau+(n+2m)^2+1} \le c_0\mu_{\nu-1}^{\frac{1}{3}} K_{\nu+1}^{(n+2m)^2\tau+(n+2m)^2+1}\\
&\quad < \min\{\frac{\gamma_\nu-\gamma_{\nu+1}}{\gamma_0}, \frac{\gamma_\nu^{2m}-\gamma_{\nu+1}^{2m}}{\gamma_0^{2m}}, \frac{\gamma_\nu^{(n+2m)^2}-\gamma_{\nu+1}^{(n+2m)^2}}{\gamma_0^{(n+2m)^2}}\},
\end{aligned}$$

which verify H4), H5) respectively for all $\nu \ge 1$.

Since

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

(4.12) clearly implies H1) for all $\nu \ge 0$. To prove H2), we choose μ_0 further small if necessary so that

$$([\log(\frac{1}{\mu_\nu})]+1)^{2\eta} \ge \frac{1+\theta}{n+l_0}\log(\frac{1}{\mu_\nu}) + 3\eta\log([\log(\frac{1}{\mu_\nu})]+1). \tag{4.14}$$

Then by (4.12) and (4.14),

$$\begin{aligned}
\frac{1}{2^{\nu+5}}([\log(\frac{1}{\mu_\nu})]+1)^{3\eta} &\ge (n+l_0)([\log(\frac{1}{\mu_\nu})]+1)^{2\eta}\\
&\ge (1+\theta)\log(\frac{1}{\mu_\nu}) + 3\eta(n+l_0)\log([\log(\frac{1}{\mu_\nu})]+1),
\end{aligned}$$

i.e.,

$$K_{\nu+1}^{n+l_0} e^{-\frac{K_{\nu+1}}{2^{\nu+5}}} \leq {\mu_\nu}^{1+\theta}.$$

This together with (4.13) implies that

$$\int_{K_{\nu+1}}^{\infty} \lambda^{n+l_0} e^{-\frac{\lambda}{2^{\nu+5}}} \mathrm{d}\lambda \leq (n+l_0+1)! 2^{(\nu+5)(n+l_0)} K_{\nu+1}^{n+l_0} e^{-\frac{K_{\nu+1}}{2^{\nu+5}}} \leq \mu_\nu^{-\theta} \mu_\nu^{1+\theta} = \mu_\nu,$$

i.e., H2) holds for all $\nu \geq 0$.

Next, we note that

$$\begin{aligned} \Gamma(r_\nu - r_{\nu+1}) = \Gamma(\frac{1}{2^{\nu+2}}) &\leq \int_1^\infty \lambda^{n+[\chi]} e^{-\frac{\lambda}{2^{\nu+5}}} \mathrm{d}\lambda \\ &\leq (n+[\chi]+1)! 2^{(\nu+5)(n+[\chi]+1)}. \end{aligned}$$

Then as μ_0 small, (4.9) and (4.13) imply that

$$\mu_\nu^{2\theta} \Gamma^3(r_\nu - r_{\nu+1}) \leq \mu_\nu^{\frac{\theta}{2}} \leq (\frac{1}{2})^b \leq (\frac{\gamma_{\nu+1}}{\gamma_\nu})^b,$$

which proves H11) for all $\nu \geq 1$. By (4.13), we also have

$$\mu_\nu^\theta \Gamma(r_\nu - r_{\nu+1}) < 1. \tag{4.15}$$

It then follows from (4.8), (4.10) that

$$\mu_\nu \Gamma(r_\nu - r_{\nu+1}) < \mu_\nu^{1-\theta} \leq \frac{\mu_0}{2^\nu} \leq \frac{1}{2^{\nu+5} c_0} = \frac{1}{8c_0}(r_\nu - r_{\nu+1}), \tag{4.16}$$

which verifies H6), H9) for all $\nu \geq 1$. Since, by (4.8),

$$\mu_\nu^{1-\frac{3}{2}\theta} \leq \mu_\nu^{\frac{3}{4}} \leq \mu_0^{\frac{3}{4}} \leq \frac{1}{(8c_0)^{\frac{3}{2}}},$$

(4.15) implies that

$$c_0 \mu_\nu \Gamma(r_\nu - r_{\nu+1}) < c_0 \mu_\nu^{1-\theta} \leq \frac{1}{8} \mu_\nu^{\frac{1}{3}},$$

i.e., H7) holds for all $\nu \geq 1$. For the verification of H10), we have by (4.8), (4.10) and (4.15) that

$$c_0 \mu_\nu \Gamma(r_\nu - r_{\nu+1}) + c_0 \mu_\nu < 2c_0 \mu_\nu^{1-\theta} \leq c_0 \frac{\mu_0}{2^{\nu-1}} \leq \frac{\beta_0}{2^{\nu+2}} = \beta_\nu - \beta_{\nu+1},\ \nu \geq 1.$$

