J Dyn Diff Equat (2008) 20:831–866
DOI 10.1007/s10884-008-9113-6

Full Measure Reducibility for Generic One-parameter Family of Quasi-periodic Linear Systems

Hai-Long Her · Jiangong You

Dedicated to Professor Zhifen Zhang on the occasion of her 80th birthday

Received: 3 July 2007 / Published online: 8 May 2008

Abstract Let $C^\omega(\Lambda, gl(m,\mathbb{C}))$ be the set of $m\times m$ matrices $A(\lambda)$ depending analytically on a parameter λ in a closed interval $\Lambda\subset\mathbb{R}$. Consider one-parameter families of quasi-periodic linear differential equations: $\dot{X}=(A(\lambda)+g(\omega_1 t,\ldots,\omega_r t,\lambda))X$, where $A\in C^\omega(\Lambda, gl(m,\mathbb{C}))$, g is analytic and sufficiently small. We prove that there is an open and dense set $\mathcal{A}$ in $C^\omega(\Lambda, gl(m,\mathbb{C}))$, such that for each $A(\lambda)\in\mathcal{A}$ the equation can be reduced to an equation with constant coefficients by a quasi-periodic linear transformation for almost all $\lambda\in\Lambda$ in Lebesgue measure sense provided that g is sufficiently small. The result gives an affirmative answer to a conjecture of Eliasson (In: Proceeding of Symposia in Pure Mathematics).

Keywords Reducibility · Quasi-periodic · KAM

1 Introduction and Main Result

A quasi-periodic linear system

$$\dot{X}=A(t)X \tag{1.1}$$

or the equivalent skew product system

$$\dot{X}=A(\varphi)X,\quad \dot{\varphi}=\omega \tag{1.2}$$

is said to be *reducible*, if there exists a so called quasi-periodic Lyapunov–Perron (L-P) transformation $X=P(\omega_1 t,\ldots,\omega_r t)Y$, such that the transformed system is a linear system with constant coefficients. We say that a linear time-varying change of variables $X=P(\omega_1 t,\ldots,\omega_r t)Y$ is a L-P transformation if P is quasi-periodic in t and non-singular for

H.-L. Her · J. You (✉)
Department of Mathematics, Nanjing University, Nanjing 210093, China
e-mail: jyou@nju.edu.cn

H.-L. Her
e-mail: hailongher@126.com

all $t \in \mathbb{R}$, P, P^{-1}, and $\dot{P} = dP/dt$ are bounded in $t \in \mathbb{R}$. A quasi-periodic linear system (1.1) is said to be *almost reducible* if for every $\varepsilon > 0$ there is a constant matrix C and an L-P transformation $X = P(t)Y$, such that the transformed system

$$\dot{Y} = B(t)Y \tag{1.3}$$

satisfying $|B - C| < \varepsilon$ for all $t \in \mathbb{R}$.

The reducibility problem of quasi-periodic linear systems has received much attention. For $r = 1$, i.e., the periodic case, the classical Floquet theory shows that there always exists a periodic change of variables $Y = P(\omega_1 t)X$ so that the transformed system has the form $\dot{Y} = \bar{A}Y, \dot{\varphi} = \omega_1$, where $\bar{A}$ is a constant matrix. For $r > 1$, i.e., quasi-periodic case, the system (1.1) or (1.2) is not always reducible (see [8] for example) or almost reducible. In this case, sufficient conditions which guarantee the reducibility had been given in earlier work by Johnson and Sell [9], and sufficient and necessary condition for almost reducibility was given by Coppel [2]. Their method fails when pure imaginery spectrum appears.

