

MINIMAL INVARIANT TORI IN THE RESONANT REGIONS FOR NEARLY INTEGRABLE HAMILTONIAN SYSTEMS

CHONG-QING CHENG

ABSTRACT. Consider a real analytical Hamiltonian system of KAM type $H(p, q) = N(p) + P(p, q)$ that has n degrees of freedom ($n > 2$) and is positive definite in p . Let $\Omega = \{\omega \in \mathbb{R}^n \mid \langle \bar{k}, \omega \rangle = 0, \bar{k} \in \mathbb{Z}^n\}$. In this paper we show that for most rotation vectors in Ω , in the sense of $(n - 1)$ -dimensional Lebesgue measure, there is at least one $(n - 1)$ -dimensional invariant torus. These tori are the support of corresponding minimal measures. The Lebesgue measure estimate on this set is uniformly valid for any perturbation.

1. INTRODUCTION

Consider a nearly integrable Hamiltonian system with n degrees of freedom ($n > 2$),

$$(1.1) \quad \dot{p} = -\frac{\partial H}{\partial q}, \quad \dot{q} = \frac{\partial H}{\partial p}, \quad p \in \mathbb{R}^n, \quad q \in \mathbb{T}^n,$$

where the Hamiltonian has the KAM form, i.e. $H(p, q) = N(p) + P(p, q)$, N is the main term and P is the small perturbation. We assume that N is a convex function in p . In this case the Hamiltonian can be transformed into a positive definite Lagrangian $L(q, \dot{q})$ by Legendre transformation,

$$L(q, \dot{q}) = \max_p \{\langle \dot{q}, p \rangle - H(p, q)\}.$$

To formulate his variational principle for Lagrange systems, Mather assumed the following conditions [M2]:

- (1) **Positive definiteness.** For every $q \in \mathbb{T}^n$, the Lagrangian function is strictly convex in velocity $\dot{q} \in T_q \mathbb{T}^n$: the Hessian H_{pp} is positive definite.
- (2) **Super-linear growth.** L has fiberwise superlinear growth: for every $q \in \mathbb{T}^n$, we have $L/\|\dot{q}\| \rightarrow \infty$ as $\|\dot{q}\| \rightarrow \infty$.
- (3) **Completeness.** All solutions of the Lagrange equations are defined on all \mathbb{R} .

In our case, these conditions are satisfied automatically. Completeness follows from energy conservation, the system under consideration is autonomous; super-linearity is not needed, due to energy conservation, what happens for large \dot{q} is irrelevant; positive definiteness follows from our assumption on N .

Received by the editors March 19, 2002 and, in revised form, March 22, 2004.

2000 *Mathematics Subject Classification.* Primary 37J40, 37J50.

Key words and phrases. KAM method, invariant torus, minimal measure.

The author was supported by the state basic research project of China “Nonlinear Sciences” (G2000077303).

Let \mathfrak{M} be the set of Borel probability measures μ on $T\mathbb{T}^n$ with compact support that are invariant with respect to the Lagrangian flow. For each $\mu \in \mathfrak{M}$ the action $A(\mu)$ of the measure μ is defined as

$$(1.2) \quad A(\mu) = \int L d\mu$$

and the rotation vector $\rho(\mu) \in H_1(\mathbb{T}^n, \mathbb{R})$ of μ is defined such that the equation

$$\langle \rho(\mu), [\lambda] \rangle = \int \lambda d\mu$$

holds for any closed 1-form λ on \mathbb{T}^n . Let $\mathfrak{M}_\rho \subset \mathfrak{M}$ be the subset of invariant measure whose rotation vector is ρ . For a smooth Lagrange system satisfying Mather's assumptions Mather proved the existence of minimal measures: for every rotation vector $\rho \in H_1(\mathbb{T}^n, \mathbb{R})$ there exists an invariant measure $\mu \in \mathfrak{M}$ which minimizes the Lagrangian action in \mathfrak{M}_ρ . Mather also proved the existence of c -minimal measures. Let λ_c be a closed smooth 1-form on \mathbb{T}^n whose de Rham cohomology is c ; we define the action

$$A_c(\mu) = \int (L - \lambda_c) d\mu.$$

It is also shown in [M2] that for each c there is an invariant measure which minimizes A_c . Its support is the graph of some Lipschitz function defined on the set $\pi(\text{support of } \mu)$, where π is the standard projection along the fiber of the tangent bundle. Consequently, an immediate problem is what the geometric structure of these minimal measures is. In the case of area-preserving twist diffeomorphisms the Aubry-Mather sets support the minimal measures [M1]. For higher-dimensional systems without further assumption, what has been proved is not beyond that KAM tori are the support of some minimal measures [M2]. It follows immediately that KAM tori are Lagrangian sub-manifolds [H2]. In a generic case, some more properties are explored in [Me], [Bo]. It is shown in [Bo], under some extra conditions besides the smallness, some lower-dimensional tori support the minimal measures.

In this paper, we are going to study the structure of the support of some minimal measures of nearly integrable systems without any assumption on the perturbations. The corresponding rotation vectors satisfy exactly one resonance condition. For simplicity, we assume that the Hamiltonian H is real analytical in a complex domain

$$\Sigma_{\sigma, \tau} = \{|\text{Im}q| \leq \tau, |p - p_o| \leq \sigma, \forall p_o \in \mathbb{D}\},$$

where \mathbb{D} is an open set in \mathbb{R}^n . The analyticity can be obviously replaced by some C^m smoothness by using some well-developed techniques, as used in [SZ]. Given an integer vector \bar{k} , let

$$\mathbb{U}_{\bar{k}} = \{p \in \mathbb{D} \mid \langle \omega(p), \bar{k} \rangle = 0\}.$$

This is an embedded $(n-1)$ -dimensional sub-manifold. For each point $\bar{p} \in \mathbb{U}_{\bar{k}}$ the unperturbed system has an invariant torus $p = \bar{p}$; it has the foliation into a family of $(n-1)$ -dimensional tori. By introducing a linear symplectic transformation $(p, q) \rightarrow (x, y)$

$$(1.3) \quad x = qI, \quad p = yI^t,$$

where $I = [k_1, k_2, \dots, k_n]$ is a uni-modular matrix, i.e. for all $k_i \in \mathbb{Z}^n$, $k_n = \bar{k}$ and $\det I = 1$, we find that in the new coordinates the rotation vector has the form $(\omega_1, \omega_2, \dots, \omega_{n-1}, 0) = (\omega, 0)$. In the following, we still use (p, q) to denote the

new coordinates. Without loss of generality, we can assume $\mathbb{U}_{\bar{k}}$ is the graph of the function $\Phi_0 : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$, $\mathbb{U}_{\bar{k}} = \{\hat{p}, \Phi_0(\hat{p})\}$, ($\hat{p} = (p_1, \dots, p_{n-1})$).

Under small perturbations, these n -tori in $\mathbb{U}_{\bar{k}}$ break down in general, but some lower-dimensional tori still exist. Strong evidence shows that for most actions in $\mathbb{U}_{\bar{k}}$, there are at least two $(n - 1)$ -dimensional tori surviving perturbations even if the corresponding n -torus is destroyed in the case that the rotation vector ω satisfies the diophantine condition

$$(1.4) \quad |\langle \omega, k \rangle| \geq D|k|^{-\nu}, \quad \forall k \in \mathbb{Z}^{n-1} \setminus \{0\}, \nu \geq n - 1.$$

In this paper we shall show that some of these $(n - 1)$ -tori support the relevant c -minimal measures. As the configuration manifold is \mathbb{T}^n , we can use cdq ($c = const.$) as the representative element of first de Rham cohomology class $[cdq]$, and identify the point $c \in \mathbb{R}^n$ with the corresponding cohomology class when necessary.

Theorem 1.1. *Let the Hamiltonian H be of KAM type $H(p, q) = N(p) + P(p, q)$ which is real analytical in $\Sigma_{\sigma, \tau}$ and convex in p . There exists a positive number $d = d(N, n, D, \sigma, \tau, \nu, |\bar{k}|) > 0$, such that if the sup-norm of P in $\Sigma_{\sigma, \tau}$, $\|P\| < d$, then there is a set $\mathbb{S} \subset \mathbb{R}^{n-1}$, a map $\mathfrak{F}_\infty : \mathbb{S} \rightarrow \mathbb{R}^{n-1}$ and a set function $\Phi_\infty : \mathbb{S} \rightarrow \{\text{non-empty closed set of } \mathbb{R}^1\}$ such that:*

- (1) *For each $\hat{p} \in \mathbb{S}$, there exists a non-empty set $\Theta(\hat{p}) \subseteq \mathbb{T}$. Corresponding to each $u \in \Theta(\hat{p})$ there exists one $(n-1)$ -dimensional invariant torus. These $(n - 1)$ -tori have the same frequency $\mathfrak{F}_\infty(\hat{p})$.*
- (2) *For each $\hat{p} \in \mathbb{S}$, the corresponding tori are the support of $c(\hat{p})$ -minimal measures, where $c(\hat{p}) = (\hat{p}, \Phi_\infty(\hat{p}))$.*
- (3) *\mathfrak{F}_∞ need not be continuous, but its inverse \mathfrak{F}_∞^{-1} is Lipschitz, and is a small perturbation of $(\frac{\partial N}{\partial \hat{p}})^{-1}$,*

$$(1.5) \quad \left\| \mathfrak{F}_\infty^{-1} - \left(\frac{\partial N}{\partial \hat{p}} \right)^{-1} \right\|_\infty \leq 2d^{\frac{6}{13}}.$$

- (4) *The set-valued function Φ_∞ is defined by $\{\Phi_\infty(\hat{p})\} = \bigcup_{u \in \Theta(\hat{p})} V_\infty(\hat{p}, u)$, where $V_\infty(\hat{p}, u)$ is smooth in u over $\Theta(\hat{p})$ in the sense of Whitney and*

$$(1.6) \quad \max_{u \in \Theta(\hat{p})} |V_\infty(\hat{p}, u) - \Phi_0(\hat{p})| \leq 2d^{\frac{9}{13}}, \quad \hat{p} \in \mathbb{S}.$$

- (5) *\mathbb{S} and its image under \mathfrak{F}_∞ have positive $(n - 1)$ -dimensional Lebesgue measure*

$$(1.7) \quad \mathbf{m}(\mathbb{S} \cap B_1) \geq \left(1 - \beta d^{\frac{2}{169}}\right) \mathbf{m}(B_1),$$

$$(1.8) \quad \mathbf{m}(\mathfrak{F}_\infty(\mathbb{S}) \cap B_1) \geq \left(1 - 2\beta d^{\frac{2}{169}}\right) \mathbf{m}(B_1),$$

where β is a positive number depending only on N and n , and B_1 is an arbitrarily centered $(n - 1)$ -dimensional unit ball. Both $\mathbf{m}(\mathbb{S})$ and $\mathbf{m}(\mathfrak{F}_\infty(\mathbb{S}))$ approach to full measure as $d \rightarrow 0$.

The small number d is uniformly valid for all perturbations, independent of their special form. The existence of these geometrically regular minimal measures requires no other conditions than what the classical KAM theory requires. It is shown in [Bo] that, under some non-degeneracy conditions, an m -dimensional torus, survived from $(n - m)$ -resonant n -torus ($[\Gamma]$), is the support of corresponding minimal measure. The condition is that the maximal point of the first order averaging

function is non-degenerate, i.e. all eigenvalues are negative. In that case, the proof is much easier, but the small number d must depend on the smallest eigenvalue in that case. $d \rightarrow 0$ as the smallest eigenvalue goes to zero.

To find Arnold diffusion orbits, one way is to construct connecting orbits between different minimal measures, for instance, along the line of Mather [M3], [M4]. To be the candidates, the support of these minimal measures cannot cover the whole configuration space under the projection along the tangent fiber. Lower-dimensional minimal tori satisfy such a condition. This motivates the author to write this paper.

Recall the result obtained in [M2] that for every rotation vector $\omega \in \mathbb{R}^n$, there exists a minimal measure μ such that $\rho(\mu) = \omega$. This does not imply that the orbits in this measure also have the same rotation vector. An example is the Hedlund geodesic flow on \mathbb{T}^3 ([H1], [Ba]). The following example shows that such a phenomenon occurs even in integrable systems. Let us consider the Hamiltonian H_ϵ

$$H_\epsilon(p, q) = \sum_{i=1}^{n-1} \omega_i^* p_i + \frac{1}{2} \sum_{i=1}^{n-1} (p_i + \epsilon \cos q_n)^2 + \frac{1}{2} p_n^2.$$

Let $\Omega_\epsilon = [\omega_1^* - \epsilon, \omega_1^* + \epsilon] \times \cdots \times [\omega_{n-1}^* - \epsilon, \omega_{n-1}^* + \epsilon] \times \{0\} \subset \mathbb{R}^n$. For each rotation vector $\omega \in \Omega_\epsilon$, there are some $(n-1)$ -dimensional invariant tori whose frequency is exactly ω , but none of them are in the support of the minimal measure $\mu(\omega)$ in the class \mathfrak{M}_ω . Any orbit in the minimal measure does not have rotation vector ω when $\omega \in \Omega_\epsilon$.

2. OUTLINE OF THE PROOF

From the technical point of view, the main part of the proof uses KAM type methods. What is new in this direction is overcoming the difficulty created by the lack of continuity of the frequency ω as a function of the action p . To our knowledge, all KAM-type work up to now treated ω either as an independent parameter or as a continuous function of actions. In the general case, such discontinuity may ruin the whole KAM iteration scheme, as one cannot get a positive measure estimate. Fortunately, the convex property of the frequency map enables us to overcome this difficulty. Section 3 is devoted to investigating those properties of the frequency map that are used to get a positive measure estimate, which is done in the section 7. In sections 4 and 5, under the condition that $\omega(p)$ satisfies the diophantine condition, the existence of $n-1$ -dimensional tori is proved by constructing symplectic change of coordinates that can cope with the degeneracy of the critical point of the first order average function. Section 6 is devoted to prove the minimality of the Lagrange action when it takes values on these tori.

In principle, the existence proof of $(n-1)$ -dimensional invariant tori is completed by employing a sequence of symplectic change of coordinates, $\mathfrak{I}_1, \mathfrak{I}_2, \dots, \mathfrak{I}_k, \dots$, converging to an identity map exponentially fast. Let $\mathfrak{S}_k = \mathfrak{I}_1 \circ \mathfrak{I}_2 \circ \cdots \circ \mathfrak{I}_k$. After k iteration steps, the Hamiltonian has the form

$$H_k = H \circ \mathfrak{S}_k = N_k + P_k,$$

with much smaller error term P_k in a suitably smaller domain. For instance,

$$\|P_{k+1}\| \leq \|P_k\|^\kappa$$

for some $\kappa > 1$. The limit of \mathfrak{S}_k maps a trivially embedded $(n - 1)$ -dimensional torus $\Gamma : (p, q_n) = (\tilde{p}, \tilde{q}_n)$ to an $(n - 1)$ -torus with its tangential map taking a constant vector field on Γ to the vector field given by (1.1).