We are now ready to prove parts 1) and 2) of the lemma. Since all conditions H1)-H11) hold for $\nu = 0$, the KAM step described in Section 3 is valid for $\nu = 0$. As an induction hypothesis, we assume that for some $\nu_* \geq 1$ the KAM steps are valid for all $\nu = 1, 2, \cdots, \nu_*$. Applying Lemma 3.3 and (4.10) for all $\nu = 0, 1, \cdots, \nu_*$, we have

$$\begin{aligned} |\partial_\omega^l M_{\nu_*+1} - \partial_\omega^l M_0|_{\mathcal{O}_{\nu_*+1}} &\leq \sum_{\nu=0}^{\nu_*} |\partial_\omega^l M_{\nu+1} - \partial_\omega^l M_\nu|_{\mathcal{O}_{\nu+1}} \leq \sum_{\nu=0}^{\nu_*} c_0 \gamma^b \mu_\nu \\ &\leq \sum_{\nu=0}^{\nu_*} c_0 \gamma^b \frac{\mu_0}{2^\nu} \leq 2c_0 \gamma^b \mu_0 \leq 2c_0 \mu_0,\ |l| \leq l_0. \end{aligned} \tag{4.17}$$

It follows from (4.8) and (4.17) that as μ_0 small M_{ν_*+1} is non-singular on $\mathcal{O}_{\nu_*+1}$ and H3) holds for $\nu = \nu_* + 1$. Thus, the KAM step described in Section 3 is also valid for $\nu = \nu_* + 1$.

Above all, by performing the KAM steps described in Section 3 for all $\nu = 0, 1, \cdots$, we obtain the desired sequences stated at the beginning of the section which satisfy all properties described in parts 1) and 2) of the lemma. Ideed, (4.1), (4.3) follow from Lemma 3.3 and (4.10). (4.2), (4.4) follow from (4.1), (4.3) respectively along with an argument similar to (4.17). Moreover, (4.5) follows from (3.54), and, (4.6) follows from Lemma 3.8 and (4.16).

Note that 3) automatically holds for $\nu = 1$. We now let $\nu > 1$. By Lemma 3.4, it is clear that

$$\begin{aligned}\mathcal{O}_{\nu-1} &= \{\omega \in \mathcal{O}_{\nu-1} : |\langle k, \omega\rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau},\ |\det \bar{L}_{1k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{2m}}{|k|^{2m\tau}},\\ &\qquad |\det \bar{L}_{2k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}, \quad \text{for all } 0 < |k| \le K_{\nu-1}\}.\end{aligned}$$

Denote

$$\begin{aligned}\hat{\mathcal{O}}_{\nu} &= \{\omega \in \mathcal{O}_{\nu-1} : |\langle k, \omega\rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau},\ |\det \bar{L}_{1k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{2m}}{|k|^{2m\tau}},\\ &\qquad |\det \bar{L}_{2k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}, \quad \text{for all } K_{\nu-1} < |k| \le K_{\nu}\}.\end{aligned}$$

Then

$$\begin{aligned}\mathcal{O}_{\nu} &= \{\omega \in \mathcal{O}_{\nu-1} : |\langle k, \omega\rangle| > \frac{\gamma_{\nu-1}}{|k|^\tau},\ |\det \bar{L}_{1k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{2m}}{|k|^{2m\tau}},\\ &\qquad |\det \bar{L}_{2k}^{\nu-1}| > \frac{\gamma_{\nu-1}^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}, \quad \text{for all } 0 < |k| \le K_{\nu}\}\\ &= \mathcal{O}_{\nu-1} \cap \hat{\mathcal{O}}_\nu = \hat{\mathcal{O}}_\nu. \qquad (4.18)\end{aligned}$$

The lemma is now complete. □

4.2. **Convergence.** Fix an arbitrary $0 < \epsilon < 1$ in (3.2). Then $\mu_0 = o(\mu^{1-\epsilon})$, and, μ_0 is small iff μ is. Therefore, there is a sufficiently small $\mu = \mu(r, s, l_0)$ such that $\mu_0 = \mu_0(r, s, l_0)$ fits in the smallness requirement in Lemma 4.1. We then obtain the following sequences

$$\begin{aligned}&\Psi_\nu = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Phi_\nu : \hat{D}_\nu \times \mathcal{O}_\nu \to \hat{D}_0,\\ &H \circ \Psi_\nu = H_\nu = N_\nu + P_\nu,\\ &N_\nu = e_\nu + \langle \omega, y\rangle + \frac{\delta}{2}\langle \begin{pmatrix} y\\ z\end{pmatrix}, M_\nu(\omega)\begin{pmatrix} y\\ z\end{pmatrix}\rangle,\end{aligned}$$

$\nu = 0, 1, \cdots$, which satisfy all properties described in Lemma 4.1, where $\Psi_0 = \Phi_0 = id$.