The first reducibility result, proved by Kolmogorov–Arnold–Moser (KAM) method, was given by Bogoljubov et al. [1]. Then Dinaburg and Sinai [3] proved that the linear Shrödinger equation

$$\ddot{x} + q(\omega_1 t, \ldots, \omega_r t)x = \lambda x \tag{1.4}$$

is reducible for "most" large enough λ, where ω is fixed and satisfy the Diophantine conditions:

$$|\langle k, \omega\rangle| > \frac{\gamma^{-1}}{|k|^{\sigma}}, \quad 0 \neq k \in \mathbb{Z}^r, \tag{1.5}$$

where $\gamma > 1,\ \sigma > r - 1$ are fixed positive constants. See also Rüssmann [17]. Later, Jorba and Simó [10] generalized the result to linear systems of the form

$$\dot{X} = (A + \lambda\bar{Q} + \lambda^2 Q(\omega_1 t, \ldots, \omega_r t))X, \quad X \in \mathbb{R}^m. \tag{1.6}$$

where $A, \bar{Q}$ are constant diagonal matrices and Q is a quasi-periodic analytic matrix with r basic frequencies, λ is a small parameter. By the KAM method, Jorba and Simó proved that there exists a positive measure Cantor set $\mathcal{E} \subset (0, \lambda_0)$, $\lambda_0 \ll 1$ such that for any $\lambda \in \mathcal{E}$ the system is reducible, provided the the following non-degeneracy conditions hold

$$\left|\alpha_i(\lambda) - \alpha_j(\lambda)\right| > \delta > 0, \quad \left|\frac{d}{d\lambda}(\alpha_i(\lambda) - \alpha_j(\lambda))\right| > \chi > 0, \quad \forall 1 \leq i < j \leq m \tag{1.7}$$

where $\alpha_i(\lambda),\ 1 \leq i \leq m$, are the eigenvalues of $\bar{A} = A + \lambda\bar{Q}$. In fact, Jorba and Simó only consider the worst case when all eigenvalues are pure imaginary.

There are two important developments since then: on one hand, the full measure reducibility of quasi-periodic linear Shrödinger equation was given by Eliasson [4] by adding a resonance-cancellation technique to the KAM iteration. The idea was firstly introduced by Moser and Pöschel [16]. More precisely, Eliasson [4] gave a full measure reducibility result for the linear Schrödinger equation

$$\frac{d^2x}{dt^2} + (\lambda + Q(\omega t))x = 0. \tag{1.8}$$

He proved that the above equation is reducible for almost all $\lambda \in (a, b)$ in Lebesgue measure sense provided that Q is small. The work was generalized by Krikorian to linear systems with coefficients in $so(3)$ and in general with coefficients in Lie algebra of compact semi-simple Lie group [12,13].

On the other hand, it is surprising that the local picture for reducibility is also available in the global case for some special systems. Rychlik [18] introduced a renormalization mechanism to research a kind of lower dimensional cocycles with golden mean rotation number where systems are not necessary to be perturbations of a constant system. More recently, using the renormalization technique, Krikorian showed that the reducible systems are global C^∞-dense for cocycles on $\mathbb{T}^1 \times SU(2)$ [14]. For higher dimensional cocycles, the global reducibility problem remains open.

In this paper we consider the following family of quasi-periodic linear systems

$$\dot{X} = (A(\lambda) + F(\omega_1 t, \ldots, \omega_r t, \lambda))X \tag{1.9}$$

or the equivalent skew product systems

$$\dot{X} = (A(\lambda) + F(\varphi, \lambda))X, \quad \dot{\varphi} = \omega, \tag{1.10}$$

where $t \in \mathbb{R}$, $X \in \mathbb{C}^m$, the parameter $\lambda \in \Lambda = (a, b)$, A is a $m \times m$ constant matrix, and $F(\varphi_1, \ldots, \varphi_r, \lambda)$ is an analytic mapping from $\mathbb{T}^r \times (a, b)$ to $gl(m, \mathbb{C})$, $(\omega_1, \ldots, \omega_r)$ is a Diophantine vector and $|F|$ is sufficiently small.

In a recent paper of Eliasson [6], he proved the almost reducibility of any quasi-periodic linear system close to constant. In this paper we shall prove that for generic family $\{A(\lambda)\}$, (1.10) is actually reducible for almost all parameters in Lebesgue measure sense. This is a conjecture of Eliasson in [6].