To describe the construction in more detail, let us drop the index k to simplify notation. For one step of iteration we introduce the generating function W so that the time-1-map of the Hamiltonian flow Φ_W^t is the coordinate transformation $\mathfrak{T} : (\bar{p}, \bar{q}) \rightarrow (p, q)$ and

$$(2.1) \quad (N + P) \circ \mathfrak{T} = N + P + \{N, W\} + \int_0^1 \{(1 - t)\{N, W\} + P, W\} \circ \Phi_W^t dt.$$

The integral is quadratic in P and W , and will be part of the new error term.

The point is to find W such that $N + (P - P_1) + \{N, W\}$ is a normal form again, where P_1 is exponentially smaller than P . The difficulty emerged here is that N depends not only on p but also on q_n after one step of iteration, as the frequency satisfies a resonance condition. To be more precise, we assume that the frequency $(\omega, 0) = \frac{\partial N}{\partial p}(p_0)$ satisfies the diophantine condition (1.4), where $p_0 \in \mathbb{U}_k$. For the first step of iteration we choose the generating function W which is the solution of the homological equation

$$(2.2) \quad \langle (\omega, 0), W_q \rangle = -P(p, q) + [P(p, q_n)],$$

where

$$[P(p, q_n)] = \frac{1}{(2\pi)^{n-1}} \int_{\mathbb{T}^{n-1}} P(p, q) dq_1 dq_2 \cdots dq_{n-1}.$$

Therefore, the new normal form $N + [P]$ depends on q_n . It implies that the invariant n -torus is destroyed in general. However, some $(n - 1)$ tori may still exist. Let $p_n(p_1, \dots, p_{n-1}, q_n)$ be the function that solves the equation $\frac{\partial N}{\partial p_n} = 0$. If the error term vanishes, each critical point of $N(p_1, \dots, p_{n-1}, p_n(p_1, \dots, p_{n-1}, q_n), q_n)$, q_n^* , corresponds to an $n - 1$ -dimensional invariant torus

$$\begin{aligned} p_i(t) &= \text{constant} & (i < n), \\ q_i(t) &= \omega_i t + q(0) & (i < n), \\ q_n(t) &= q_n^*, & p_n(t) = p_n(p_1, \dots, p_{n-1}, q_n^*). \end{aligned}$$

What we are interested in is the point where N reaches its maximum as the function of q_n .

To simplify notation, we shall use (p, q) to denote $(p_1, \dots, p_{n-1}, q_1, \dots, q_{n-1})$ and use (u, v) to denote (q_n, p_n) in the following.

The change of coordinates for the following iteration steps is constructed in a neighborhood of an $(n - 1)$ -torus instead of the n -torus. We design two types of coordinate transformations according to whether the critical point has “strong persistency” with respect to the perturbation. The “strong persistency” is such a condition that guarantees that the critical point of $N_k + [P_k]$ remains sufficiently close to the corresponding critical point of N_k . There is no precise criterion to distinguish whether the critical point has strong persistency or not. It depends on how we run the KAM machine. In this paper, we say it has strong persistency when the following holds:

$$\left| \frac{\partial N_k}{\partial u}(p, v(p, u^*), u^*) \right| \geq |P_k|^{\frac{2}{13}},$$

otherwise, we say it has “weak persistency”.

The construction of coordinate transformation in the case of “weak persistency” is much more complicated than in the case of “strong persistency”. We need to retain a sufficiently large neighborhood of the $(n - 1)$ -torus so that the critical point of $N_k + [P_k]$, via the same variational principle, still falls into this domain. On the other hand, this domain cannot be too large, in order not to destroy the exponentially fast convergence of the error term. This domain is chosen in an implicit way, different from the usual KAM way. The details are shown in section 4.

When we run the KAM iteration scheme the frequency vector $\omega(p)$ is assumed to satisfy the diophantine condition at every step. We need to show there is such a positive measure set in $\mathbb{U}_{\bar{k}}$, where this assumption is really true. Towards that goal, let us consider the frequency map $\mathfrak{F}_j: p \rightarrow \omega$ at j -th step of iteration

$$(2.3) \quad \omega(p) = \left\{ \frac{\partial N_j}{\partial p}(p, u^*, v(p, u^*)) \mid u^* \text{ is the maximal point of } N_j(p, u, v(p, u)) \right\}.$$

This map may be discontinuous in p and may be multi-valued when $j \neq 0$. Without any further conditions, such discontinuity may ruin the positive measure proof. Thanks to the convexity, the inverse of this frequency map is single-valued, Lipschitz and structurally stable in the following sense:

$$(2.4) \quad \|\mathfrak{F}_k^{-1} - \mathfrak{F}_j^{-1}\| \leq O(\sqrt{\delta}),$$

if $\|N_k - N_j\| \leq \delta$. This property guarantees that the set $\bigcap_{j=0}^\infty \mathfrak{F}_j^{-1}(\Omega(n - 1, D))$ has positive measure. Here we use $\Omega(m, D)$ to denote all m -dimensional diophantine vectors with coefficient D . We can see from (1.4) that Ω also depends on the diophantine exponent ν . Since $\nu \geq n - 1$ is already fixed, we omit the dependence of Ω on ν . The choice of D is carefully done so that we obtain the measure estimate of the KAM tori. Also thanks to the convexity, we find that the image of $\mathbb{U}_{\bar{k}}$ under \mathfrak{F}_j has positive Lebesgue measure. Indeed, if $\|N_j - N_0\|_{C^2} \leq \varepsilon$, we have

$$(2.5) \quad \begin{aligned} m\left(\mathfrak{F}_0(\mathbb{U}_{\bar{k}}) \setminus \mathfrak{F}_j(\mathbb{U}_{\bar{k}})\right) &= O(\varepsilon) m\left(\mathfrak{F}_0(\mathbb{U}_{\bar{k}})\right), \\ \mathfrak{F}_0(\mathbb{U}_{\bar{k}}) + \varepsilon &\supseteq \mathfrak{F}_j(\mathbb{U}_{\bar{k}}), \end{aligned}$$

where we use $S + a$ to denote the set $\{x \mid \text{dist}(x, S) \leq a\}$.

The final step is to show the minimality of these tori. We extend each coordinate transformation \mathfrak{F}_j to a neighbourhood covering the n -dimensional torus $(p, v) = (\tilde{p}, \tilde{v}) \in \mathbb{U}_{\bar{k}}$. Under the map $\mathfrak{S}_\infty = \lim_{j \rightarrow \infty} \mathfrak{S}_j$, the Hamiltonian $H_\infty = H \circ \mathfrak{S}_\infty$ has some $(n - 1)$ -tori $\{p = \tilde{p}, u = \tilde{u}, v = v(\tilde{p}, \tilde{u})\}$. By introducing the Legendre transformation,

$$\mathfrak{L} : (p, q, u, v) \in T^*\mathbb{T}^n \rightarrow (\dot{q}, \dot{u}, q, u) \in T\mathbb{T}^n, \quad (\dot{q}, \dot{u}) = \left(\frac{\partial H}{\partial p}, \frac{\partial H}{\partial v} \right).$$

We obtain the corresponding Lagrangian function

$$(2.6) \quad L(q, u, \dot{q}, \dot{u}) = \max_{p, v} \left(\langle (\dot{q}, \dot{u}), (p, v) \rangle - H_\infty(p, q, u, v) \right).$$

Let

$$L_\lambda = L(q, u, \dot{q}, \dot{u}) - \eta_\lambda,$$

where η_λ is a closed 1-form, in local coordinates, $\eta_\lambda = \langle p, \dot{q} \rangle + v(p, u)\dot{u}$ ($p = \text{const.}$). If we can show that the Lagrangian L_λ takes its minimal value at every point of the torus Γ , then the following holds:

$$\int_\Gamma L_\lambda d\mu_\Gamma \leq \int L_\lambda d\mu$$

for any invariant measure μ , where μ_Γ is the $(n - 1)$ -Lebesgue measure on Γ . Obviously, ν is invariant to the Hamiltonian flow.

Since L_λ is strictly convex in (\dot{q}, \dot{u}) , the equation $\partial_{\dot{q}}L_\lambda = \partial_{\dot{u}}L_\lambda = 0$ has the solution $(p = \tilde{p}, u = \tilde{u}, v = v(\tilde{p}, \tilde{u}))$. We obtain from (2.6) that $\min_{\dot{q}, \dot{u}} L_\lambda(q, u, \dot{q}, \dot{u}) = -H_\infty(p, q, u, v(p, u))$. We shall show that for p fixed, $H_\infty(p, q, u, v(p, u))$ takes its maximum in the variable of (q, u) everywhere on an $(n - 1)$ -torus $\{u^*\} \times \mathbb{T}^{n-1}$. It implies that such an $(n - 1)$ -torus is the support of corresponding minimal measures.

3. THE FREQUENCY MAP \mathfrak{F}

At each step of KAM iteration we have the frequency map $\mathfrak{F}: \mathbb{R}^{n-1} \rightarrow \mathbb{R}^{n-1}$

$$(3.1) \quad \omega = \mathfrak{F}(p) = \frac{\partial N}{\partial p}(p, v(p, u^*), u^*),$$

where $N(p, u)$ attains its maximum at $u = u^*$. In this section we show some properties of the frequency map that are used for measure estimate, in particular, in the case of weaker persistency.

Let us consider a function $H(x, y) = H_0(y) + H_1(x, y)$, $x \in S^1$, $y \in \mathbb{R}^n$. We assume H_0 is convex in y , $\|H_1\|_{C^2} \leq \epsilon$. Let $\lambda > 0$ be the smallest eigenvalue of the Hessian matrix of H_0 . If ϵ is suitably small, H is also convex in y . Since we are considering the existence of codimension 1 tori which correspond to the maximum of $H(x, y)$ as the function of x , the frequency is given by the value $\partial_y H$ takes at a maximum point, it is natural to study the function

$$F(y) = \max_{x \in S^1} H(x, y).$$

Clearly, F is also a convex function of y , for any $0 \leq \lambda \leq 1$, $y_1, y_2 \in \mathbb{R}^n$, and there exists $x_0 \in S^1$ such that

$$\begin{aligned} F(\lambda y_1 + (1 - \lambda)y_2) &= H(x_0, \lambda y_1 + (1 - \lambda)y_2) \\ &\leq \lambda H(x_0, y_1) + (1 - \lambda)H(x_0, y_2) \\ &\leq \lambda F(y_1) + (1 - \lambda)F(y_2). \end{aligned}$$

To go further, we assume first that $y \in \mathbb{R}^1$. In this case, F has left and right derivative everywhere,

$$F_y^-(y) = \lim_{y_1 \nearrow y} \frac{F(y_1) - F(y)}{y_1 - y}, \quad F_y^+(y) = \lim_{y_1 \searrow y} \frac{F(y_1) - F(y)}{y_1 - y}.$$

F_y is monotonically increasing, discontinuous only at countably many points, and and differentiable almost everywhere. Let

$$\begin{aligned} \mathbb{M}(y) &= \{x | H(x, y) = \max_{\theta \in S^1} H(\theta, y)\}, \\ \mathbb{M}^+(y) &= \{x | \partial_y H(x, y) = \max_{\theta \in \mathbb{M}} \partial_y H(\theta, y)\}, \\ \mathbb{M}^-(y) &= \{x | \partial_y H(x, y) = \min_{\theta \in \mathbb{M}} \partial_y H(\theta, y)\}. \end{aligned}$$

We claim that

$$(3.2) \quad \begin{aligned} F_y^+(y) &= \partial_y H(x, y), \quad \forall x \in \mathbb{M}^+(y), \\ F_y^-(y) &= \partial_y H(x, y), \quad \forall x \in \mathbb{M}^-(y). \end{aligned}$$

Obviously, we have $H(x, y') - H(x, y) \leq F(y') - F(y)$, where $x \in \mathbb{M}^+(y)$. It follows that $F_y^+ \geq \partial_y H(x, y)$. To show that $F_y^+ < \partial_y H(x, y) + \delta$ for arbitrary small positive δ , let us make use of the fact that there exists at least one $x \in \mathbb{M}^+(y)$ so that (x, y) is the accumulation point of (x', y') as $y' \searrow y$, where $x' \in \mathbb{M}(y')$. Let $y' > y$ be sufficiently close to y so that $|\partial_y H(x', y) - \partial_y H(x, y)| < \delta$ for some $x' \in \mathbb{M}(y')$. The conclusion follows from

$$\begin{aligned} H(x', y') - H(x, y) &\leq H(x', y') - H(x', y) \\ &< (\partial_y H(x, y) + \delta)(y' - y) + O(|y' - y|^2). \end{aligned}$$

Therefore, the condition that F is differentiable at y is $\mathbb{M}^+(y) = \mathbb{M}^-(y)$; we use ∇F to denote its differential.

Let us assume that $y \in \mathbb{R}^n$ again. The above arguments hold when we study the directional derivatives, from which we see that the differential of F , $\nabla F = (\partial_{x_1} F, \partial_{x_2} F, \dots, \partial_{x_n} F)$, is precisely the frequency map when restricted to the set where F is differentiable. Obviously, this set is a full Lebesgue measure set (cf. [R]). The frequency map \mathfrak{F} is multi-valued on the set where F is not differentiable, but the measure of this set is zero. If F is not differentiable at y , the set of all subgradients at y , $\{\nabla F(y)\}$ is a closed convex set. In this case, $\{\mathfrak{F}(y)\} \subseteq \{\nabla F(y)\}$.

In the m -th step of KAM iteration, the frequency map \mathfrak{F}_m is the perturbation of the frequency map \mathfrak{F}_{m-1} for the last step of the iteration. The next lemma shows how the image of \mathfrak{F}^{-1} changes when the function H undergoes small perturbation.

Lemma 3.1. *Let $F_i(y) = H_0(y) + \max_{x \in \mathbb{T}} H_i(x, y)$ ($i = 1, 2$). We assume the following:*

- (1) H_0 is C^2 -differentiable and convex, and $\lambda > 0$ is the smallest eigenvalue of the Hessian matrix of H_0 .
- (2) Both H_1 and H_2 are small, $\|H_k\|_{C^2} \leq \frac{\lambda}{2}$ and $\|H_1 - H_2\|_\infty \leq \delta$.