Let

$$\mathcal{O}_* = \bigcap_{\nu=0}^{\infty} \mathcal{O}_\nu.$$

Then $\mathcal{O}_* \subset \hat{\mathcal{O}}_\gamma$. By (4.1), (4.3), it is clear that e_ν, M_ν converge uniformly on $\mathcal{O}_*$ as $\nu \to \infty$. We denote their limits by $e_\infty(\omega)$, $M_\infty(\omega)$, respectively. Fix a $0 < \sigma_0 < 1$. For all ν and $(x, y, z) \in D_\nu$, by a similar application of the Whitney extension theorem as in Lemma 3.6, we can extend P_ν uniformly to functions of class $C^{l_0-1+\sigma_0}$ with respect to $\omega \in \mathcal{O}_0$, whose Hölder norms satisfy the same estimates (4.5), up to multiplication of a constant. With such extended functions P_ν, we can use the same formula (3.37) inductively to define $C^{l_0-1+\sigma_0}$ extensions of M_ν on

$\mathcal{O}_0$. Then the Hölder norms $\|M_\nu - M_{\nu-1}\|_{C^{l_0-1+\sigma_0}(\mathcal{O}_0)}$ and $\|M_\nu - M_0\|_{C^{l_0-1+\sigma_0}(\mathcal{O}_0)}$ still satisfy the same estimates (4.3) and (4.4), respectively, up to multiplication of a constant. This implies that for all $|l| \le l_0 - 1$, $\partial_\omega^l M_\nu$ are (Hölder) equally continuous and uniformly convergent on on $\mathcal{O}_0$. In particular, the limit of M_ν on $\mathcal{O}_0$ is of class C^{l_0-1} which coincides with the original limit on $\mathcal{O}_*$. Thus, $M_\infty(\omega)$ is C^{l_0-1} Whitney smooth on $\mathcal{O}_*$, and,

$$|\partial_\omega^l M_\infty - \partial_\omega^l M_0|_{\mathcal{O}_*} = O(\gamma^b \mu_0) = o(\gamma^b \mu^{1-\epsilon}),\ |l| \le l_0 - 1,$$

in the sense of Whitney (see [11]). Similarly, $e_\infty(\omega)$ is C^{l_0-1} Whitney smooth on $\mathcal{O}_*$, and,

$$|\partial_\omega^l e_\infty - \partial_\omega^l e_0|_{\mathcal{O}_*} = o(\gamma^b \mu^{1-\epsilon}),\ |l| \le l_0 - 1,$$

in the sense of Whitney.

Set

$$\mathcal{D}_* = D(\frac{1}{2}r_0, \frac{1}{4}\beta_0) \times \mathcal{O}_*, \quad \mathcal{D}_0 = D(\frac{1}{2}r_0, \frac{1}{4}\beta_0) \times \mathcal{O}_0$$

and let Ψ_ν be defined as above using the extended Φ_ν on $\hat{D}_\nu \times \mathcal{O}_0$ as in Lemma 4.1. Since

$$\Psi_\nu = id + \sum_{i=1}^{\nu} (\Psi_i - \Psi_{i-1}), \tag{4.19}$$

and,

$$\begin{aligned} \Psi_1 - \Psi_0 &= \Phi_1 - id, \\ \Psi_i - \Psi_{i-1} &= \Phi_1 \circ \cdots \circ \Phi_i - \Phi_1 \circ \cdots \circ \Phi_{i-1} \\ &= \int_0^1 D(\Phi_1 \circ \cdots \circ \Phi_{i-1})(id + \theta(\Phi_i - id)) \mathrm{d}\theta (\Phi_i - id),\ i \ge 2, \end{aligned} \tag{4.20}$$

we have by (4.6) that

$$|\Phi_i - id|_{\mathcal{D}_0} \le \frac{\mu_0}{2^i},$$

and,

$$\begin{aligned} &|D(\Phi_1 \circ \cdots \circ \Phi_{i-1})(id + \theta(\Phi_i - id))|_{\mathcal{D}_0} \\ &\quad \le |D\Phi_1(\Phi_2 \circ \cdots \circ \Phi_{i-1})(id + \theta(\Phi_i - id))|_{\mathcal{D}_0} \cdots \\ &\qquad |D\Phi_{i-1}(id + \theta(\Phi_i - id))|_{\mathcal{D}_0} \\ &\quad \le (1 + \frac{c_0\mu_0}{2}) \cdots (1 + \frac{c_0\mu_0}{2^{i-1}}) \\ &\quad \le e^{\frac{c_0\mu_0}{2} + \cdots + \frac{c_0\mu_0}{2^{i-1}}} \le e^{c_0\mu_0} \le e^{\beta_0}. \end{aligned}$$

Thus,

$$|\Psi_i - \Psi_{i-1}|_{\mathcal{D}_0} \le e^{\beta_0} |\Phi_i - id|_{\mathcal{D}_0} \le c_0 e^{\beta_0} \frac{\mu_0}{2^i}$$

for all $i = 1, 2, \cdots$. It follows that Ψ_ν converges uniformly on $\mathcal{D}_0$. We denote its limit by Ψ_∞. Then

$$\Psi_\infty = id + \sum_{i=1}^{\infty} (\Psi_i - \Psi_{i-1}), \tag{4.22}$$

and, $|\Psi_\infty - id|_{\mathcal{D}_0} = O(\mu_0) = o(\mu^{1-\epsilon})$. Hence, Ψ_∞ is uniformly close to the identity.

By differentiating (4.19)-(4.22) and applying (4.6) and the above arguments inductively, one can further show that Ψ_ν converges in $C^{l_0+\sigma_0, l_0-1+\sigma_0}(\mathcal{D}_0)$ norm to Ψ_∞. It follows that Ψ_∞ is C^{l_0} smooth in $\xi = (x, y, z) \in D(\frac{1}{2}r_0, \frac{1}{4}\beta_0)$ and C^{l_0-1}

Whitney smooth in $\omega \in \mathcal{O}_*$. Similarly, Ψ_∞ is also C^{l_0} uniformly close to the identity. We omit the details.