For the convenience of stating and proving our result, we introduce the following equivalent formulation of reducibility. Given an analytic quasi-periodic linear system

$$\dot{X} = a(t)X \tag{1.11}$$

or the skew-product system:

$$\dot{X} = A(\varphi)X, \quad \dot{\varphi} = \omega \tag{1.12}$$

where a, A are in the Lie algebra $g = gl(m, \mathbb{C})$, and the solutions of the equations take values in the Lie group $G = GL(m, \mathbb{C})$. If A is analytic on a complex neighborhood $W_h(\mathbb{T}^r)$ we denote $A \in C_h^\omega(\mathbb{T}^r, g)$. We say that two analytic g-valued functions $A_1, A_2 \in C_h^\omega(\mathbb{T}^r, g)$ are conjugated, if there exists a G-valued function $P \in C_h^\omega(\mathbb{T}^r, G)$, which is also L-P transformation, such that for two solutions X_1, X_2 corresponding to A_1, A_2, we have the relation

$$X_2 = P(\varphi)X_1$$

and denote the conjugate relation by

$$A_1 \equiv A_2 \; (mod \;\, P).$$

It is easy to see that $A_1 \equiv A_2 \; (mod \;\, P)$ is equivalent to the following equality:

$$A_2 = L_\omega P \cdot P^{-1} + P A_1 P^{-1} \tag{1.13}$$

where $L_\omega = \frac{\partial}{\partial \varphi} \cdot \dot{\varphi}$ is the derivative along frequency vector ω. A_1 is said to be reducible if it conjugates to a constant A_2. In this paper, we will prove that, under the following non-degeneracy conditions, for any parameter in a set of full measure, there exists a L-P transformation $P^*(\varphi)$ such that the system $A(\lambda) + F(\lambda, \varphi)$ is conjugate to a constant one A^*.

From now on, for a matrix $A(\lambda)$ defined on Λ, we denote its spectrum by $\sigma(A(\lambda))$. For $u \in \mathbb{R}$ we denote

$$g(\lambda, u) = g(\sigma(A(\lambda)), u) = \prod_{\alpha_i, \alpha_j \in \sigma(A); i \neq j} [(\alpha_i(\lambda) - \alpha_j(\lambda)) - \mathrm{i}u], \tag{1.14}$$

Remark The value of g will be used to estimate the size of the so called small divisor. We only need consider the worst situation when all $\alpha_i - \alpha_j$ are pure imaginary. If they have real part, the size of small divisor will be bigger than the one in the pure imaginary case, and it is easier to solve the homological equation. Therefore it is enough to restrict to the case when $\mathrm{Re}(\alpha_i - \alpha_j) = 0$ for all $1 \le i, j \le m$, not only for the initial constant A, but also for all A_n in the following iterative steps. For the analyticity of g in λ, we refer the reader to Appendix A.

We firstly introduce the non-degeneracy conditions for $A(\lambda)$.

Definition 1.1 We say that $A(\lambda)$(in the original system (1.10)) satisfies the non-degeneracy conditions on the interval Λ, if there exist $d \in \mathbb{Z}^+$, $\chi > 0$ such that for all $u \in \mathbb{R}$ the following inequalities hold uniformly for all $\lambda \in \Lambda$,

$$\max_{0 \le l \le d} \left| \frac{\partial^l g(\lambda, u)}{\partial \lambda^l} \right| > \chi. \tag{1.15}$$

Now we state our main result. Suppose that F is analytic g-valued function defined on a complex neighborhood of $\mathbb{T}^r \times \Lambda$:

$$W_{h,\delta}(\mathbb{T}^r \times \Lambda) = \{(\varphi, z) \in \mathbb{C}^r \times \mathbb{C}^1 \mid \mathrm{dist}(\varphi, \mathbb{T}^r) < h;\ \mathrm{dist}(z, \Lambda) < \delta\},$$

where $\Lambda = (a, b)$. The norm of F is defined as

$$|F|_{h,\delta} = \sup_{(\varphi,z) \in W_{h,\delta}(\mathbb{T}^r \times \Lambda)} \|F(\varphi, z)\|.$$

Similarly,

$$|A|_\delta = \sup_{z \in W_\delta(\Lambda)} \|A(z)\|,$$

where $\|\cdot\|$ denotes the matrix norm.