Then

$$(3.3) \quad \|(\nabla F_1)^{-1} - (\nabla F_2)^{-1}\|_\infty \leq 2\sqrt{\frac{\delta}{\lambda}}.$$

Proof. Clearly, $(\nabla F)^{-1}$ is single-valued although ∇F might be multi-valued. Let $z \in \nabla F_1(y_1) \cap \nabla F_2(y_2)$. If we rewrite the function F_i in the form

$$\begin{aligned} F_i(y) &= \frac{1}{2}H_0(y) - \frac{1}{2}\langle \nabla H_0(y_i), y - y_i \rangle + G_i(y), \\ G_i(y) &= \frac{1}{2}H_0(y) + \frac{1}{2}\langle \nabla H_0(y_i), y - y_i \rangle + \max_{x \in \mathbb{T}} H_i(x, y), \quad i = 1, 2, \end{aligned}$$

we can see that G_i is convex. By virtue of this observation, as z is a sub-gradient of F_1 at y_1 and of F_2 at y_2 , we have

$$\begin{aligned} F_1(y_2) - F_1(y_1) &\geq \langle z, y_2 - y_1 \rangle + \frac{\lambda}{4}\|y_2 - y_1\|^2, \\ F_2(y_1) - F_2(y_2) &\geq \langle z, y_1 - y_2 \rangle + \frac{\lambda}{4}\|y_2 - y_1\|^2, \end{aligned}$$

from which we find

$$(F_1(y_2) - F_2(y_2)) - (F_1(y_1) - F_2(y_2)) \geq \frac{\lambda}{2} \|y_2 - y_1\|^2.$$

On the other hand, as $\exists x_i \in \mathbb{T}$ such that $F_i(y_i) = H_0(y_i) + H_i(x_i, y_i)$, we have $|F_1(y) - F_2(y)| \leq \delta$. This proves the lemma.

Let B be an arbitrary set in the frequency space; this lemma implies that $\nabla F_2^{-1}(B) \subset \nabla F_1^{-1}(B) + 2\sqrt{\frac{\delta}{\lambda}}$. This property plays the role the Kolmogorov non-degeneracy does in the paper of V.I. Arnold [A]. It is crucial for the positive measure estimate.

Lemma 3.2. *Assume $H(x, y) = H_0(y) + H_1(x, y)$ ($x \in \mathbb{T}^1, y \in \mathbb{R}^n$), in which H_0 is convex in $y, \|H_1\|_{C^2} \leq \epsilon$, where ϵ is sufficiently small. Let $\lambda > 0$ be the smallest eigenvalue of the Hessian matrix of H_0 . Then the image of the frequency map \mathfrak{F} has near full Lebesgue measure, i.e. for each measurable set B , we have*

$$(3.4) \quad m(\mathfrak{F}(B)) \geq (1 - \frac{\epsilon}{\lambda})^n m\left(\frac{\partial H_0}{\partial y}(B)\right),$$

$$(3.5) \quad \mathfrak{F}(B) \subset \frac{\partial H_0}{\partial y}\left(B + \frac{\epsilon}{\lambda}\right).$$

Moreover, if we assume $\xi > 0$ be a small number and let $B_\xi \subset B$ be a set such that the map ∇F expands its volume at least ξ^{-1} times, i.e.

$$(3.6) \quad m(\nabla F(B_\xi)) \geq \frac{1}{\xi} m\left(\frac{\partial H_0}{\partial y}(B_\xi)\right),$$

then we have

$$(3.7) \quad m(B_\xi) \leq c_0 \xi \epsilon m(B),$$

where c_0 is a positive number, independent of ξ and ϵ .

From now on we use c_i to denote some positive constant, which is independent of involved small parameters.

Proof. We have shown that $\mathfrak{F} = \nabla F$ on the set where F is differentiable, thus we only need to study the map ∇F . If we rewrite F in the form $F(y) = (1 - \frac{\epsilon}{\lambda})H_0(y) + G(y)$, where $G(y) = \frac{\epsilon}{\lambda}H_0(y) + \max_{x \in \mathbb{T}} H_1(x, y)$, then G is a convex function of y .

We introduce a smoothing function

$$\alpha_a(x) = \begin{cases} e^{-\frac{1}{a^2 - \|x\|^2}} & \|x\| < a, \\ 0 & \text{otherwise,} \end{cases}$$

$$\phi_a(x) = \alpha_a(x) \left(\int_{\mathbb{R}^n} \alpha_a(x) dx \right)^{-1},$$

and define

$$G_a(y) = \int_{\mathbb{R}^n} G(y + x)\phi_a(x)dx, \quad F_a(y) = (1 - \frac{\epsilon}{\lambda})H_0(y) + G_a(y).$$

It is easy to check that $G_a \in C^\infty$ is also convex. For a bounded set $B \subset \mathbb{R}^n$ and $\delta > 0 \exists a(B, \delta) > 0$ such that $\|G_a - G\|_{C^0(B)} < \delta$ if $a < a(B, \delta)$.

Obviously, ∇F_a^{-1} is a Lipschitz map, the Lipschitz constant not bigger than $(\lambda - \epsilon)^{-1}$, as $\|\nabla F_a(y_1) - \nabla F_a(y_2)\| \geq (\lambda - \epsilon)\|y_1 - y_2\|$. By virtue of Lemma 3.1,

we see that ∇F^{-1} is also a Lipschitz function. As both G_a and H_0 are convex and differentiable, $\nabla H_0 \circ \nabla F_a^{-1}$ is a Lipschitz map from $\nabla F_a(B)$ to $\nabla H_0(B)$, the Lipschitz constant not bigger than $(1 - \epsilon/\lambda)^{-1}$. By Lemma 3.1 we see that $\nabla H_0 \circ \nabla F^{-1}$ is also a Lipschitz map with constant not bigger than $(1 - \epsilon/\lambda)^{-1}$. Let B be the set where F is differentiable. This set is clearly measurable. Thus $\nabla F(B)$ is also measurable, consequently. $\nabla F(B)^{-1}$ is continuous. Therefore, we have

$$(3.8) \quad m(\nabla H_0(B)) \leq (1 - \frac{\epsilon}{\lambda})^{-n} m(\nabla F(B)).$$

Since ∇H_0 is a diffeomorphism, ∇F is differentiable almost everywhere, and (3.4) follows from (3.8). (3.5) is obvious.

To show (3.7) let us apply (3.4) as well as (3.5). We find that (3.7) is an immediate consequence of the following argument:

$$\begin{aligned} m\left(\nabla H_0\left(B + \frac{\epsilon}{\lambda}\right)\right) &\geq m(\nabla F(B_\epsilon)) + m(\nabla F(B \setminus B_\epsilon)) \\ &\geq \frac{1}{\xi} m(\nabla H_0(B_\epsilon)) + \left(1 - \frac{\epsilon}{\lambda}\right)^n m(\nabla H_0(B \setminus B_\epsilon)). \end{aligned}$$

In the next section we shall construct two types of KAM iteration schemes according to whether the critical point of $N(p, u, v(p, u))$ is degenerate or not. For the degeneracy case, we choose a neighborhood for u , denoted by $\mathbb{L}(p)$, big enough so that $\max_{u \in \mathbb{L}} N(p, u, v(u)) - N(p, \partial \mathbb{L}(p), v(p, \partial \mathbb{L}(p))) \geq 2[P(p, u, v(p, u))]$. This guarantees that the maximal point of $N + [P](p, u, v(p, u))$ remains in the same neighborhood. In this case, $\mathbb{L}(\tilde{p})$ might be as big as of order $O(1)$, consequently, the variation of $\frac{\partial N}{\partial p}(p, u, v(p, u))$ over $\mathbb{L}(p)$ might not be small. This would ruin the KAM iteration scheme. Fortunately, it does not happen for most p ; we show it in the following.

Lemma 3.3. *Assume $H(x, y) = H_0(y) + H_1(x, y)$, where H_0 is convex, $\|H_1\|_{C^2} < \epsilon$ which is small. Let $\mathbb{L}(s, y) = \{x \in \mathbb{T} \mid F(y) - H(y, x) \leq s^8\}$, $A(s) = \{y \in \mathbb{R}^n \mid \exists x \in \mathbb{L}(s, y) \text{ s.t. } |F_y(y) - H_y(x, y)| > s^2\}$. There exists $\epsilon_0 > 0$, such that if $\epsilon \leq \epsilon_0$, then the s^4 -neighborhood of $A(s)$ has small Lebesgue measure:*

$$(3.9) \quad m((A(s) + s^4) \cap B_1) \leq c_1 s^2 \epsilon m(B_1),$$

where B_1 is a unit ball arbitrarily centered, both ϵ_0 and c_1 depend only on H_0 and n .

Proof. By the arguments in the proof of Lemma 3.2, we know that the Lipschitz constant of the map ∇F^{-1} is not bigger than $(\lambda - \epsilon)^{-1}$. Denote by $\partial_e^\pm F$ the one-side derivative of F in the direction of $e \in \mathbb{R}^n$ and define $J_s = \{y \in \mathbb{R}^n : \exists e \in \mathbb{R}^n, \|e\| = 1, \text{ s.t. } |\partial_e^+ F(y + s^4 e) - \partial_e^- F(y)| \geq \frac{1}{2} s^2\}$. By the smallness assumption on ϵ , we see that $J_s = \emptyset$ if $s \geq \epsilon^{\frac{1}{3}}$. Obviously, given $y \in J_s$ and relevant $e, \exists z_1^* \in \nabla F(y), z_2^* \in \nabla F(y + s^4 e)$ such that

$$\|z_1^* - z_2^*\| = \max_{z_1 \in \nabla F(y), z_2 \in \nabla F(y + s^4 e)} \|z_1 - z_2\| \geq \frac{1}{2} s^2.$$

Denote by $I(y, e, s)$ the segment $\{y + te : 0 \leq t \leq s^4\}$, let $\delta_1 = \delta(1 - \frac{\epsilon}{\lambda})$ and let $C(z_1, z_2)$ be a curve joining z_1^* and z_2^* and passing through $\nabla F(y + te)$ ($0 \leq t \leq s^4$).

We see that for any $z \in C(z_1, z_2)$, $\nabla H_0 \circ \nabla F^{-1}$ maps $B_{\delta_1}(z)$ into $\nabla H_0(I) + \delta$. This implies that

$$(3.10) \quad \mathbf{m}(\nabla F(I(y, e, s)) + \delta_1) \geq \frac{\frac{1}{2}s^2 + 2\delta_1}{\Lambda(s^4 + 2\delta)} \left(\frac{\delta_1}{\delta}\right)^{n-1} \mathbf{m}(\nabla H_0(I(y, e, s)) + \delta),$$

where Λ is the largest eigenvalue of the Hessian matrix of H_0 . Let $E(y, s) = \{e \in \mathbb{R}^n : \|e\| = 1, |\partial_e^+ F(y + s^4 e) - \partial_e^- F(y)| \geq \frac{1}{2}s^2\}$ and

$$\mathbb{J} = \bigcup_{y \in J_s, e \in E(y, s)} I(y, e, s) + s^4.$$

Let $\delta = s^4$; it follows from (3.10) that

$$(3.11) \quad \mathbf{m}(\nabla H_0(\mathbb{J})) \leq 6\Lambda s^2 \left(1 - \frac{\epsilon}{\lambda}\right)^{1-n} \mathbf{m}(\nabla F(\mathbb{J})).$$

By (3.5) we find that $\nabla F \circ \nabla H_0^{-1}$ maps B_1 into $B_{1+\epsilon/\lambda}$. Thus we have

$$B_{1+\epsilon/\lambda} \setminus \nabla F \circ \nabla H_0^{-1}(B_1 \setminus \nabla H_0(\mathbb{J})) \supseteq \nabla F(\mathbb{J} \cap \nabla H_0(B_1)).$$

By applying (3.4) on $\nabla H_0^{-1}(B_1) \setminus \mathbb{J}$, we obtain from (3.11) that

$$\begin{aligned} \mathbf{m}(B_{1+\epsilon/\lambda}) - \left(1 - \frac{\epsilon}{\lambda}\right)^n \mathbf{m}(B_1 \setminus \nabla H_0(\mathbb{J})) &\geq \mathbf{m}(B_{1+\epsilon/\lambda}) - \mathbf{m}(\nabla F(\nabla H_0^{-1}(B_1) \setminus \mathbb{J})) \\ &\geq \frac{\mathbf{m}(B_1 \cap \nabla H_0(\mathbb{J}))}{6s^2 \left(1 - \frac{\epsilon}{\lambda}\right)^{n-1}}. \end{aligned}$$

Consequently

$$\mathbf{m}(\nabla H_0(\mathbb{J}) \cap B_1) \leq \frac{6s^2 \left(1 - \frac{\epsilon}{\lambda}\right)^{1-n}}{1 - 6s^2 \left(1 - \frac{\epsilon}{\lambda}\right)^{1-n}} \left(\left(1 + \frac{\epsilon}{\lambda}\right)^n - \left(1 - \frac{\epsilon}{\lambda}\right)^n \right) \mathbf{m}(B_1).$$

Thus, $\exists \epsilon_0 > 0$ and $c_1 > 0$ such that if $\epsilon \leq \epsilon_0$, then the term on the right-hand side of the above inequality is not bigger than $c_1 s^2 \epsilon \mathbf{m}(B_1)$. As ∇H_0 is a diffeomorphism, we have

$$\mathbf{m}(\mathbb{J} \cap B_1) \leq c_1 s^2 \epsilon \mathbf{m}(B_1).$$

What remains to prove is $A \subseteq J_s$; this implies that $A + s^4 \subseteq \mathbb{J}$. If not, there would be $y \in A$ but $y \notin \mathbb{J} - s^4$. By the definition of $A(s)$, the following hold for some $x \in \mathbb{T}$, $\|e\| = 1$ and $z \in \nabla F(y)$:

$$0 \leq F(y) - H(x, y) \leq s^8, \quad |\langle z - \nabla_y H(x, y), e \rangle| > s^2.$$

We assume that $\langle \nabla_y H(x, y) - z, e \rangle > s^2$; the proof for the case $\langle \nabla_y H(x, y) - z, e \rangle < s^2$ is similar. Because of $y \notin J_s$, we have

$$F(y + s^4 e) - F(y) \leq \max_{0 \leq \eta \leq s^4} \partial_e^+ F(y + \eta e) s^4 \leq (\partial_e^+ F(y) + \frac{1}{2}s^2) s^4.$$

If s is suitably small, we would have

$$\begin{aligned} H(x, y + s^4 e) - F(y + s^4 e) &\geq H(x, y) - F(y) - \frac{1}{2}\Lambda s^8 \\ &\quad + \left(\langle \nabla_y H(x, y), e \rangle - (\partial_e^+ F(y) + \frac{1}{2}s^2) \right) s^4 \\ &> -\left(1 + \frac{1}{2}\Lambda\right) s^8 + s^6 - \frac{1}{2}s^6 > 0, \end{aligned}$$

but this contradicts the definition of F . This completes the proof of the lemma.