Now, by the uniform convergence of

$$N_\nu = e_\nu(\omega) + \langle \omega, y\rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M_\nu(\omega)\begin{pmatrix} y \\ z \end{pmatrix}\rangle, \ \Psi_\nu$$

to

$$N_\infty = e_\infty(\omega) + \langle \omega, y\rangle + \frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M_\infty(\omega)\begin{pmatrix} y \\ z \end{pmatrix}\rangle, \ \Psi_\infty$$

respectively on $\mathcal{D}_*$, we see that, on $\mathcal{D}_*$,

$$P_\nu = H \circ \Psi_\nu - N_\nu$$

and their derivatives, converge uniformly to

$$P_\infty = H \circ \Psi_\infty - N_\infty$$

and its corresponding derivatives. It follows that P_∞ is C^{l_0} smooth in $\xi = (x, y, z) \in D(\frac{1}{2}r_0, \frac{1}{4}\beta_0)$ and C^{l_0-1} Whitney smooth in $\omega \in \mathcal{O}_*$. Since Φ_ν is real analytic on D_ν for each $\omega \in \mathcal{O}_\nu$, P_∞ is also analytic in x when $y = 0, z = 0, \omega \in \mathcal{O}_*$.

Recall that, for any $\omega \in \mathcal{O}_*$,

$$|P_\nu|_{D_\nu} \le \delta\gamma_\nu^b s_\nu^2 \mu_\nu.$$

Using the Cauchy estimate, we have that

$$|\partial_y^j \partial_z^k P_\nu|_{D(r_\nu, \frac{1}{2}s_\nu)} \le \delta\gamma_\nu^b \mu_\nu \tag{4.23}$$

for all $j \in Z_+^n, k \in Z_+^{2m}$ with $|j| + |k| \le 2$. By (4.7), it is easy to see that, the right hand side of (4.23) converges to 0 as $\nu \to 0$. Therefore, on $D(\frac{r_0}{2}, 0) \times \mathcal{O}_*$,

$$\partial_y^j \partial_z^k P_\infty = 0, \ j \in Z_+^n, k \in Z_+^{2m}, \ |j| + |k| \le 2.$$

In particular, for each $\omega \in \mathcal{O}_*$, the equation of motion restricted to $D(\frac{r_0}{2}, 0)$ reduces to

$$\begin{aligned} \dot{x} &= \omega + \frac{\partial}{\partial y}\frac{\delta}{2}\langle \begin{pmatrix} y \\ z \end{pmatrix}, M_\infty \begin{pmatrix} y \\ z \end{pmatrix}\rangle + \frac{\partial P_\infty}{\partial y} = \omega, \\ \dot{y} &= -\frac{\partial P_\infty}{\partial x} = 0, \\ \dot{z} &= J(\frac{\partial N_\infty}{\partial z} + \frac{\partial P_\infty}{\partial z}) = 0. \end{aligned}$$

Thus, for each $\omega \in \mathcal{O}_*$, the perturbed system (1.2) possesses an analytic, quasi-periodic, (Floquet) invariant torus with the Diophantine frequency vector ω, which is slightly deformed from the unperturbed torus corresponding to ω. Moreover, the perturbed tori form a C^{l_0-1} Whitney smooth family.

4.3. **Measure estimate.**

Lemma 4.2. *Suppose that $g \in C^p(\bar{I})$, $p \ge 2$, where $I \subset R^1$ is a finite interval. Let $I_h = \{x \in I : |g(x)| \le h\}, h > 0$. If on I, $|g^{(p)}(x)| \ge c > 0$ for some constant c, then $|I_h| \le c' h^{\frac{1}{p}}$, where $c' = p + 2 + \dfrac{2}{c}$.*

Proof. See [16], Lemma 2.1. □

We now prove the measure estimate

$$|\mathcal{O}_0 \backslash \mathcal{O}_*| = O(\gamma).$$

By Lemma 4.1 3),

$$\mathcal{O}_{\nu+1} = \mathcal{O}_\nu \setminus \bigcup_{K_\nu < |k| \le K_{\nu+1}} R_k^{\nu+1}(\gamma_\nu),$$

where

$$\begin{aligned} R_k^{\nu+1}(\gamma_\nu) \quad = \quad & \{\omega \in \mathcal{O}_\nu : |\langle k, \omega\rangle| \le \frac{\gamma_\nu}{|k|^\tau}, \text{ or } |\det \bar{L}_{1k}^\nu| \le \frac{\gamma_\nu^{2m}}{|k|^{2m\tau}}, \\ & \text{or } |\det \bar{L}_{2k}^\nu| \le \frac{\gamma_\nu^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}\}. \end{aligned} \tag{4.24}$$

Thus,

$$\mathcal{O}_0 \backslash \mathcal{O}_* = \bigcup_{\nu=0}^{\infty} \bigcup_{K_\nu < |k| \le K_{\nu+1}} R_k^{\nu+1}(\gamma_\nu).$$

Thus, the desired measure estimate amounts to the estimate of $|R_k^{\nu+1}(\gamma_\nu)|$.