Theorem 1 *Suppose that the frequency ω is fixed and satisfy the Diophantine condition* (1.5) *and A satisfies the non-degeneracy conditions* (1.15) *on Λ. Then there exist positive real numbers ε_0, h, δ, such that if $F \in C^\omega_{h,\delta}(\mathbb{T}^r \times \Lambda, g)$ satisfying $|F|_{h,\delta} \le \varepsilon_0$, for almost every $\lambda \in \Lambda$, the system* (1.10) *is reducible i.e., there exists $B_\lambda \in C^\infty(\mathbb{T}^r, G)$ s.t.*

$$L_\omega B_\lambda(\varphi) \cdot B_\lambda^{-1}(\varphi) + B_\lambda(\varphi)[A(\lambda) + F(\varphi, \lambda)]B_\lambda^{-1}(\varphi) = \text{constant} \in g. \tag{1.16}$$

Remark In fact, for the real case i.e., $g = gl(m, \mathbb{R})$, $G = GL(m, \mathbb{R})$, the same conclusion holds. However, we will face the phenomenon of period-doubling, and the final transformation B_λ will generally be defined on $\mathbb{R}^r / C_G \mathbb{Z}^r$ instead of $\mathbb{T}^r$, where C is a constant only depending on the group G (see [12] for details). For simplicity and emphasizing our method, we will only give a proof for complex case.

Remark In above theorem, the coefficient matrices are in Lie algebra of noncompact Lie group. For systems with additional structure such as $sp(n)$, $su(n)$ etc., our method can not preserve their Lie algebraic structure. However, controlling eigenvalues in noncompact case is more difficult.

Remark As a concrete application, we claim that the system $\dot{X} = (\lambda A_0 + F(\varphi, \lambda))X$ is full measure reducible, provided all eigenvalues of the parameter-independent matrix A_0 are distinct. For this special case, all assumptions in the theorem are easy to verify. For more general case, $\dot{X} = (A(\lambda) + F(\varphi, \lambda))X$, if the differences of any two eigenvalues of $A(\lambda)$ are all polynomials of λ with order no less than 1, then all assumptions of Theorem 1 are satisfied. Moreover, our theorem can apply to the reducibility of linear shrödinger equations with sufficiently small potentials in which the coefficient matrix of constant part in the corresponding linear system is $\begin{pmatrix} 0 & 1 \\ -\lambda & 0 \end{pmatrix}$, and it is easy to get $g(\lambda, u) = u^2 + 4\lambda$, so $|\frac{\partial}{\partial \lambda} g(\lambda, u)| = 4$. Thus Theorem 1 also implies the result obtained by Eliasson [4].

Genericity of the non-degeneracy condition (1.15). Denote by

$$C^{\omega}(\Lambda, g) = \{A(\lambda) : \quad A(\lambda) \text{ is analytic in } \lambda \in \Lambda\},$$

by $\mathcal{A}$ the set of $A(\lambda) \in C^{\omega}(\Lambda, g)$ satisfying (1.15). We aim to prove that $\mathcal{A}$ is a open and dense set in $C^{\omega}(\Lambda, g)$. It is obvious that $\mathcal{A}$ is C^{ω}-open in $C^{\omega}(\Lambda, g)$.

To prove $\mathcal{A}$ is C^{ω}-dense in $C^{\omega}(\Lambda, g)$, let's recall that

$$\begin{aligned} g(\lambda, u) &= \prod_{\alpha_i, \alpha_j \in \sigma(A); i \neq j} [(\alpha_i(\lambda) - \alpha_j(\lambda)) - \mathrm{i}u] \\ &= \det[A(\lambda) \otimes I_m - I_m \otimes A(\lambda) - \mathrm{i}u I_{m^2}]/(-\mathrm{i}u)^m, \end{aligned}$$