4. THE KAM ITERATION

In this section we construct one step of KAM iteration. It has two types according to whether the critical point has weak persistency or not. We assume that the frequency $\omega = \mathfrak{F}(p)$ satisfies the diophantine condition (1.4).

4.1. **Weak persistency.** By definition, the following holds at \tilde{u} , the maximum point of $N(p, u, v(u))$ in u :

$$(4.1) \quad \left| \frac{d^2 N}{du^2}(p, \tilde{u}, v(p, \tilde{u})) \right| \leq d^{\frac{2}{13}}.$$

Under such a condition the critical point of $(N + [P])(p, u, v(p, u))$, determined by the same variational principle, might be far away from that of $N(p, u, v(p, u))$, thus we need a larger domain for $\text{Re } u$. Consequently, we can not expand the Hamiltonian in Taylor series of (u, v) as in the previous work studying lower dimensional tori (cf. [E], [G]). Instead, we use a different scheme.

Let $G_K \subset \mathbb{R}^{n-1}$ be an open set; its connected components are denoted by $G_K^{(i)}$. We assume that the Hamiltonian

$$H(p, q, u, v) = N(p, u, v) + P(p, q, u, v)$$

is real analytic on each domain

$$\mathbb{D}_{s,t}^{(i)} = \bigcup_{\tilde{p} \in G_K^{(i)}} \mathbb{D}_{s,t}(\tilde{p}),$$

$$\mathbb{D}_{s,t}(\tilde{p}) = \left\{ |\text{Im}q| \leq t, |p - \tilde{p}| \leq s^4, |v - v(\tilde{p}, u)| \leq s^3, |\text{Im}u| \leq s^2, \text{Re}u \in \mathbb{L}(\tilde{p}) + s^2 \right\},$$

where $v(p, u)$ is the solution of the equation $\frac{\partial N}{\partial v}(p, u, v) = 0$. In this section it is enough to show one step of KAM iteration for one connected component of G_K . So we drop the sup-index (i) to simplify the notation.

We assume the following:

(4.1a) $\mathbb{L}(p)$ is either an interval contractible to one point or the whole \mathbb{T} . The interval is defined so that it contains a maximal point of $N(p, u, v(p, u))$ in u and the following holds in $\mathbb{L}(p)$:

$$(4.2) \quad \max_{u \in \mathbb{L}(p)} N(p, u, v(p, u)) - N(p, u, v(p, u)) \leq s^8, \quad \forall u \in \mathbb{L}(p), p \in G_K.$$

It becomes an equality at the boundary of $\mathbb{L}(p)$. Otherwise, we let $\mathbb{L}(p) = \mathbb{T}$.

There might be such a case that $N(p, u, v(p, u))$ reaches its global maximum at several points, e.g. at the point p_0 when $\nabla F(p_0)$ is a non-trivial convex set. In this case, $\mathbb{L}(p)$ may contain several disjoint intervals $\mathbb{L}(p) = \bigcup_i \mathbb{L}_i(p)$ and (4.2) holds for each $\mathbb{L}_i(p)$. For the proof in this section, it is sufficient to assume that in each \mathbb{L}_i there is a local maximum such that (4.2) holds. For simplicity, we assume that $\mathbb{L}(p)$ is an interval.

(4.1b) On each line $p = \text{constant} \in G_K, u \in L(p)$

$$(4.3) \quad \max_{u_1, u_2 \in L(p)} \left| \frac{\partial N}{\partial p}(p, u_1, v(p, u_1)) - \frac{\partial N}{\partial p}(p, u_2, v(p, u_2)) \right| \leq s^2.$$

According to Lemma 3.3 this assumption holds for most p .

(4.1c) The diameter of each connected component of G_K is not bigger than $2s^3$. In each connected component there is a point p^* such that the following holds:

$$(4.4) \quad \frac{\partial N}{\partial v}(p^*, u, 0) = 0, \quad \frac{\partial^2 N}{\partial u \partial v}(p^*, u, 0) = 0, \quad u \in L(p^*).$$

Indeed, under small perturbation $[P](p, u, v)$ the equation

$$\frac{\partial}{\partial v}(N + [P])(p^*, u, v) = 0$$

has solution $v = v(p^*, u)$. By using the coordinate translation $(p, q, u, v) \rightarrow (p, q, u, v + v(p^*, u))$ we obtain (4.4). This transformation is symplectic if we fix p^* as a parameter.

(4.1d) The maximum point \tilde{u} has weak persistency:

$$(4.5) \quad \left| \frac{d^2 N}{du^2}(p, \tilde{u}, v(p, \tilde{u})) \right| \leq \bar{s}^2, \quad \forall p \in G_K.$$

(4.1e) The positivity of N : The smallest eigenvalue of the Hessian matrix of N in (p, v) has positive lower bound $\lambda > 0$ for all u .

(4.1f) The smallness of the perturbation:

$$(4.6) \quad \|P(p, q, u, v)\|_{s,t} \leq \delta,$$

where

$$\delta \leq \min \left\{ \frac{D(t - \bar{t})^{\nu+1}(s^3 - \bar{s}^3)(s^2 - \bar{s}^2)}{8c_2}, \lambda \bar{s}^3(s^3 - \bar{s}^3) \right\}.$$

$\|\cdot\|_{s,t}$ denotes the supnorm on the domain $\mathbb{D}_{s,t}$. Note that since the dependence of N on u comes from the perturbation, it is reasonable to assume that

$$(4.7) \quad \max_{l_1+l_2+l_3 \leq 3, l_2 \geq 1} \left| \frac{\partial^{l_1+l_2+l_3} N}{\partial p^{l_1} \partial u^{l_2} \partial v^{l_3}} \right|, \max \left| \frac{d^3 N}{du^3}(p, u, v(p, u)) \right| \leq \frac{1}{16}.$$

The parameters s, \bar{s}, t, \bar{t} are assumed to satisfy the following conditions:

$$(4.8) \quad 2\bar{s} \leq s, \quad s \leq t \leq \frac{1}{2}, \quad \bar{s} \leq \bar{t}, \quad s^5 \leq \bar{s}^4,$$

$$(4.9) \quad \frac{D}{4s^2} \geq \left(\frac{32}{3(t - \bar{t})} \log \frac{1}{\bar{s}} \right)^{\nu+1}.$$

To construct the coordinate transformation $\mathfrak{T} : (\bar{p}, \bar{q}, \bar{u}, \bar{v}) \rightarrow (p, q, u, v)$, we introduce a generating function W which is the solution of the equation

$$(4.10) \quad \left\langle \frac{\partial N}{\partial p}, \frac{\partial W}{\partial q} \right\rangle = -T_K P + [P].$$

Here $T_K P$ is the truncation of the Fourier expansion of P up to the order K , i.e.

$$T_K P = \sum_{k \in \mathbb{Z}^{n-1}, |k| \leq K} P_k(p, u, v) e^{i\langle k, q \rangle},$$

and P_k is the corresponding Fourier coefficient. As in [A] we choose

$$(4.11) \quad K = \left\lceil \frac{32}{3(t - \bar{t})} \log \frac{1}{\bar{s}} \right\rceil + 1,$$

where $[a]$ denotes the largest integer among those smaller than a . In this case,

$$(4.12) \quad \|P - T_K P\|_{s, t - \frac{t - \bar{t}}{4}} \leq \left(\frac{2\pi}{e} \right)^{n-1} \left(\frac{16}{t - \bar{t}} \right)^n \delta \bar{s}^2.$$

Let \mathfrak{T} be the time-1-map of the Hamiltonian flow Φ_W^t . Then

$$(4.13) \quad H \circ \mathfrak{T} = (N + [P])(p, u, v) + P_+(p, q, u, v),$$

where

$$\begin{aligned} P_+ &= M_1 + M_2 + M_3, \\ M_1 &= \left\langle \frac{\partial N}{\partial p}, \frac{\partial W}{\partial q} \right\rangle + T_K P - [P] + \frac{\partial N}{\partial v} \frac{\partial W}{\partial u} - \frac{\partial N}{\partial u} \frac{\partial W}{\partial v} \\ M_2 &= \int_0^1 \{(1-t)\{N, W\} + P, W\} \circ \Phi_W^t dt, \\ M_3 &= P - T_K P. \end{aligned}$$

To complete one step of iteration we introduce some intermediate domains for those relevant variables. Let $\mathbb{L}_1(p) = [l_1(p), r_1(p)]$, and choose $\tilde{u} - l_1$ and $r_1 - \tilde{u}$ as big as possible under the condition

$$(4.14) \quad \max_u N(p, u, v(p, u)) - N(p, u, v(p, u)) \leq \bar{s}^8, \quad \forall u \in \mathbb{L}_1.$$

Obviously, $\mathbb{L}_1 \subseteq \mathbb{L}$. Let

$$\mathbb{D}_1 = \bigcup_{\tilde{p} \in G_K} \mathbb{D}_1(\tilde{p}), \quad \mathbb{D}_2 = \bigcup_{\tilde{p} \in G_K} \mathbb{D}_2(\tilde{p}),$$

where

$$\begin{aligned} \mathbb{D}_2(\tilde{p}) &= \left\{ |\operatorname{Im} q| \leq \frac{1}{2}(t + \bar{t}), |p - \tilde{p}| \leq \frac{1}{2}(s^4 + \bar{s}^4), |v - v(p, u)| \leq \frac{1}{2}(s^3 + \bar{s}^3), \right. \\ &\quad \left. |\operatorname{Im} u| \leq \frac{1}{2}(s^2 + \bar{s}^2), \operatorname{Re} u \in \mathbb{L}(\tilde{p}) + \frac{1}{2}(s^2 + \bar{s}^2) \right\}, \\ \mathbb{D}_1(\tilde{p}) &= \left\{ |\operatorname{Im} q| \leq \bar{t}, |p - \tilde{p}| \leq \bar{s}^4, |v - v(p, u)| \leq 2\bar{s}^3, |\operatorname{Im} u| \leq \bar{s}^2, \right. \\ &\quad \left. \operatorname{Re} u \in \mathbb{L}_1(\tilde{p}) + \bar{s}^2 \right\}. \end{aligned}$$

Let

$$G'_K = \left\{ p \in \mathbb{R}^{n-1} : \left| \left\langle \frac{\partial N}{\partial p}(p, \tilde{u}, v(p, \tilde{u})), k \right\rangle \right| \geq D|k|^{-\nu}, \forall |k| \leq K \right\}.$$

It shall be shown in section 7 that G_K is constructed so that $G_K \subset G'_K$. With such a condition we can convince ourselves that

$$(4.15) \quad \left| \left\langle \frac{\partial N}{\partial p}(p, u, v), k \right\rangle \right| \geq \frac{1}{2} D|k|^{-\nu}, \quad \forall (p, q, u, v) \in \mathbb{D}_2.$$

In fact, under the assumption of (4.3), the condition (4.9) and the definition (4.11), we have $|\partial_p N(\tilde{p}, u_1, v(\tilde{p}, u_1)) - \partial_p N(\tilde{p}, u_2, v(\tilde{p}, u_2))| \leq 4^{-1} DK^{-\nu-1}$ for all $u_1, u_2 \in \mathbb{L}(\tilde{p})$ and for all $\tilde{p} \in G_K$. Thus the assumptions (4.7) and (4.8) guarantee that the following hold:

$$\left| \partial_p N(p, u, v) - \partial_p N(\tilde{p}, u(\tilde{p}), v(\tilde{p}, u)) \right| \leq \frac{1}{2} DK^{-\nu-1}, \quad \forall (p, q, u, v) \in \mathbb{D}_2(\tilde{p}).$$

This verifies (4.15). If we restrict (p, q, u, v) in a smaller domain $\mathbb{D}_{s,t} \cap \{|\operatorname{Im}| \leq t - \frac{1}{4}(t - \bar{t})\}$, the standard arguments lead to

$$(4.16) \quad \|W(p, q, u, v)\|_{\mathbb{D}_{s,t-\frac{1}{4}(t-\bar{t})}} \leq \theta = \frac{c_2 \delta}{D(t - \bar{t})^{\nu+1}}.$$

Shrinking it further to \mathbb{D}_2 and using Cauchy's estimate we obtain

$$\begin{aligned}
 (4.17) \quad & \|W_p\|_{\mathbb{D}_2} \leq 2\theta(s^4 - \bar{s}^4)^{-1}, \\
 & \|W_q\|_{\mathbb{D}_2} \leq 4\theta(t - \bar{t})^{-1}, \\
 & \|W_u\|_{\mathbb{D}_2} \leq 2\theta(s^2 - \bar{s}^2)^{-1}, \\
 & \|W_v\|_{\mathbb{D}_2} \leq 2\theta(s^3 - \bar{s}^3)^{-1}, \\
 (4.18) \quad & \max_{l_1+l_2+l_3+l_4=2} \left\{ \left\| \frac{\partial^2 W}{\partial p^{l_1} \partial q^{l_2} \partial u^{l_3} \partial v^{l_4}} \right\|_{\mathbb{D}_2} \right\} \leq 8\theta(s^4 - \bar{s}^4)^2.
 \end{aligned}$$

Let $(\bar{p}, \bar{q}, \bar{u}, \bar{v}) = \mathfrak{F}(p, q, u, v)$. As

$$(4.19) \quad (\bar{p}, \bar{q}, \bar{u}, \bar{v}) - (p, q, u, v) = \int_0^1 J\nabla W(\Phi_W^t(p, q, u, v)) dt,$$

we are convinced by the smallness condition (4.8) on δ that \mathfrak{F} does map \mathbb{D}_1 into \mathbb{D}_2 . Furthermore, it is obvious in \mathbb{D}_1 that

$$(4.20) \quad |\mathfrak{F} - id| \leq \frac{2\theta}{s^4 - \bar{s}^4}, \quad |D\mathfrak{F} - I| \leq \frac{8\theta}{(s^4 - \bar{s}^4)^2},$$

where $D\mathfrak{F}$ stands for the Jacobian of \mathfrak{F} .

What remains to do is to figure out the upper bound of the new error term. We state the following lemma first.

Lemma 4.1. *Assume a function $f(x) \in C^2[a, b]$ satisfies the following conditions: $\exists A \geq 0, B \geq 0, \xi \geq 0$ such that $|f(x)| \leq A, |f''(x)| \leq B$, for all $x \in [a, b]$; there exists $x_0 \in [a, b]$ such that $|f'(x_0)| \leq \xi$. Then*

$$|f'(x)| \leq 2\sqrt{AB} + \xi, \quad \forall x \in [a, b].$$

Proof. Suppose $|f'(x)|$ reaches its maximum in $[a, b]$ at $x_1, |f'(x_1)| = M \geq \xi \geq 0$. Without losing generality, we assume $f'(x_1) = M$. By our assumption, there exists a subinterval $[x_1, x_2] \subset [a, b]$ (or $[x_2, x_1] \subset [a, b]$) such that $f'(x_2) = \xi, \xi \leq f'(x) \leq M$ for all $x \in [x_1, x_2]$ (or $[x_2, x_1]$). Since both at x_1 and at x_2 $|f(x)| \leq A$,

$$2A \geq |f(x_1) - f(x_2)| = \left| \int_{x_2}^{x_1} f'(x) dx \right| \geq \frac{1}{2B} |M - \xi|^2,$$

which leads to our conclusion.