To begin with, we fix a $\nu = 0, 1, \cdots$ and a $k = (k_1, k_2, \cdots, k_n)^\top \in Z^n \backslash \{0\}$, and, without loss of generality, assume that $k_1 = \max_{1 \le i \le n}\{|k_i|\}$. Let $g(\omega) = \det(\bar{L}_{2k}^\nu(\omega))$. For any $\hat{\omega} \in \tilde{\mathcal{O}}_\nu = \{(\omega_2, \cdots, \omega_n)$: there is a ω_1 such that $\omega = (\omega_1, \omega_2, \cdots, \omega_n)^\top \in \mathcal{O}_\nu\}$, we consider the set $I(\hat{\omega}) = \{\omega_1 : \omega = (\omega_1, \hat{\omega})^\top \in \mathcal{O}_\nu\}$ and define

$$S_1(\hat{\omega}) = \{\omega_1 \in I(\hat{\omega}) : |g(\omega_1, \hat{\omega})| \le \frac{\gamma_\nu^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}\}.$$

Then, it follows from (4.4) that, on $I(\hat{\omega})$,

$$A(\hat{\omega}) = |\frac{\partial^{(n+2m)^2}}{\partial \omega_1^{(n+2m)^2}} g(\omega_1, \hat{\omega})| = |k_1|^{(n+2m)^2}(1 + O(\delta)),$$

where, $O(\delta)$ is independent of k, ν, ω, l_0 (note that, since only up to the $(n+2m)^2$-th order derivatives were taken, we can choose $l_0 = (n+2m)^2$ in (4.4) so that b is also independent of l_0). Hence if δ is small, say $0 < \delta \le \delta_0$ for some δ_0, then

$$A(\hat{\omega}) \ge \frac{1}{2}|k_1|^{2(n+2m)^2} > \frac{1}{2}, \tag{4.25}$$

for all $\omega_1 \in I(\hat{\omega})$. Applying Lemma 4.2 to $g(\omega_1, \hat{\omega})$ and making use of (4.25), we see that

$$|S_1(\hat{\omega})| \le ((n+2m)^2 + 6)\frac{\gamma_\nu}{|k|^\tau}$$

for all $\hat{\omega} \in \tilde{\mathcal{O}}_\nu$.

Thus, by Fubini's theorem, there is a constant $c_9 > 0$ such that

$$\begin{aligned} |\{\omega \in \mathcal{O}_\nu : \det(\bar{L}_{2k}^\nu) \quad & \le \quad \frac{\gamma_\nu^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}\}| = |\{\omega \in \mathcal{O}_\nu : |g(\omega)| \le \frac{\gamma_\nu^{(n+2m)^2}}{|k|^{(n+2m)^2\tau}}\}| \\ & \le \quad c_9 \max_{\hat{\omega} \in \tilde{\mathcal{O}}_\nu} |S_1(\hat{\omega})| \le 4c_9((n+2m)^2 + 6)\frac{\gamma}{|k|^\tau}. \end{aligned}$$

Similarly, there are constants $c_{10}, c_{11} > 0$ such that

$$|\{\omega \in \mathcal{O}_\nu : |\langle k, \omega\rangle| \le \frac{\gamma_\nu}{|k|^\tau}\}| \le c_{10}\frac{\gamma}{|k|^\tau},$$
$$|\{\omega \in \mathcal{O}_\nu : |\det \bar{L}^\nu_{1k}| \le \frac{\gamma_\nu^{2m}}{|k|^{2m\tau}}\}| \le c_{11}\frac{\gamma}{|k|^\tau}.$$

Therefore, there is a constant $c_{12} > 0$ such that

$$|R_k^{\nu+1}(\gamma_\nu)| \le c_{12}\frac{\gamma}{|k|^\tau}. \tag{4.26}$$

In fact, it is easily seen from the above that (4.26) holds for all $k \in Z^n \setminus \{0\}$ and $\nu = 0, 1, \cdots$. Consequently,

$$\begin{aligned} |\mathcal{O}_0 \backslash \mathcal{O}_*| &\le |\bigcup_{\nu=0}^{\infty} \bigcup_{K_\nu < |k| \le K_{\nu+1}} R_k^{\nu+1}(\gamma_\nu)| \le c_{12}\gamma \sum_{\nu=0}^{\infty} \sum_{K_\nu < |k| \le K_{\nu+1}} \frac{1}{|k|^\tau} \\ &= O(\gamma), \text{ as } \gamma \to 0. \end{aligned}$$

The proof of Theorem 2 is now complete.

5. Proof of Theorem 1

Let ρ be sufficiently small such that $O_\rho(g, G) \ne \emptyset$. By using finite open coverings on $O_{\frac{\rho}{2}}(g, G)$, it is easy to see that $O_\rho(g, G)$ consists of finitely many connected components. Therefore, to prove the theorem, we can assume without loss of generality that $O_\rho(g, G)$ itself is connected. It then follows from the implicit function theorem that there is a positive integer j_0 such that each $y \in O_\rho(g, G)$ admits exactly j_0 many non-degenerate critical points of $h_0(\cdot, y)$. In fact, there are analytic functions $\varphi_j : O_\rho(g, G) \to T^m$, $j = 1, 2, \cdots, j_0$, such that, for each $1 \le j \le j_0$ and $y \in O_\rho(g, G)$, $\varphi_j(y)$ is a non-degenerate critical point of $h_0(\cdot, y)$.