since $A \in C^{\omega}(\Lambda, g)$, by the Appendix A we know that $g(\lambda, u)$ is an analytic function of the parameter λ. If $A(\lambda)$ does not satisfy the non-degeneracy conditions (1.15), then there exist $u_0 \in \mathbb{R}$ and $\lambda_0 \in \Lambda$, such that for $\forall\, d \in \mathbb{Z}^+ \cup \{0\}$ we have

$$\frac{\partial^d g(\lambda_0, u_0)}{\partial \lambda^d} \equiv 0.$$

Since $g(\lambda, u_0)$ is analytic for λ, so we can see $g(\lambda, u_0) \equiv 0$ for all $\lambda \in \Lambda$. Considering the expression of g, that means there exists $i \neq j$, such that for all $\lambda \in \Lambda$ the difference of two eigenvalues $\alpha_i(\lambda) - \alpha_j(\lambda) = \mathrm{i}u_0 \equiv \text{const}$.

For convenience we set $\Lambda = [0, 1]$, and $A([0, 1])$ is a C^{ω}-path in the matrix space $gl(m, \mathbb{C})$. We will construct a sequences of non-degenerate C^{ω}-paths $A_{\varepsilon}(\lambda)$, which converge to $A(\lambda)$, $\lambda \in [0, 1]$. We denote the eigenvalues of $A(0)$ by $\alpha_1, \alpha_2, \ldots, \alpha_m$, and eigenvalues of $A(1)$ by $\beta_1, \beta_2, \ldots, \beta_m$. Choosing m real numbers $b_1, b_2, \ldots, b_m$ such that for a sufficiently small positive number $\epsilon > 0$ and $\forall\, i < j$,

$$\alpha_i - \alpha_j \neq \beta_i' - \beta_j',$$

where $\beta_i' = \beta_i + b_i \epsilon$, $1 \leq i \leq m$. Moreover, we know that there exists invertible matrix $P(1)$ transforming the matrix $A(1)$ to Jordan form $J(1) = P^{-1}(1)A(1)P(1)$. So we can set $B = P(1)\mathrm{diag}(b_1, b_2, \ldots, b_m)P^{-1}(1)$ and write the sequences of C^{ω}-paths as

$$A_{\epsilon}(\lambda) = A(\lambda) + \lambda \epsilon B.$$

It is easy to see that when $\epsilon \to 0$, $A_{\epsilon}(\lambda) \to A(\lambda)$, and $A_{\epsilon}(\lambda)$ is non-degenerate if ϵ is sufficiently small. In fact, the difference of any two eigenvalues of $A_{\epsilon}(1) = A(1) + \epsilon B$ is different from the corresponding difference of $A_{\epsilon}(0) = A(0)$, so any difference of two eigenvalues of $A_{\epsilon}(\lambda)$ is impossible to be constant. Now we see that $\mathcal{A}$ is a C^{ω}-dense subset of $C^{\omega}(\Lambda, g)$.

2 Outline of the Proof

As it is well known, the KAM method can efficiently deal with the "small divisor" problem. However, at each iterative step, if we want to get a desired bound for solutions of the linearized

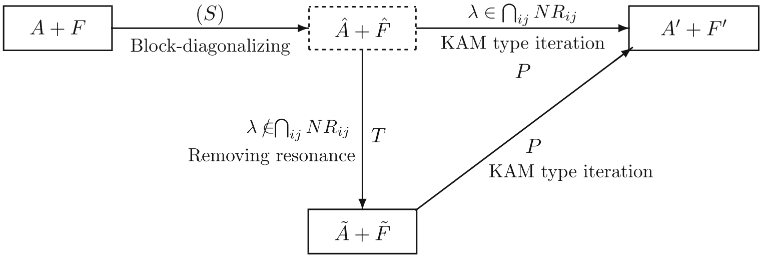

Fig. 1 One step of iteration

equations, we have to discard some sub-intervals of parameters on which the so-called non-resonant conditions are not satisfied. So there is at most a positive measure parameter set preserved after infinitely many times of iteration, i.e., we can only obtain the reducibility for a positive measure parameter set [7,10].