Applying this lemma to $N(p, u, v(p, u))$, in view of (4.7) and the fact that $\partial_u N(p, u, v(p, u)) = \frac{d}{du} N(p, u, v(p, u))$, we obtain that

$$\begin{aligned}
 (4.21) \quad & \left| \frac{\partial N}{\partial u}(p, u, v(p, u)) \right| \leq 2\sqrt{\sup \frac{d^2 N}{du^2} \sup N} + \left| \frac{\partial N}{\partial u}(p, u(p), v(p, u)) \right| \\
 & \leq \bar{s}^4, \quad (\forall u \in \mathbb{L}_1, p \in G_K).
 \end{aligned}$$

With this and (4.7) in hand and applying the lemma again to $\frac{\partial N}{\partial u}(p, u, v(p, u))$ we obtain

$$\begin{aligned}
 (4.22) \quad & \left| \frac{d^2 N}{du^2}(p, u, v(p, u)) \right| \leq 2\sqrt{\sup \frac{d^3 N}{du^3} \sup \frac{\partial N}{\partial u}} + \left| \frac{d^2 N}{du^2}(p, u(p), v(p, u)) \right| \\
 & \leq \frac{3}{2} \bar{s}^2, \quad (\forall u \in \mathbb{L}_1, p \in G_K).
 \end{aligned}$$

According to assumption (4.1c), we find that

$$|v(p, u)| \leq c_3|p - p^*|,$$

$$\left| \frac{\partial^2 N}{\partial u \partial v}(p, u, v(p, u)) \right| \leq \frac{1}{8}|p - p^*|.$$

Consequently,

$$\left| \frac{\partial^2 N}{\partial u^2} \right| = \left| \frac{d^2 N}{du^2} + \left(\frac{\partial^2 N}{\partial v^2} \right)^{-1} \left(\frac{\partial^2 N}{\partial u \partial v} \right)^2 \right|$$

$$\leq 2\bar{s}^2, \quad u \in \mathbb{L}_1(p).$$

Therefore, by direct calculation we obtain that

$$(4.23) \quad \left\| \frac{\partial N}{\partial u} \right\|_{\mathbb{D}_1} \leq \left| \frac{\partial N}{\partial u}(p, u, v) - \frac{\partial^2 N}{\partial u \partial v}(p, u, v(p, u))(v - v(p, u)) \right|$$

$$+ \left| \frac{\partial^2 N}{\partial u \partial v}(p, u, v(p, u)) - \frac{\partial^2 N}{\partial u \partial v}(\tilde{p}, u, v(\tilde{p}, u)) \right| |v - v(p, u)|$$

$$\leq 8\bar{s}^4,$$

$$(4.24) \quad \left\| \frac{\partial N}{\partial v} \right\|_{\mathbb{D}_1} \leq \max_{l_1+l_2+l_3=2, l_3 \geq 1} \left| \frac{\partial^2 N}{\partial p^{l_1} \partial u^{l_2} \partial v^{l_3}} \right| \max\{|p - \text{Rep}|, |v - v(p, u)|\}$$

$$\leq c_4 \bar{s}^3 \quad (\text{see (4.4)}).$$

Thus, from (4.8), (4.13), (4.23) and (4.24) we obtain

$$(4.25) \quad \|M_1\|_{\mathbb{D}_1} \leq c_5 \frac{\bar{s}^3 \theta}{s^2 - \bar{s}^2}.$$

Note that the term M_2 is of square order of the error term, and that M_3 is the remainder of the truncation of the Fourier series of P (see (4.12)). We finally obtain

$$(4.26) \quad \|\tilde{P}\|_{\mathbb{D}_1} \leq c_6 \left\{ \frac{\bar{s}^3 \theta}{s^2 - \bar{s}^2} + \frac{s^2 \delta}{(t - \bar{t})^n} + \frac{\theta^2}{(s^4 - \bar{s}^4)^2} \right\} := \bar{\delta}.$$

Under the coordinate transformation the Hamiltonian has the form

$$H \circ \mathfrak{T} = (N + [P])(p, u, v) + \tilde{P}(p, q, u, v).$$

Let $\hat{v}(\tilde{p}, u)$ be the solution of $\partial_v(N + [P])(\tilde{p}, u, \hat{v}(\tilde{p}, u)) = 0$. In view of assumption (4.4) and the smallness of P , we have $|\hat{v}(\tilde{p}, u) - v(\tilde{p}, u)| < \bar{s}^3$. By introducing the coordinate transformation $(p, q, u, v) \rightarrow (p, q, u, v + \hat{v}(u))$ we can set (4.4) to hold for $N + [P]$. Let $\bar{\mathbb{L}} \subseteq \mathbb{T}$ be an interval as big as possible under the condition

$$\max_u (N + [P])(\tilde{p}, u, \hat{v}) - (N + [P])(\tilde{p}, u, \tilde{v}) \leq \hat{s}^8, \quad \forall u \in \bar{\mathbb{L}},$$

where \hat{s}^8 plays the role in the next step of KAM iteration as \bar{s} does in this step. Obviously, $\bar{\mathbb{L}} \subseteq \mathbb{L}_1$. Define

$$\mathbb{D}_{\bar{s}, \bar{t}} = \bigcup_{\tilde{p} \in G_{\bar{K}}} \mathbb{D}_{\bar{s}, \bar{t}}(\tilde{p}),$$

where

$$\mathbb{D}_{\bar{s}, \bar{t}}(\tilde{p}) = \left\{ |\text{Im } q| \leq \bar{t}, |p - \tilde{p}| \leq \bar{s}^4, |v - \bar{v}(p, u)| \leq \bar{s}^3, |\text{Im } u| \leq \bar{s}^2, \text{Re } u \in \bar{\mathbb{L}} + \bar{s}^2 \right\}.$$

$G_{\bar{K}}$ is constructed by cutting off some small measure set from G_K so that the diameter of each connected component is not bigger than $2\bar{s}^2$. We postpone the construction to section 7. The above arguments imply $\mathbb{D}_{\bar{s}, \bar{\epsilon}} \subseteq \mathbb{D}_1$ on which $H \circ \mathfrak{T}$ is real analytical and (4.26) holds.

For the next step of iteration we need to check the assumptions from (4.1a) to (4.1f). Clearly all of them are satisfied except for (4.1d) and (4.1f). When (4.1d) does not hold, it means the critical point has strong persistency, so we switch the iteration process to another type, which is shown below. We postpone the verification of (4.1f), the smallness for \bar{P} in $\mathbb{D}_{\bar{s}, \bar{\epsilon}}$. It is closely related to the convergence proof, which we shall address in the next section. So far, one step of iteration for the case of weak persistency is completed.

4.2. Strong persistency. By definition, the following holds at \tilde{u} , the maximum point of $N(p, u, v(p, u))$:

$$(4.27) \quad \left| \frac{d^2 N}{du^2}(p, \tilde{u}, v(p, \tilde{u})) \right| \geq \bar{s}^2.$$

Under such a condition the critical point of $(N + [P])(p, u, v)$, determined by the same variational principle, remains close to that of $N(p, u, v)$. In this case, the coordinate transformation is constructed in the standard way.

Assume that the last step of iteration is done under the condition of weak persistency, so the Hamiltonian is analytical in the domain $\{|u - \tilde{u}| \leq s^2, |v| \leq s^3\}$. The main part takes the form

$$N(p, u, v) = N(p, \tilde{u}, 0) + \frac{1}{2}N_{vv}(\tilde{p}, \tilde{u}, 0)v^2 - \frac{1}{2}N_{uu}(\tilde{p}, \tilde{u}, 0)u^2 + O\left((p - \tilde{p})(u, v), (u, v)^3\right).$$

To set N in a normal form we introduce a linear transformation $\mathbf{T} : (u, v) \rightarrow (x, y)$ for rescaling

$$(4.28) \quad \begin{pmatrix} u - \tilde{u} \\ v \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{\alpha}^{-1} & \sqrt{\alpha}^{-1} \\ \sqrt{\alpha} & -\sqrt{\alpha} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

where $\alpha = \sqrt{-\frac{N_{uu}}{N_{vv}}}$ is a positive number. In the new variable (x, y) the main part has the form

$$N(p, x, y) = N(p, 0, 0) + \alpha xy + O\left((p - \tilde{p})(x, y), (x, y)^3\right).$$

The facts that $N_{vv} = O(1)$ and $\bar{s}^2 \leq -N_{uu} \ll 1$ convince us that $\{|x| \leq s^3, |y| \leq s^3\} \subset \{|u - \tilde{u}| \leq s^2, |v| \leq s^3\}$, on which H is analytical. Thus we can restrict (x, y) in a smaller domain $\{|x| \leq s^3, |y| \leq s^3\}$ and expand both P and the generating function W into Taylor series in (x, y)

$$(4.29) \quad \begin{aligned} P(p, q, x, y) &= \sum P_{ij}(p, q)x^i y^j, \\ W(p, q, x, y) &= \sum_{|i+j| \leq 5} W_{ij}(p, q)x^i y^j. \end{aligned}$$

The generating function W is constructed so that the main part of P can be cancelled by the main part of $\{N, W\}$. Towards that goal we set W be the solution of the following equation:

$$(4.30) \quad \langle \omega, W_q \rangle + (x, y)\text{diag}(\alpha, -\alpha) \begin{pmatrix} W_x \\ W_y \end{pmatrix} = R,$$

where R is the truncation of P 's Taylor series, up to order $|i + j| = 5$ in (x, y) , not including the average part of P in q , i.e. P_{0l} . Equation (4.30) is solved by Fourier expansion. If we write both W and R in the form of (4.29), then

$$(4.31) \quad W_{kl} = \frac{R_{kl}}{i\langle k, \omega \rangle + \langle (l_1 - l_2), \alpha \rangle}.$$

Since α is a real number and ω is a diophantine vector, the generating function W is well defined. Consequently, the coordinate transformation \mathfrak{Z} is also well defined.

Now let us show the estimates needed for the convergence proof. We assume that

1. The Hamiltonian $H = N(y, x, y) + P(p, q, x, y)$ is real analytic on the complex domain $\mathbb{D}_{s,t}$

$$\mathbb{D}_{s,t} = \{|\operatorname{Im}q| \leq t, |p - \tilde{p}| \leq s^4, |x|, |y| \leq s^3\}.$$

2. The perturbation P satisfies the smallness condition

$$\|P\|_{\mathbb{D}_{s,t}} \leq \delta,$$

$$(4.32) \quad \delta \leq \min \left\{ \frac{3D}{2c_7}(t - \bar{t})^{\nu+1}(s^3 - \bar{s}^3)\bar{s}^3, \frac{1}{2\sqrt{\lambda}}s^4\bar{s}^3 \right\},$$

where \bar{t} and \bar{s} are small positive numbers satisfying (4.33) below.

To complete one step of iteration, we need several intermediate domains. We choose small positive numbers σ and τ such that

$$(4.33) \quad 0 < 2\bar{s} < s \leq \frac{t - \bar{t}}{5}, \quad 0 < t < 1.$$

We define

$$\begin{aligned} \mathbb{D}_2 &= \{|\operatorname{Im}q| \leq \frac{1}{2}(t + \bar{t}), |p - \tilde{p}| \leq \frac{1}{2}(s^4 + \bar{s}^4), |x|, |y| \leq \frac{1}{2}(s^3 + \bar{s}^3)\}, \\ \mathbb{D}_1 &= \{|\operatorname{Im}q| \leq \bar{t}, |p - \tilde{p}| \leq \frac{3}{2}\bar{s}^4, |x|, |y| \leq \frac{3}{2}\bar{s}^3\}. \end{aligned}$$

Note that the function R in (4.30) is the truncation of the Taylor series of P with respect to (p, x, y) . It is easy to check that

$$(4.34) \quad \begin{aligned} \|R\|_{\mathbb{D}_{s,t}} &\leq 2\|P\|_{\mathbb{D}_{s,t}}, \\ \|P - R\|_{\mathbb{D}_1} &\leq \frac{3^5\bar{s}^{15}}{(s^3 - \bar{s}^3)^5}\|P\|_{\mathbb{D}_{s,t}}. \end{aligned}$$

By the construction of the generating function W and Cauchy's estimate on the derivative of analytic functions we get

$$(4.35) \quad \|W\|_{\mathbb{D}_2} \leq \theta = \frac{c_7\|P\|_{\mathbb{D}_{s,t}}}{D(t - \bar{t})^{\nu+1}},$$

$$(4.36) \quad \begin{aligned} \|W_p\|_{\mathbb{D}_2} &\leq 2\theta(s^4 - \bar{s}^4)^{-1}, \\ \|W_q\|_{\mathbb{D}_2} &\leq 4\theta(t - \bar{t})^{-1}, \\ \|W_x\|_{\mathbb{D}_2} &\leq 2\theta(s^3 - \bar{s}^3)^{-1}, \\ \|W_y\|_{\mathbb{D}_2} &\leq 2\theta(s^3 - \bar{s}^3)^{-1}, \end{aligned}$$

$$(4.37) \quad \|D^2W\|_{\mathbb{D}_2} \leq \frac{10c_7\|P\|_{\mathbb{D}_{s,t}}}{D(t - \bar{t})^{\nu+1}(s^4 - \bar{s}^4)^2}.$$

The coordinate transformation \mathfrak{T} is defined as the time-1-map of the Hamiltonian flow Φ_W^t ; it follows that

$$\mathfrak{T}(p, q, x, y) - (p, q, x, y) = \int_0^1 J\nabla W(\Phi_W^t(p, q, x, y))dt.$$

Thus, the conditions (4.32), (4.33) and the upper bound for W in (4.35) and (4.36) guarantee that \mathfrak{T} maps \mathbb{D}_1 into \mathbb{D}_2 .