5.1. **Treshchev's Reduction.** Let φ_j be one of the analytic functions above. We now use the idea of Treshchev ([15]) to reduce (1.1) to a Hamiltonian of form (1.2) near the tori $\{T_y(\varphi_j(y)) : y \in O_\rho(g, G)\}$. For simplicity, we shall omit all constant terms appearing in Hamiltonians after each transformation. Below, all "O" will be independent of j.

The reduction process will consist of three steps. The first step is the separation of 'resonant' and 'non-resonant' parts in both angular and action variables. To do so, we first use the Taylor expansion for H_0 at $y_0 \in O_\rho(g, G)$ to rewrite (1.1) into the form

$$H_j(x, y) = \langle \omega(y_0), y - y_0\rangle + \frac{1}{2}\langle \Gamma_1(y_0)(y - y_0), y - y_0\rangle + \varepsilon P(x, y, \varepsilon) + O(|y - y_0|^3),$$

where

$$\Gamma_1(y_0) = K_0^\top \frac{\partial^2 H_0}{\partial y^2}(y_0) K_0$$

and $K_0 = (K_1, K_2)$ is as in Section 1. We then employ the linear symplectic transformation: $y - y_0 = K_0 p, q = K_0^\top(x - x_0)$, where x_0 is such that $K_2^\top x_0 = \varphi_j(y_0)$, to further transform the above Hamiltonian into the form

$$\begin{aligned} H_j(q, p) &= \langle \omega'(y_0), p'\rangle + \frac{1}{2}\langle p, \Gamma_1(y_0) p\rangle + \varepsilon \bar{P}_j(q, p, y_0) \\ &\quad + O(|p|^3) + O(\varepsilon^2), \end{aligned} \tag{5.1}$$

where $\omega'(y_0) = K_1^\top \omega(y_0)$, $p = (p', p'')^\top$, $q = (\psi, v)^\top$ with $p' \in R^n, p'' \in R^m, \psi \in T^n, v \in R^m$, and

$$\bar{P}_j(q, p, y_0) = P(x_0 + (K_0^\top)^{-1} q, y_0 + K_0 p, 0).$$

Consider

$$\mathcal{O}(g, G) = \{K_1^\top \omega(y) \in R^n : y \in O_\rho(g, G)\}.$$

Since $\mathcal{O}(g, G)$ is diffeomorphic to $O_\rho(g, G)$, for simplicity, we shall use $\omega \in \mathcal{O}(g, G)$ as a parameter instead of $y_0 \in O_\rho(g, G)$. Hence (5.2) can be rewritten as

$$(5.2)\quad H_j(\psi, v, p', p'') = \langle \omega, p' \rangle + \frac{1}{2}\langle p, \Gamma_1(\omega) p \rangle + \varepsilon \bar{P}_j(\psi, v, p, \omega) + O(|p|^3) + O(\varepsilon^2),$$

where $\omega \in \mathcal{O}(g, G)$ serves as a parameter, and, $\Gamma_1(\omega)$ and $\bar{P}_j(\psi, v, p, \omega)$ are defined through $\Gamma_1(y_0)$ and $\bar{P}_j(\psi, v, p, y_0)$ respectively, via the diffeomorphism between $\mathcal{O}(g, G)$ and $O_\rho(g, G)$. In the sequel, we also denote the matrix function

$$\Gamma_2^j(y_0) = \frac{\partial^2 h_0}{\partial \varphi^2}(\varphi_j(y_0), y_0)$$

by $\Gamma_2^j(\omega)$ accordingly.

Let $\tau > n - 1$ be fixed as in Theorem 2 and consider the set

$$\hat{\mathcal{O}}_\gamma = \{\omega \in \mathcal{O}(g, G) : |\langle k, \omega \rangle| > \frac{\gamma}{|k|^\tau} \quad \text{for all} \quad k \in Z^n \setminus \{0\}\}.$$

The next step of the reduction is to separate the first order resonant terms from the perturbations in (5.2). Consider the family of functions $\{S_\omega^j\}$, $\omega \in \hat{\mathcal{O}}_\gamma$, on $(T^n \times R^m) \times R^d$ defined by

$$S_\omega^j(q, Y) = \langle Y, q \rangle + \varepsilon \sum_{k \in Z^n \setminus \{0\}} S_k^j e^{\sqrt{-1}\langle k, \psi \rangle},$$

where $S_k^j = \frac{\sqrt{-1} h_k^j(v, \omega)}{\langle \omega, k \rangle}$ and $h_k^j(v, \omega) = \int_{T^m} \bar{P}_j(\psi, v, 0, \omega) e^{\sqrt{-1}\langle k, \psi \rangle} \mathrm{d}\psi$. Since each $\omega \in \hat{\mathcal{O}}_\gamma$ is Diophantine, S_ω^j is analytic and induces a family of symplectic transformations $R^d \times (T^n \times R^m) \longrightarrow R^d \times (T^n \times R^m)$: $(p, q) \mapsto (Y, q)$ with

$$p = \frac{\partial S_\omega^j(q, Y)}{\partial q}, \quad q = \frac{\partial S_\omega^j(q, Y)}{\partial Y}.$$