To prove a full measure reducibility result, one has to improve the KAM iterative method. At each KAM iteration, one cannot discard any parameter whenever the non-resonant conditions are satisfied or not. For this purpose, a procedure of cancelling resonances must be introduced. In fact, for the original system $A(\lambda)+F(\varphi,\lambda)$, if $A(\lambda)$ is of block diagonal form, one can always find a linear transformation $T(\varphi)$, which may not be close to the identity, to move some eigenvalues of $A(\lambda)$ such that the resonance does not happen. It means that the transformed system $\tilde{A}(\lambda)+\tilde{F}(\varphi,\lambda)$ satisfies the non-resonance conditions for all parameters. Then one step of KAM type iteration (using linear transformation $P(\varphi,\lambda)$) can be done for any parameter, i.e., the solution of the linearized equation is well bounded for all parameters $\lambda \in \Lambda$. We thus obtain a new system $A' + F'$ with F' smaller. Figure 1 shows what we will do in one step of iteration:

In Fig. 1, $\bigcap_{ij} NR_{ij}$ is the non-resoance set defined in (2.6). At each step of iteration, we have to divide the parameter interval into smaller sub-intervals in order to move eigenvalues. The above iteration step is done for each sub-interval. The transformation $S(\lambda)$ is used to make the constant matrix $A(\lambda)$ block diagonal: $\hat{A}(\lambda) = \text{ diag }(A_1,\ldots,A_s)$ where the set of eigenvalues of any two different blocks $\sigma(A_i(\lambda))$, $\sigma(A_j(\lambda))$ are sufficiently separated. This can always be done if the parameter interval is small enough. In each sub-interval of parameters, one has to use different change of variables to make the perturbation smaller depending on weather or not there are resonances. The change of variables for non-resonant case is different from the resonant case. Thus in total the change of variables depends only piece-wise analytically on $\lambda \in \Lambda$. In the iterative process we will also see that the new perturbation term F' can be made much smaller than it is in the classical KAM iterative scheme, e.g., at the $(n+1)$th step if we have got $|F_n| < \varepsilon_n$, then we can make $|F_{n+1}| < \varepsilon_n^{m^{4n}}$ instead of the normal estimation $|F_{n+1}| < \varepsilon_n^{1+\kappa},\ 0<\kappa<1$. This kind of super-exponential decaying of perturbation in the process of KAM-type iteration will overcome the "bad" influence resulting from the transformation of removing resonance which makes the perturbation much bigger, it is a critical point for our proof.

Repeating the above process, we can make the perturbation smaller and smaller not throwing away any parameter. At the same time, at the nth iterative step, we obtained $A_n + F_n$ with $|F_n|$ decaying super-exponentially and the non-degeneracy of A_n is preserved. Using the KAM theory to this system, we get the positive measure reducibility result for $A_n(\lambda) + F_n(\varphi,\lambda)$ (see [7]), i.e., $A_n(\lambda) + F_n(\varphi,\lambda)$ can be reduced to a constant system for all parameter points except a small positive measure subset $\Lambda - R_n$ of the parameter interval

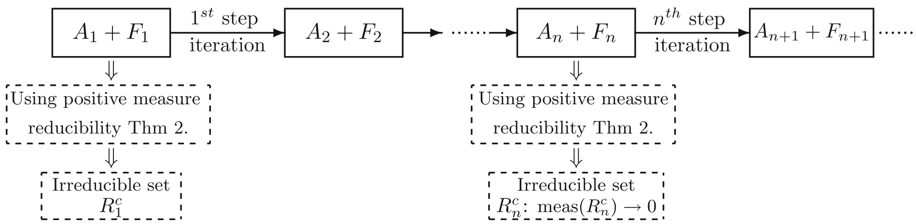

Fig. 2 Iteration procedure

Λ (R_n denotes the reducible parameter set). From the following Theorem 2 in this section we know that the Lebesgue measure of the exceptional subset $\Lambda - R_n$ is proportional to the size of the perturbation F_n. It follows that meas$(\Lambda - R_n) \to 0$ when $n \to \infty$. Thus, as the iteration goes on, the measure of the set of the irreducible parameter points gets smaller and smaller, in the end, we obtain the reducibility on a full measure subset of parameter interval Λ. Figure 2 clearly shows the iteration method.