Because the generating function W solves (4.30), we find that

$$(4.38) \quad H \circ \mathfrak{T} = (N + [P])(p, x, y) + P_+(p, q, x, y),$$

where

$$P_+ = M_1 + M_2 + M_3,$$

$$\begin{aligned} M_1 &= \{N, W\} - \langle \omega, W_q \rangle + (x, y)\text{diag}(\alpha, -\alpha) \begin{pmatrix} W_x \\ W_y \end{pmatrix} + R \\ &= O\{x^2, y^2\}\{W_x, W_y\} + \|p - \tilde{p}\|O(W_x, W_y, W_q) \\ &\quad (\text{see (4.30)}), \end{aligned}$$

$$M_2 = \int_0^1 \{(1-t)\{N, W\} + P, W\} \circ \Phi_W^t dt,$$

$$M_3 = P - R.$$

It follows that, by virtue of the estimates (4.34), (4.35) and (4.36),

$$(4.39) \quad \|P_+\|_{\mathbb{D}_1} \leq c_8 \left\{ \frac{\bar{s}^4 \theta}{(s^3 - \bar{s}^3)} + \frac{\bar{s}^{15} \delta}{(s^3 - \bar{s}^3)^5} + \frac{\theta^2}{(s^4 - \bar{s}^4)^2} \right\} = \bar{\delta}.$$

The estimates (4.36) and (4.37) provide an upper bound for the transformation \mathfrak{T} :

$$(4.40) \quad \begin{aligned} \|\mathfrak{T} - Id\|_{\mathbb{D}_1} &\leq \frac{2c_7 \delta}{D(t - \bar{t})^{\nu+1}(s^4 - \bar{s}^4)}, \\ \|D\mathfrak{T} - I\|_{\mathbb{D}_1} &\leq \frac{8c_7 \delta}{D(t - \bar{t})^{\nu+1}(s^4 - \bar{s}^4)^2}. \end{aligned}$$

To complete this step of iteration we need to show that once the critical point begins to have strong persistency, the critical point of the follow-up iteration steps shall possess strong persistency as well. Indeed, in this case, the singular point of the vector field $J\nabla_{(x,y)}(N + [P])(p, x, y)$, $(\bar{x}(p), \bar{y}(p))$ keeps very close to the origin, $\|(\bar{x}(p), \bar{y}(p))\| \leq \frac{\sqrt{\lambda}\delta}{s^3\bar{s}}$. Consequently, the eigenvalues of the matrix $JD_{(x,y)}^2(N + [P])$, $(\bar{\alpha}, -\bar{\alpha})$ keep close to $(\alpha, -\alpha)$:

$$(4.41) \quad |\bar{\alpha} - \alpha| \leq c_9 s^8.$$

It implies that the new critical point has strong persistency also. As N is strictly convex in y and the critical point \tilde{u} is non-degenerate, there exists \hat{p} with $|\tilde{p} - \hat{p}| \leq \frac{1}{2}\bar{s}^4$ such that

$$(4.42) \quad \partial_p N(\tilde{p}, 0, 0) = \partial_p(N + [P])(\hat{p}, \bar{x}(\hat{p}), \bar{y}(\hat{p})).$$

We define

$$\mathbb{D}_{\bar{s}, \bar{t}} = \{|\text{Im}q| \leq t, |p - \hat{p}| \leq \bar{s}^4, |x - \bar{x}(\hat{p})|, |y - \bar{y}(\hat{p})| \leq s^3\}.$$

By the smallness (4.32), it is contained in \mathbb{D}_1 .

5. THE CONVERGENCE

Let us consider the Hamiltonian $H_0(p, q, u, v) = N_0(p, v) + P_0(p, q, u, v)$ which is real analytical in

$$\mathbb{D}_0 = \{|\operatorname{Im} q| \leq t_0, |p - \tilde{p}| \leq s_0^4, \tilde{p} \in \mathbb{D}, |\operatorname{Im} u| \leq s_0^2, \operatorname{Re} u \in L_0 = \mathbb{T}, |v - \tilde{v}| \leq s_0^3\},$$

in which $s_0 = s$ and $t_0 = t$.

To simplify the notation, we always use (p, q, u, v) to denote the symplectic coordinates at each step of iteration, and put the subscript m to H, N and P to specify the Hamiltonian functions at m -th step of iteration.

After the m iteration steps, the Hamiltonian H of (1.1) has the form

$$(5.1) \quad H_m(p, q, u, v) = N_m(p, u, v) + P_m(p, q, u, v).$$

It is analytical in the domain \mathbb{D}_m which is defined in two ways according to whether the maximum point \tilde{u} of the function $N(p, u, v(p, u))$ has weak persistency or not. When it has weak persistency we choose

$$\mathbb{D}_m = \bigcup_{\tilde{p} \in G_{\kappa_m}} \mathbb{D}_m(\tilde{p}),$$

where

$$\begin{aligned} \mathbb{D}_m(\tilde{p}) = \{ & |\operatorname{Im} q| \leq t_m, |p - \tilde{p}| \leq s_m^4, |v - v_m(\tilde{p}, u)| \leq s_m^3, \\ & |\operatorname{Im} u| \leq s_m^2, \operatorname{Re} u \in \mathbb{L}_m(\tilde{p}) + s_m^2 \}. \end{aligned}$$

$v_m(p, u)$ is the solution of the equation $\partial_v N_m(p, u, v) = 0$. In the case of strong persistency we choose

$$\mathbb{D}_m = \{|\operatorname{Im} q| \leq \tau_m, |p - \tilde{p}| \leq s_m^4, |x| \leq s_m^3, |y| \leq s_m^3\}.$$

On these domains we have

$$(5.2) \quad \|P_m(p, q, u, v)\|_m \leq \delta_m = s_m^{13},$$

where $\|\cdot\|_m$ denotes the sup-norm on the domain \mathbb{D}_m . We say that the critical point has weak persistency when $|\frac{d^2 N_m}{du^2}(p, u, v_m(p, u))| \leq s_{m+1}^2$, otherwise it is treated as the strong persistency case. In the case of weak persistency, \mathbb{L}_m is chosen as long as possible under the condition that $|\max_{u \in \mathbb{L}_m} N_m(p, u, v_m(p, u)) - N_m(p, u, v_m(p, u))|_{u \in \mathbb{L}_m} \leq s_m^8$ is satisfied. In the first few iteration steps we have to deal with the case of weak persistency. For instance, in the first step, the main term N_0 is independent of u .

We set

$$t_m = \frac{t_0}{2}(1 + 2^{-m}), \quad \delta_m = s_m^{13}, \quad \delta_m = \delta_{m-1}^{\frac{14}{13}}.$$

In order to use the arguments in the last section we use s_m, t_m and δ_m to replace s, t and δ , use s_{m+1}, t_{m+1} and δ_{m+1} to replace \bar{s}, \bar{t} and $\bar{\delta}$ respectively at the $(m+1)$ -th step of KAM iteration. We assume that the diophantine condition (1.4) is satisfied for $|k| \leq K_m$ in the case of weak persistency. The verification of (1.4) is the main task of section 7. In the weak persistency case we also assume the small fluctuation of frequency along L_m .

$$\left| \frac{\partial N_m}{\partial p}(p, u_1, v(p, u_1)) - \frac{\partial N_m}{\partial p}(p, u_2, v(p, u_2)) \right| \leq s_m^2, \quad \forall u_1, u_2 \in L_m.$$

By the work of the last section, we know that a generating function $W_m(p, q)$ is constructed by solving the relevant homological equation. The $(m+1)$ -th step of

symplectic coordinate transformation \mathfrak{T}_{m+1} is defined as the time-1-map of the Hamiltonian flow

$$\mathfrak{T}_{m+1} = \Phi_{W_m}^t \Big|_{t=1}.$$

It transforms the Hamiltonian H_m into $H_{m+1} = N_{m+1}(p, u, v) + P_{m+1}(p, q, u, v)$ which is real analytical in the domain in \mathbb{D}_{m+1} . The main point we need to show is that $\exists d_1 = d_1(n, D, \nu, t_0) > 0$, such that if $\delta_0 \leq d_1$, then $\|P_m\| \leq \delta_m \forall m \in \mathbb{N}$. We set $d_2 = \min\{2^{-13}, t_0/20\}$. If $\delta_0 \leq d_2$, then $s_{m+1}/s_m \leq \frac{1}{2}$, $s_m \leq t_m \leq 1$, $s_m^5 \leq \bar{s}_{m+1}^4$, and $s_m \leq (t_m - t_{m+1})/5$; these verify (4.8) and (4.33). Thus, in the case of weak persistency, we obtain from (4.26)

$$\|P_{m+1}\| \leq c_6 \left\{ \frac{4c_2 2^{(\nu+1)(m+2)}}{3Dt_0^{\nu+1}} \left(\frac{\delta_{m+1}}{\delta_m} \right)^{\frac{3}{13}} + \frac{2^{n(m+2)} \delta_m^{\frac{1}{13}}}{t_0^n} + \left(\frac{16}{15} \right)^2 \delta_m^{\frac{4}{13}} \right\} \delta_{m+1},$$

while in the case of strong persistency, we obtain from (4.39)

$$\|P_{m+1}\| \leq c_8 \left\{ \frac{8c_7 2^{(\nu+1)(m+2)}}{7Dt_0^{\nu+1}} \left(\frac{\delta_{m+1}}{\delta_m} \right)^{\frac{4}{13}} + \left(\frac{8}{7} \right)^5 \delta_m^{\frac{2}{169}} + \left(\frac{16}{15} \right)^2 \delta_m^{\frac{4}{13}} \right\} \delta_{m+1}.$$

The fact that δ_m decreases to zero faster than exponentially implies that there exists such a $d_1 = d_1(n, D, \nu, t_0) \leq d_2$. In fact, if we choose

$$(5.3) \quad d_1 = \min \left\{ 2^{-13^3 \frac{\nu+1}{3}}, \left(\frac{t_0^{2\nu+2}}{4 \max\{c_2 c_6, c_7 c_8\} 2^{2\nu+2}} \right)^{169} \right\}, \quad D = d_1^{\frac{2}{169}},$$

then

$$\|P_{m+1}\| \leq \delta_{m+1}.$$

When δ_0 is set smaller than d_1 , the prerequisite conditions (4.7) and (4.9) are also satisfied.

By such an iteration procedure we obtain a series of coordinate transformations \mathfrak{T}_m . It follows from (4.20) and (4.40) that

$$(5.4) \quad \begin{aligned} \|\mathfrak{T}_m - id\| &\leq s_m^8, \\ \|D\mathfrak{T}_m - I\| &\leq s_m^4, \end{aligned}$$

where $D\mathfrak{T}_m$ is the Jacobian of \mathfrak{T}_m . Let $\mathfrak{S}_m = \mathfrak{T}_1 \circ \mathfrak{T}_2 \circ \dots \circ \mathfrak{T}_m$. Under the transformation \mathfrak{S}_m , the Hamiltonian approaches closer and closer to an integrable one as $m \rightarrow \infty$,

$$H \circ \mathfrak{S}_m = N_m(p, u, v) + P_m(p, q, u, v), \quad \|P\|_{\mathbb{D}_m} \leq \delta_m,$$

which is real analytical in the domain \mathbb{D}_m . If the critical point always has weak persistency as $m \rightarrow \infty$, the domain \mathbb{D}_m shrinks to a set containing the sub-manifold

$$\mathbb{D}_\infty \supseteq \left\{ \text{Im}q \leq \frac{\tau_0}{2} \right\} \times \{p = \tilde{p}, u \in \mathbb{L}_\infty, v = 0\},$$

where $\mathbb{L}_\infty = \lim_{m \rightarrow \infty} \mathbb{L}_m$ is either a point or an interval. The transformation \mathfrak{S}_m maps \mathbb{D}_m into \mathbb{D}_0 and converges C^1 uniformly to an embedding $\mathfrak{S}_\infty: \mathbb{D}_\infty \rightarrow \mathbb{D}_0$. The set $\text{Re}\mathbb{D}_\infty$ contains trivially embedded $(n - 1)$ -dimensional tori $p = \tilde{p}, u \in \mathbb{L}_\infty, v = 0$. By the standard arguments (cf. [A]), we see that \mathfrak{S}_∞ maps this torus (these tori) (depending on whether \mathbb{L}_∞ is a point or an interval) to an $(n - 1)$ manifold (manifolds) in $\text{Re}\mathbb{D}_0$, with its tangential map taking constant vector field on $p = \tilde{p}, u \in \mathbb{L}_\infty, v = 0, \dot{q} = \omega, \dot{p} = \dot{u} = \dot{v} = 0$, to the vector field governed by $J\nabla H$. Therefore, this manifold (these manifolds) is (are) invariant to the Hamiltonian flow Φ_H^t . If the critical point obtains the stronger persistency at some step of iteration

once, it shall have strong persistency in the following steps as well. In this case, the domain \mathbb{D}_m shrinks to a set

$$\mathbb{D}_\infty = \left\{ \text{Im}q \leq \frac{\tau_0}{2} \right\} \times \{p = \tilde{p}, u = \tilde{u}, v = 0\};$$

it contains an $(n - 1)$ -torus. The transformation \mathfrak{S}_m with its differential converges uniformly on this set, which implies the existence of the invariant torus.

6. THE MINIMALITY

In this section we show that these $(n - 1)$ -dimensional tori are the support of the corresponding minimal measures. We study this problem in the real part of the domain \mathbb{D}_m . For the simplicity of notation, let us still use \mathbb{D}_m to denote it. We first consider the case when the critical point at each step always has weak persistency; the proof for other cases can be done in a similar way. We assume $(p_0, u_0, 0) \times \mathbb{T}^{n-1} \in \bigcap_{j=0}^\infty \mathbb{D}_j$.

Lemma 6.1. *There exists a symplectic transformation of $\tilde{\mathfrak{S}}_\infty: \mathbb{D}_0 \rightarrow \mathbb{D}_0$, such that under this transformation the new Hamiltonian function $H_\infty(p, q, u, v) = H \circ \tilde{\mathfrak{S}}_\infty(p, q, u, v)$ has the following properties:*

- (1) *the $(n - 1)$ -dimensional torus $\Gamma = \{p = p_0, u = u_0, v = 0\}$ is invariant to the Hamiltonian flow $\Phi_{H_\infty}^t$;*
- (2) *the function $h = H_\infty(p_0, q, u, 0) = H \circ \tilde{\mathfrak{S}}_\infty(p_0, q, u, 0)$ attains its maximum on Γ .*

Proof. Under the m iteration steps, the Hamiltonian H of (1.1) has the form

$$(6.1) \quad H_m(p, q, u, v) = H \circ \mathfrak{S}_m = N_m(p, u, v) + P_m(p, q, u, v).$$

By the construction of the KAM iteration, we have a series of nested domains $\mathbb{D}_0 \supset \mathbb{D}_1 \supset \dots \supset \mathbb{D}_m \dots$, and (6.1) holds in each \mathbb{D}_m . Note that the Hamiltonian flow $\Phi_{N_m}^t$ is integrable; it has an invariant $(n - 1)$ -dimensional torus Γ_m which corresponds to the maximum point of $N_m(p, u, v)$ as the function of u . Unless N_m is constant in u , there exist an unstable manifold \mathcal{W}_m^u and a stable manifold \mathcal{W}_m^s which intersect each other at Γ_m even if the critical point is degenerate. These two Lagrangian sub-manifolds are horizontal, consequently, are the graph of some closed one form η_m^u and η_m^s , respectively. Obviously, we have $\eta_m^u = \eta_m^s|_{\Gamma_m}$. If we write

$$\mathcal{W}^{u,s} = \{(p, q, u, v) : p = p_0, v = V^{u,s}(p_0, u)\},$$

by observing the convexity of N_m in v and the fact $\frac{\partial N_m}{\partial v}(p_0, u, v_m(p_0, u)) \equiv 0$ we find

$$(6.2) \quad |V^{u,s}(p_0, u) - v(p_0, u)| \geq \sqrt{\frac{2}{\Lambda} (N_m(p_0, u_0, v(p_0, u_0)) - N_m(p_0, u, v(p_0, u)))}.$$

Here Λ is the biggest eigenvalue of the Hessian matrix of N_m in (p, v) , and u_0 is the maximum point. However, the coordinate transformation \mathfrak{S}_m , and consequently (6.1) and (6.2), are valid only in \mathbb{D}_m .