In fact, if we denote $Y = (Y', Y'') \in R^n \times R^m$, then

$$p' = Y' + \sqrt{-1}\varepsilon \sum_{k \in Z^n} k S_k^j e^{\sqrt{-1}\langle k, \psi \rangle} = Y' + O(\frac{\varepsilon}{\gamma}),$$

and,

$$p'' = Y'' + \sqrt{-1}\varepsilon \sum_{k \in Z^n \setminus \{0\}} \frac{1}{\langle k, \omega \rangle} \frac{\partial h_k^j}{\partial v} e^{\sqrt{-1}\langle k, \psi \rangle} = Y'' + O(\frac{\varepsilon}{\gamma}).$$

By using the Fourier expansion, we have

$$\begin{aligned}
\bar{P}_j(\psi, v, p, \omega) &= \sum_{k \in Z^n} h_k^j(v, \omega) e^{\sqrt{-1}\langle k, \psi \rangle} + O(|p|^2) \\
&= h_0^j(0, \omega) + \frac{1}{2}\langle v, \Gamma_2^j(\omega) v \rangle + \sum_{k \in Z^n \setminus \{0\}} h_k^j(v, \omega) e^{\sqrt{-1}\langle k, \psi \rangle} \\
&\quad + O(|p|^2) + O(|v|^3).
\end{aligned}$$

Thus, the Hamiltonian (5.2) becomes

$$\begin{aligned} H_j(q,Y) &= \langle\omega,Y'\rangle+\frac{1}{2}\langle Y,\Gamma_1(\omega)Y\rangle+\frac{\varepsilon}{2}\langle v,\Gamma_2^j(\omega)v\rangle \\ &\quad +O(\frac{\varepsilon}{\gamma}Y)+O(\frac{\varepsilon^2}{\gamma^2})+O(|Y|^3)+\varepsilon O(|v|^3). \end{aligned}$$

We now consider the rescaling $Y'=\varepsilon^{\frac12}\bar{Y'}$, $Y''=\varepsilon^{\frac12}\bar{Y''}$ and denote $\bar{Y}=(\bar{Y'},\bar{Y''})^\top$. Then the rescaled Hamiltonian reads

$$\begin{aligned} H_j(\psi,\bar{Y'},\bar{Y''},v) &= \frac{H_j(q,\varepsilon^{\frac12}\bar{Y})}{\varepsilon^{\frac12}} \\ &= \langle\omega,\bar{Y'}\rangle+\frac{\varepsilon^{\frac12}}{2}\Big(\langle\bar{Y},\Gamma_1(\omega)\bar{Y}\rangle+\langle v,\Gamma_2^j(\omega)v\rangle\Big) \\ &\quad +\varepsilon^{\frac12}(\frac{\varepsilon^{\frac12}}{\gamma}O(\bar{Y})+O(\frac{\varepsilon}{\gamma^2})+\varepsilon^{\frac12}O(|\bar{Y}|^3)+\varepsilon^{\frac12}O(|v|^3)). \end{aligned} \tag{5.3}$$

Let $M_j(\omega)=\operatorname{diag}(\Gamma_1(\omega),\Gamma_2^j(\omega))$. We replace $\psi,\bar{Y'},\bar{Y''},\varepsilon^{\frac12}$ above by x,y,u,ε respectively and denote $z=(u,v)^\top$. Then (5.3) becomes the desired Hamiltonian:

$$H_j(x,y,z)=N_j+P_j, \tag{5.4}$$

where

$$\begin{aligned} N_j &= \langle\omega,y\rangle+\frac{\varepsilon}{2}\langle\begin{pmatrix}y\\z\end{pmatrix},M_j(\omega)\begin{pmatrix}y\\z\end{pmatrix}\rangle, \\ P_j &= O(\frac{\varepsilon^2}{\gamma}|y|)+O(\frac{\varepsilon^3}{\gamma^2})+\varepsilon^2O(|y|^3)+\varepsilon^2O(|v|^3). \end{aligned}$$

Summarizing up, we have found a family of real analytic, symplectic transformations $\{\mathcal{F}_\omega^j\}$, $j=1,2,\cdots,j_0$, which depend analytically on ε and transform the original Hamiltonian into (5.4) for ω lying in the Cantor set $\hat{\mathcal{O}}_\gamma$. By the Whitney extension theory developed in [11] (see also [3]), the family, or equivalently, $\{S_\omega^j\}$, $\omega\in\hat{\mathcal{O}}_\gamma$, admits a C^∞ extension with respect to $\omega\in\mathcal{O}(g,G)$. This allows us to work with the Hamiltonian (5.4) by assuming that M_j, P_j are well defined and C^∞ in $\omega\in\mathcal{O}(g,G)$.

Tracing back the construction above, we also see that, on a fixed domain $D(r_*,s_*)\times\mathcal{O}(g,G)$,

$$|\partial_\omega^l P_j|=O(\frac{\varepsilon^2}{\gamma^{|l|+1}})+O(\frac{\varepsilon^3}{\gamma^{|l|+2}}),\ |l|\le l_*, \tag{5.5}$$

where $l_*=\max\{(n+2m)^2+1,l_0\}$, l_0 is as in Theorem 1.