Recall the construction of the generating function W_m at each step of iteration. It is easy to see that we can extend W_m with C^2 -differentiability to the whole \mathbb{D}_0

such that $\|W_m\|_{C^2} \leq s_m^4$. Indeed, let

$$\mathbb{D}_m^{(2)} = \left\{ |p - p_0| \leq \frac{1}{2}(s_m^4 + s_{m+1}^4), |v - v_m(p_0, u)| \leq \frac{1}{2}(s_m^3 + s_{m+1}^3), \right. \\ \left. u \in L_m + \frac{1}{2}(s_m^2 - s_{m+1}^2) \right\},$$

$a_m = \frac{1}{2}(s_m^4 - s_{m+1}^4)$ and $\rho(z) = \text{dist}(z, \mathbb{D}_m^{(2)})$ in which $z = (p, u, v)$. We define the extension of the generating function

$$(6.3) \quad \widetilde{W}_m = \left(1 - \gamma \left(\frac{\rho(z)}{a_m} \right) \right) W_m(p, q, u, v),$$

where the function $\gamma(x) : \mathbb{R} \rightarrow \mathbb{R}^+$ is a C^∞ function defined as follows:

$$\gamma(x) = \frac{\int_{-\infty}^x \alpha(t) dt}{\int_{-\infty}^{\infty} \alpha(t) dt}, \\ \alpha(t) = \begin{cases} e^{-\frac{1}{t(1-t)}} & 0 < t < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Obviously, $\widetilde{W}_m \in C^\infty$, $\widetilde{W}_m = W_m(p, q, u, v)$ in $\mathbb{D}_m^{(2)}$ and $\widetilde{W}_m \equiv 0$ outside of $\mathbb{D}_m^{(2)} + a_m \subset \mathbb{D}_m - a_m$. By the construction of the generating function W we have

$$\|D^j \widetilde{W}_m\| \leq \frac{c_2 2^{(m+2)(\nu+1)+1}}{Dt_0^{\nu+1}} \delta_m^{1-\frac{4j}{13}}, \quad |j| \leq 2.$$

Thus, the map \mathfrak{I}_m can be extended to \mathbb{D}_0 , denoted by $\widetilde{\mathfrak{I}}_m$, with

$$\|\widetilde{\mathfrak{I}}_m - Id\|_{\mathbb{D}_0} \leq \frac{c_2 2^{(m+2)(\nu+1)+1}}{Dt_0^{\nu+1}} s_m^9 := \epsilon_m, \\ \|D\widetilde{\mathfrak{I}}_m - I\|_{\mathbb{D}_0} \leq s_m^4.$$

Let $\widetilde{\mathfrak{S}}_m = \widetilde{\mathfrak{I}}_1 \circ \widetilde{\mathfrak{I}}_2 \circ \dots \circ \widetilde{\mathfrak{I}}_m$. Together with its differential, it converges uniformly to a diffeomorphism $\widetilde{\mathfrak{S}}_\infty = \lim_{m \rightarrow \infty} \widetilde{\mathfrak{S}}_m$ over the whole \mathbb{D}_0 . Indeed, we have

$$(6.4) \quad \|\widetilde{\mathfrak{S}}_\infty - \widetilde{\mathfrak{S}}_m\|_{\mathbb{D}_0} \leq \sum_{k=m}^{\infty} \epsilon_k, \\ \|\widetilde{\mathfrak{S}}_\infty - \widetilde{\mathfrak{S}}_m\|_{\mathbb{D}_0 \setminus \mathbb{D}_m} \leq \epsilon_m,$$

since $\widetilde{W}_{m+1} \equiv 0$ when z is outside of \mathbb{D}_m . Under the coordinate transformation $\widetilde{\mathfrak{S}}_m$, the Hamiltonian $\widetilde{H}_m = H \circ \widetilde{\mathfrak{S}}_m$ does not have the form of (6.1) on the whole \mathbb{D}_0 ; this form is valid only in \mathbb{D}_m . Obviously, H_∞ is independent of q when $p = p_0$, $u = u_0 \in \bigcap_{k=1}^\infty (\mathbb{L}_k + s_k^2)$, $v = 0$.

To show the maximum property, let us consider two points $z = (p_0, q, u_0, 0) \in \Gamma$ and $z^* = (p_0, q, u^*, 0) \notin \Gamma$ and compare the values H_∞ takes at these two points. As $z^* \notin \Gamma$, $\exists m < \infty$ such that $u^* \in \mathbb{L}_m + s_m^2 \setminus (\mathbb{L}_{m+1} + s_{m+1}^2)$. Let $\widetilde{\mathfrak{S}}_\infty^m = \lim_{k \rightarrow \infty} \widetilde{\mathfrak{I}}_m \circ \widetilde{\mathfrak{I}}_{m+1} \circ \dots \circ \widetilde{\mathfrak{I}}_{m+k}$. In view of (6.4) we find that their images $z_m = \widetilde{\mathfrak{S}}_\infty^m(z)$ and $z_m^* = \widetilde{\mathfrak{S}}_\infty^m(z^*)$, under the map $\widetilde{\mathfrak{S}}_\infty^m$, remain close to themselves respectively,

$$(6.5) \quad \|z_m - z\| \leq \epsilon_m, \quad \|z_m^* - z^*\| \leq \epsilon_m.$$

Therefore,

$$\begin{aligned}
 H_\infty(z) - H_\infty(z^*) &= H_m(z_m) - H_m(z_m^*) \\
 &= N_m(z) - N_m(z^*) + P_m(z_m) - P_m(z_m^*) \\
 &\quad - (N_m(z_m^*) - N_m(z^*)) + (N_m(z_m) - N_m(z)) \\
 (6.6) \qquad &\geq s_{m+1}^8 - 2 \max \left\| \frac{\partial N_m}{\partial z} \right\| \epsilon_m - 2\delta_m \\
 &\geq \frac{1}{2} s_{m+1}^8.
 \end{aligned}$$

This completes the proof.

Let $L_\infty(\dot{q}, \dot{u}, q, u) = \langle p, \dot{q} \rangle + \dot{u}v - H_\infty(p, q, u, v)$ be the Lagrangian corresponding to the Hamiltonian H_∞ through the Legendre transformation \mathfrak{L} . Let $L_\lambda = L_\infty - \eta_\lambda$ where $\eta_\lambda = p_0 dq$ is a closed 1-form. Since L_λ^λ is convex in (\dot{q}, \dot{u}) it attains its minimum in (\dot{q}, \dot{u}) at the point where

$$\frac{\partial L_\lambda}{\partial \dot{q}} = 0, \qquad \frac{\partial L_\lambda}{\partial \dot{u}} = 0.$$

Its solution is $(p, v)(\dot{q}, \dot{u}, q, u) = (p_0, 0)$, where the Lagrangian

$$L_\lambda(\dot{q}, \dot{u}, q, u)|_{(p,v)=(p_0,0)} = -H_\infty(p_0, q, u, 0).$$

By virtue of Lemma 6.1, it implies that the Lebesgue measure μ_Γ distributed on $\mathfrak{L}(\Gamma)$ minimizes the action $A_\lambda(\mu) = \int (L_\infty - \lambda) d\mu$. Therefore, the measure $\tilde{\mathfrak{S}}_{\infty*} \mu_\Gamma$ minimizes the action $A_\lambda(\mu) = \int (L - \lambda) d\mu$, because the time-1-map of Hamiltonian flow is an exact symplectic diffeomorphism and $\tilde{\mathfrak{S}}_\infty$ is the limit of the composition of these exact symplectic diffeomorphisms.

7. THE MEASURE ESTIMATE

In the previous sections the existence of some $(n - 1)$ -dimensional tori is proved under the assumption that $\mathfrak{F}_k(p)$ satisfies the diophantine condition for all steps of the iteration. In this section we shall show that this assumption is true for the majority of the action $p \in \mathbb{R}^{n-1}$ in the sense of Lebesgue measure.

First, we assume that the maximum point of $N_k(p, u, v(p, u))$, as the function of u , has weak persistency at each point p .

After m steps of KAM iteration the main part of the Hamiltonian has the form

$$N_m(p, u, v) = N_0 \left(p, v + \sum_{i=1}^{m-1} v_i(p_k^i, u) \right) + \sum_{i=0}^{m-1} [P_i] \left(p, u, v + \sum_{l=i}^{m-1} v_l(p_k^l, u) \right),$$

where $v_l(p_k^l, u)$ is the solution of the equation $\frac{\partial N_l}{\partial v}(p_k^l, u, v) = 0$. To run the $(m + 1)$ -th step of KAM iteration we need to cut off a small measure subset from G_{m-1} to get $G_m \subset G_{m-1}$, so that the hypotheses (4.1b) and (4.1c) hold for N_m when $Rep \in G_m$.

Before the construction of G_m , we should be aware of the following fact. Let $G_m^{(i)}$ and $G_m^{(j)}$ be two different connected components of G_m . The two sets $\bigcup_{p \in G_m^{(i)}} \mathbb{D}_m(p)$ and $\bigcup_{p \in G_m^{(j)}} \mathbb{D}_m(p)$ may have non-empty overlap, so the Hamiltonian may be multi-valued on the overlap, due to different translations $v \rightarrow v + v(p_\sigma, u)$, $\sigma = i, j$. But it does not make any trouble for us. For the measure estimate, we are only interested

in the value $\partial_p N_m$ takes on the set G_m , and we do not care about how to extend N_m , as the function of p , to the whole space \mathbb{R}^{n-1} .

There are several steps to construct G_m and to get its measure estimate. As the first step, we assume the distance between any two connected components of G_i ($i < m$) is not smaller than s_i^4 .

Let $\{p_k^i\}$ be a point in the k -th connected component of G_i . We introduce the coordinate transformation $\mathfrak{P}_i(p_k^i): (p, q, u, v) \rightarrow (p, q, u, v + v(p_k^i, u))$, and (4.1c) is then satisfied at the i -th connected component of G_i . We introduce a coordinate transformation $\mathfrak{U}_m = \mathfrak{P}_1^{-1} \circ \dots \circ \mathfrak{P}_{m-1}^{-1} \circ \mathfrak{P}_m^{-1}$. Under this map the main term of H_m has the form

$$N_m \circ \mathfrak{U}_m = N_0(p, v) + \sum_{i=0}^{m-1} [P_i] \left(p, u, v - \sum_{l=1}^i v_l(p_k^l, u) \right),$$

where $[P_i]$ is real analytical on $\bigcap_{l=1}^i G_l + s_i^4$ and $|[P_i]| \leq s_i^{13}$. As the distance among different connected components of G_l is not smaller than s_l^4 , we are convinced that $[P_i]$ can be C^2 -smoothly extended to the whole space such that $\|P_i\|_{C^2} \leq c_{10} s_i^5$. For simplicity, we still use $N_m \circ \mathfrak{U}_m$ to denote the function extended to the whole space.

Let $\bar{v}_m(p, u)$ be the solution of $\frac{\partial}{\partial v} N \circ \mathfrak{U}_m(p, u, v) = 0$. Obviously, $\bar{v}_m(p, u) = v_m(p, u) + \sum_{i=1}^m v_i(p_k^i, u)$ if $p \in \bigcap_{i=1}^m G_i$. Therefore,

$$(7.1) \quad \begin{aligned} N_m \circ \mathfrak{U}_m(p, u, \bar{v}_m(p, u)) &= N_m(p, u, v_m(p, u)), \\ \frac{\partial}{\partial p} N_m \circ \mathfrak{U}_m(p, u, \bar{v}_m(p, u)) &= \frac{\partial N_m}{\partial p}(p, u, v_m(p, u)), \end{aligned}$$

whenever $p \in \bigcap_{i=1}^m G_i$. It is easy to see that the function $N_m(p, u, v(p, u))$ is a convex function of p for fixed u . Indeed, it follows from $\frac{\partial N}{\partial v}(p, u, v(p, u)) = 0$ that

$$(7.2) \quad \frac{DN}{Dp} = \frac{\partial N}{\partial p} \Big|_{\partial_v N=0}.$$

Thus the Hessian matrix

$$(7.3) \quad \frac{D^2 N}{Dp^2} = \frac{\partial^2 N}{\partial p^2} - \frac{\partial^2 N}{\partial p \partial v} \left(\frac{\partial^2 N}{\partial v^2} \right)^{-1} \frac{\partial^2 N}{\partial v \partial p}$$

is a positive definite matrix because

$$\begin{aligned} &\begin{pmatrix} I & -H_{pv}H_{vv}^{-1} \\ 0 & H_{vv}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} H_{pp} & H_{pv} \\ H_{vp} & H_{vv} \end{pmatrix} \begin{pmatrix} I & 0 \\ -H_{vv}^{-1}H_{vp} & H_{vv}^{-\frac{1}{2}} \end{pmatrix} \\ &= \begin{pmatrix} H_{pp} - H_{pv}H_{vv}^{-1}H_{vp} & 0 \\ 0 & I \end{pmatrix}. \end{aligned}$$

We use $\Omega(k, D) = \{\omega \in \mathbb{R}^{n-1} : |\langle k, \omega \rangle| \leq D|k|^{-\nu}\}$ to denote the k -resonant strip, $S(k) = \{\omega \in \mathbb{R}^{n-1} : \langle k, \omega \rangle = 0\}$ to denote the k -resonant hyperplane, and define $F_m(p) = \max_{u \in S^1} N_m \circ \mathfrak{U}_m(p, u, \bar{v}_m(p, u))$, since $N_m \circ \mathfrak{U}_m$ can be extended to the whole circle S^1 as the function of u . For the m -th step of iteration, we need to cut off a set

$$A_m^{(1)} = \bigcup_{k \in \mathbb{Z}^{n-1}, |k| \leq K_m} \nabla F_m^{-1}(\Omega(k, D)) \cup \left(\nabla F_m^{-1}(S(k)) + \frac{1}{2}s_m^4 \right)$$

from the p -action space. $\nabla F_m^{-1}(\Omega(k, D))$ is cut off so that the image of the complement set of $A_m^{(1)}$, under the map \mathfrak{F}_m , does not fall into the resonant strips, and $\nabla F_m^{-1}(S(k)) + \frac{1}{2}s_m^4$ is cut off so that the distance between any two different connected components of the complement of $A_m^{(1)}$ is not smaller than s_m^4 . If \mathfrak{F}_m is a diffeomorphism, $\nabla F_m^{-1}(S(k)) + \frac{1}{2}s_m^4 \subset \nabla F_m^{-1}(\Omega(k, D))$. Because ∇F_m^{-1} is a Lipschitz map, the upper bound of the Lipschitz constant is independent of m and δ_m . We find that

$$(7.4) \quad \mathbf{m} \left(A_m^{(1)} \cap B_1 \right) \leq c_{11} \Sigma_m D \mathbf{m}(B_1),$$

where

$$\Sigma_m = \sum_{|k| \leq K_m, k \in \mathbb{Z}^{n-1}} \frac{1}{|k|^\nu}.$$

Obviously, $\Sigma_\infty = \lim_{m \rightarrow \infty} \Sigma_m < +\infty$.