5.2. **Application of Theorem 2.** Let $r=r_*$, $\mathcal{O}=\mathcal{O}(g,G)$ and define

$$\delta=\varepsilon,\ s=\gamma^{b_*}=\varepsilon^{b_*a},\ \mu=\varepsilon^{1-a_*(l_*+1+5b_*)},$$

where $b_*=(2l_*+3)(n+2m)^2$, and, $0<a<\frac{1}{l_*+1+5b_*}$ and $a_*\in(a,\frac{1}{l_*+1+5b_*})$ are fixed. By (5.5), we have

$$|\partial_\omega^l P_j|_{D(r,s)\times\mathcal{O}}\le\delta\gamma^{3b_*}s^2\mu,\ |l|\le l_*,$$

provided that ε is sufficiently small, say, $0<\varepsilon\le\varepsilon_0$ for some $\varepsilon_0=\varepsilon_0(g,G,\rho,l_0)>0$. Therefore, (1.3) holds with $l_0:=l_*$, $b:=b_*$ uniformly for all $P=P_j$, $j=1,2,\cdots,j_0$, $0\le\varepsilon\le\varepsilon_0$.

At the beginning of Section 3, choose

$$M^* = \max_{l \le l_0, 1 \le j \le j_0} |\partial_\omega^l M_j(\omega)|_{\mathcal{O}}, \quad M_* = \max_{1 \le j \le j_0} |M_j^{-1}(\omega)|_{\mathcal{O}}.$$

Then it is easy to see that all arguments in Sections 3, 4 hold uniformly for $M = M_j$, $P = P_j$, $j = 1, 2, \cdots, j_0$. That is, Theorem 2 is applicable to (5.4) uniformly for all $j = 1, 2, \cdots, j_0$, as long as ε_0 (hence δ and μ) is (are) sufficiently small. In particular, if $\mathcal{O}_\gamma^j \subset \hat{\mathcal{O}}_\gamma$ denotes the Cantor set in Theorem 2 for $j = 1, 2, \cdots, j_0$ respectively, then $|\mathcal{O} \backslash \mathcal{O}_\gamma^j| = O(\gamma)$, uniformly for $j = 1, 2, \cdots, j_0$. Let $\mathcal{O}_\gamma = \cap_{j=1}^{j_0} \mathcal{O}_\gamma^j$. Then $|\mathcal{O} \backslash \mathcal{O}_\gamma| = O(\gamma)$, and, $\mathcal{O}_\gamma$ corresponds to j_0 many C^{l_0} Whitney smooth families of real analytic, invariant, quasi-periodic n-tori $\{T_j(\omega) : \omega \in \mathcal{O}_\gamma\}$, $j = 1, 2, \cdots, j_0$, of (5.4). Finally, let $\Lambda_\varepsilon \subset O_\rho(g, G)$ be associated to $\mathcal{O}_\gamma \subset \mathcal{O} = \mathcal{O}(g, G)$ through the defeomorphism between $O_\rho(g, G)$ and $\mathcal{O}(g, G)$. Then $|O_\rho(g, G) \setminus \Lambda_\varepsilon| = O(\varepsilon^a) \to 0$, as $\varepsilon \to 0$. The proof of Theorem 1 is now complete.

References

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

[2] S. V. Bolotin, Homoclinic orbits to invariant tori of Hamiltonian systems, *Dynamical Systems in Classical Mechanics,* Amer. Math. Soc. Transl. Vol.**168,** No 2 (1995), 21-90

[3] H. Broer, G. Huitema and M. Sevryuk, Quasi-Periodic Motions in Families of Dynamical Systems, *Lect. Notes Math.,* **1645**, Springer-Verlag, 1996

[4] C.-Q. Cheng, Birkhoff-Kolmogorov-Arnold-Moser tori in convex Hamiltonian systems, *Comm. Math. Phys.* **177** (1996), 529-559

[5] L. Chierchia and G. Gallavotti, Drift and diffusion in phase space, *Ann. Inst. H. Poincaré Phy. Th.* **69** (1994), 1-144

[6] F. Z. Cong, T. Küpper, Y. Li and J. G. You, KAM-type theorem on resonant surfaces for nearly integrable Hamiltonian systems, *J. Nonl. Sci.* to appear

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

[8] 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

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

[10] H. Poincaré, Les Méthodes Nouvelles de la Mécaniques Céleste, I-III, Gauthier-Villars, 1892, 1893, 1899 (The English translation: New Methods of Celestial Mechanics, AIP Press, Williston, 1992)

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

[12] M. Rudnev and S. Wiggins, KAM theory near multiplicity one resonant surfaces in perturbations of A-priori stable Hamiltonian systems, *J. Nonl. Sci.* **7** (1997), 177-209

[13] C. L. Siegel and J. K. Moser, Lectures on Celestial Mechanics, Grundlenhren **187,** Springer, Berlin, 1971

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

[15] D. V. Treshchev, The mechanism of destruction of resonance tori of Hamiltonian systems, *Math. USSR Sb.* **68** (1991), 181-203

[16] 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

Yong Li: Department of Mathematics, Jilin University, Changchun 130023, PRC
E-mail address: `ylimd@21cn.com or ylimd@mail.jl.cn`

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