As we are treating the case of weak persistency, we need to cut off another set $A_m^{(2)}$ so that condition (4.1b) holds for the remaining part. Thanks to Lemma 3.3 we have

$$(7.5) \quad \mathbf{m} \left((A_m^{(2)} + s_m^4) \cap B_1 \right) \leq c_0 s_m^2 \delta_0 \mathbf{m}(B_1).$$

In order for condition (4.1c) to hold for the $(m + 1)$ -th step of iteration, we need to cut off one more set, $A_m^{(3)}$, from the complement set of $A_m^{(1)} \cup A_m^{(2)}$. Let us consider a connected component of $G_{m-1} \setminus \left(A_m^{(1)} \cup (A_m^{(2)} + s_m^4) \right)$. If its diameter is not greater than $2s_m^3$, then we keep it in G_m . If its diameter is great than $2s_m^3$, we cut it into several parts by as few as possible straight strips with the width s_m^4 , so that the diameter of each part is between s_m^3 and $2s_m^3$. Thus,

$$(7.6) \quad \mathbf{m} \left(A_m^{(3)} \cap B_1 \right) \leq c_{12} s_m \mathbf{m}(B_1).$$

We define

$$G_m = G_{m-1} \setminus \left(A_m^{(1)} \cup (A_m^{(2)} + s_m^4) \cup A_m^{(3)} \right).$$

Obviously, the object in this section is to obtain the measure estimate on the set $\lim_{m \rightarrow \infty} G_m$. To do it, we should be aware that the two sets $\nabla F_m^{-1}(\Omega(k, D))$ and $\nabla F_{m-1}^{-1}(\Omega(k, D))$ do not coincide with each other in the general case. However, thanks to Lemma 3.1, we have

$$(7.7) \quad \|\nabla F_m^{-1} - \nabla F_{m-1}^{-1}\| \leq 2\sqrt{\frac{s_m^{13}}{\lambda}} < s_m^6.$$

It follows that

$$(7.8) \quad \nabla F_m^{-1} \circ \mathfrak{F}_{m-1} \left(A_{m-1}^{(1)} \right) \subset A_{m-1}^{(1)} + s_m^6,$$

$$(7.9) \quad A_{m+1}^{(1)} \subset \left(A_m^{(1)} + s_m^6 \right) \cup \tilde{A}_{m+1},$$

where

$$(7.10) \quad \begin{aligned} \tilde{A}_{m+1} &= \bigcup_{K_m < |k| \leq K_{m+1}} \nabla F_{m+1}^{-1}(\Omega(k, D)) \cup \left(\nabla F_{m+1}^{-1}(S(k)) + \frac{1}{2}s_{m+1}^4 \right), \\ \mathbf{m} \left(\tilde{A}_{m+1} \cap B_1 \right) &\leq c_{11} (\Sigma_{m+1} - \Sigma_m) D \mathbf{m}(B_1). \end{aligned}$$

Now we are ready to show that $\mathbb{S} = \lim_{m \rightarrow \infty} G_m$ has near full measure. By using (7.9) inductively we obtain that

$$(7.11) \quad \lim_{m \rightarrow \infty} \mathbf{m} \left(\bigcup_{m \in \mathbb{Z}^+} A_m^{(1)} \cap B_1 \right) \leq c_{11} \mathbf{m} \left(\bigcup_{k \in \mathbb{Z}^{n-1} \setminus \{0\}} \Omega(k, D) \cap B_1 \right) + \sum_{i=0}^{\infty} s_m^6 \mathbf{m}(B_1) \leq (c_{11} \Sigma_{\infty} D + 2s_0^6) \mathbf{m}(B_1).$$

Using (7.5) and (7.6) inductively we obtain that

$$(7.12) \quad \lim_{m \rightarrow \infty} \mathbf{m} \left(\bigcup_{m \in \mathbb{Z}^+} A_m^{(2)} \cap B_1 \right) \leq \sum_{m=0}^{\infty} c_0 s_m^2 \delta_0 \mathbf{m}(B_1) \leq 2c_0 s_0^{15} \mathbf{m}(B_1),$$

$$(7.13) \quad \lim_{m \rightarrow \infty} \mathbf{m} \left(\bigcup_{m \in \mathbb{Z}^+} A_m^{(3)} \cap B_1 \right) \leq \sum_{m=0}^{\infty} c_{12} s_m \mathbf{m}(B_1) \leq 2c_{12} s_0 \mathbf{m}(B_1).$$

Recall (5.3) where we choose $D = d_1^{\frac{2}{169}}$. With (7.10)~(7.13), we finally convince ourselves of the following measure estimate for \mathbb{S} (see (1.7)):

$$\mathbf{m}(\mathbb{S} \cap B_1) \geq \left(1 - \beta d_1^{\frac{2}{169}}\right) \mathbf{m}(B_1),$$

where $\beta = \max\{(c_{11} \Sigma_{\infty})^{\frac{2}{13}}, 3c_{12}\}$. Recall the translation $v \rightarrow v + v_m(p, u)$ in each step of iteration. Given $p \in \mathbb{S}$, let $\Theta(p) \subset \mathbb{T}$ be the set where $V_{\infty}(p, u) = \sum_{m=1}^{\infty} v_m(p, u)$ is well defined for u . Clearly, each $u \in \Theta$ corresponds to a minimal $(n - 1)$ -torus. Thus, $\Phi_{\infty}(p) = \bigcup_{u \in \Theta(p)} V_{\infty}(p, u)$ is a set function. Nevertheless, we have (1.6)

$$\max_{u \in \Theta(p)} |V_{\infty}(p, u) - \Phi_0(p)| \leq \sum_{m=1}^{\infty} |v_m| \leq \sum_{m=0}^{\infty} \frac{s_m^{10}}{\lambda} \leq 2s_0^9.$$

Let $\mathfrak{F}_{\infty} = \lim_{m \rightarrow \infty} \mathfrak{F}_m$, it is single valued on \mathbb{S} . Clearly, the inverse of \mathfrak{F}_{∞} is Lipschitz. By applying Lemma 3.1 we obtain (1.5)

$$\left\| \mathfrak{F}_{\infty}^{-1} - \left(\frac{\partial N_0}{\partial p} \right)^{-1} \right\|_{\infty} \leq \sum_{m=0}^{\infty} \sqrt{\frac{s_m^{13}}{\lambda}} \leq 2s_0^6.$$

To obtain (1.8) we apply Lemma 3.2 to the frequency maps $\frac{\partial N_0}{\partial p}$ and \mathfrak{F}_{∞} and consider the images of the set $\mathbb{S} \cap \frac{\partial N_0}{\partial p}^{-1}(B_1)$

$$(7.14) \quad \mathbf{m} \left(\mathfrak{F}_{\infty} \left(\mathbb{S} \cap \frac{\partial N_0}{\partial p}^{-1}(B_1) \right) \right) \geq \mathbf{m} \left((1 - 2s_0)^n \frac{\partial N_0}{\partial p}(\mathbb{S}) \cap (B_1) \right).$$

Obviously, (1.8) follows from (1.7) as well as (7.14). This completes the whole proof of the theorem in the case that the critical points at each step of iteration have weak persistency.

The remaining part is to show the positive measure estimate in the case when some critical points have strong persistency. We assume that after the $(m - 1)$ -th

step of the iteration the critical point of the function $N_m(p, u, v(p, u))$ has strong persistency at \bar{p} , i.e. if we write

$$\left| \frac{d^2 N_m}{du^2}(\bar{p}, u, v(\bar{p}, u)) \right| = \beta_m(u) s_{m+1}^2,$$

then $\beta_m(u) \geq 1$. In this case, the critical points nearby also have strong persistency. For instance,

$$(7.15) \quad \beta_m(u) \geq \frac{3}{4}, \quad \forall \|p - \bar{p}\| \leq s_{m+1}^5.$$

When $\frac{1}{2} \leq \beta_m(u(\bar{p})) \leq 1$, we can treat it either by the iteration scheme for weak persistency or by the iteration scheme for strong persistency. In this case, if the frequency $\omega(p^*) = \partial_p N(p^*, u(p^*), v(p^*, u(p^*)))$ satisfies the diophantine condition (1.4) for all $k \in \mathbb{Z}^{n-1}$, the work of Graff [G] shows that the lower-dimensional torus of hyperbolic type exists under small perturbation. The frequency of the flow on this torus is exactly $\omega(p^*)$ (see (4.42)). According to the demonstration in this section above and (7.15), we can see that

$$\mathfrak{F}_{m_1}^{-1} \circ \mathfrak{F}_m \left\{ p \in \mathbb{R}^{n-1} : \beta_m(p) \leq \frac{3}{4} \right\} \supset \left\{ p \in \mathbb{R}^{n-1} : \beta_m(p) \leq \frac{1}{2} \right\}, \quad \forall m_1 \geq m.$$

On one hand, we cut off a set $\{\beta_m(p) \geq 1\} + s_{m+1}^5$ from G_m and carry out the following KAM iteration for weaker persistency on the complementary set. On the other hand, we can pick out every point $p \in \{\beta(p) \geq \frac{1}{2}\}$ which satisfies the diophantine condition (1.4), and run the KAM machine for strong persistency. In this way we can see that the measure estimate (1.6) and (1.7) also hold. This proves the theorem completely.

ACKNOWLEDGEMENT

This work was partially done when the author visited the Institute of Mathematical Sciences at The Chinese University of Hong Kong. The author thanks the hospitality of the Institute. This work is under the support of the state basic research project of China “Nonlinear Sciences” (G2000077303). The author would like to thank the referees—the revised version of this paper is made under their suggestions.

REFERENCES

- [A] V.I. Arnold, *Proof of a theorem of A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian* Russ. Math. Surv., **18** (1963), 9-36. MR0163025 (29:328)
- [Ba] V. Bangert, *Minimal geodesics* Ergod. Th. & Dynam. Sys., **10** (1989), 263-286. MR1062758 (91j:58126)
- [Bo] S. Bolotin, *Homoclinic orbits to invariant tori of Hamiltonian systems* Amer. Math. Soc. Transl. (2), **168** (1995), 21-90. MR1351032 (97b:58124)
- [C1] C.Q. Cheng, *Birkhoff-Kolmogorov-Arnold-Moser tori in convex Hamiltonian systems* Commun. Math. Phys. **177** (1996), 529-559. MR1385075 (97a:58161)
- [C2] C.Q. Cheng, *Lower dimensional invariant tori in the regions of instability for nearly integrable Hamiltonian systems* Commun. Math. Phys., **203** (1999), 385-419. MR1697603 (2000g:37085)
- [E] L. Eliasson, *Perturbations of stable invariant tori for Hamiltonian systems* Ann. Scuola Norm. Su. Pisa, **15** (1988), 115-148. MR1001032 (91b:58060)
- [G] S.M. Graff, *On the continuation of hyperbolic invariant tori for Hamiltonian systems* J. Diff. Eqns., **15** (1974), 1-69. MR0365626 (51:1878)

- [H1] G. Hedlund, *Geodesics on a two-dimensional Riemannian manifold with periodic coefficients* Annals Math., (2)**3** (1932), 719-739. MR1503086
- [H2] M. Herman, *Existence et non existence de tores invariants par des diffeomorphisms symplectiques*, (1988). Séminaire sur les Équations aux Dérivées Partielles 1987–1988, Exp. No. XIV, École Polytech., Palaiseau, 1988, 24 pp. MR1018186 (90m:58070)
- [M1] J. Mather, *Existence of quasiperiodic orbits for twist homeomorphisms of annulus* Topology, **21** (1982), 457-467. MR0670747 (84g:58084)
- [M2] J. Mather, *Action minimizing invariant measures for positive definite Lagrangian systems* Math. Z., **207** (1991), 169-207. MR1109661 (92m:58048)
- [M3] J. Mather, *Variational construction of connecting orbits* Ann. Inst. Fourier, **43** (1993), 1349-1386. MR1275203 (95c:58075)
- [M4] J. Mather, *Arnold diffusion* talk at Oberwolfach (2001).
- [Me] R. Mañé, *Generic properties and problems of minimizing measures of Lagrangian systems* Nonlinearity, **9** (1996), 273-310. MR1384478 (97d:58118)
- [Mo] J. Moser, *On invariant curves of area-preserving mappings of an annulus* Nachr. Akad. Wiss. Gött. Math. Phys., **K1** (1962), 1-20. MR0147741 (26:5255)
- [R] Rockfeller *Convex analysis*, Princeton Press, 1970. MR0274683 (43:445)
- [SZ] D. Salamon & E. Zehnder, *KAM theory in configuration space* Comment. Math. Helv., **64** (1989), 181-203. MR0982563 (90d:58045)
- [T] D.V. Treschev, *A mechanism for the destruction of resonance tori in Hamiltonian systems* Math. Sbornik, **68** (1991), 181-203. MR1025685 (91i:58124)

DEPARTMENT OF MATHEMATICS, NANJING UNIVERSITY, NANJING 210093, PEOPLE'S REPUBLIC OF CHINA – AND – THE INSTITUTE OF MATHEMATICAL SCIENCES, THE CHINESE UNIVERSITY OF HONG KONG, HONG KONG, CHINA

E-mail address: chengcq@nju.edu.cn