In general, the above iteration is a formal process. In fact, all linear transformations T_n which are used to remove resonances are far from the identity, so the composition of infinitely many such transformations does not converge on whole parameter interval Λ. However we can prove the convergence on Λ except a set $\Lambda - R^*$ with measure zero, since only finite many removing-resonance transformations are used in each $\lambda \in R^*$. Therefore the composition of finitely many such transformations and infinitely many "KAM-type" transformations, which is more and more close to identity with "super-exponential" velocity, will ultimately converge. We remark that the number of the transformations for removing the resonances depends on λ.

One subtle thing by the above method is: every time when we apply the positive measure reducibility theorem, we need to verify the so-called non-degeneracy conditions. So we have to spend many words on discussing the preserving of non-degeneracy conditions along the iteration. Another subtle thing is: In each interval we have to exclude a small set of parameter of small Lebesgue measure proportional to the size of the perturbation by the positive measure reducibility result. Thus the number of intervals is crucial. We have to control the bound of the number of intervals at each iteration step.

In the following of the section, we give more detailed outline of proof.

2.1 One Step of KAM Type Iteration

At each iterative step, we have a system whose coefficient matrix is of the form $A + F$. For any parameter point λ, we will try to find a L-P transformation close to identity as

$$P(\varphi) = I + Y(\varphi), \tag{2.1}$$

where $Y \in C_h^\omega(\mathbb{T}^r, g)$ is very small and $P \in C_h^\omega(\mathbb{T}^r, G)$, which transforms the quasi-periodic system $\dot{X} = (A + F)X$ into

$$\dot{X} = (L_\omega P \cdot P^{-1} + P(A + F)P^{-1})X.$$

Since Y will be small, P^{-1} can be expanded as

$$P^{-1} = I - Y + Y^2 + O(\|Y\|^3).$$

It follows that

$$\begin{aligned} &L_\omega P \cdot P^{-1} + P(A+F)P^{-1} \\ &\quad = L_\omega Y \cdot (I - Y + Y^2 + O(\|Y\|^3)) + (I+Y)(A+F)(I - Y + Y^2 + O(\|Y\|^3)) \\ &\quad = A + L_\omega Y + [Y, A] + F - L_\omega Y \cdot Y + [Y, F] + AY^2 - YAY + O(\|Y\|^3) \end{aligned} \tag{2.2}$$

Roughly speaking, we need to find a small Y so that the transformed system is still of the form $\dot{X} = (A^+ + F^+)X$, where A^+ is block diagonal as A and F^+ is much smaller than F.

For this purpose, we need to find a Y solving the following linearized equation[1]

$$L_\omega Y - [A, Y] = -F \tag{2.3}$$

where $[A, Y] = AY - YA$, and prove

$$F^+ = -L_\omega Y \cdot Y + [Y, F] + AY^2 - YAY + O(\|Y\|^3),$$

is much smaller.

Expanding Y, F into Fourier series, substituting them into the Eq. (2.3) and comparing the corresponding Fourier coefficients in both sides of the equation, we have

$$\mathrm{i}\langle k, \omega\rangle Y_k - (AY_k - Y_k A) = -F_k. \tag{2.4}$$

The eigenvalues of the linear operator $\mathrm{i}\langle k, \omega\rangle Id + [A, \cdot]$ in the left part are

$$\mathrm{i}\langle k, \omega\rangle - (\alpha_i - \alpha_j), \quad 1 \le i, j \le m, \alpha_i, \alpha_j \in \sigma(A).$$

Especially when $k = 0$, the eigenvalues are $\alpha_i - \alpha_j$.

Without loss of generality, we assume that $A(\lambda) = \mathrm{diag}(A_1(\lambda), \ldots, A_s(\lambda))$ is block-diagonal (see Appendix B for proof),moreover, we suppose its spectrum

$$\sigma(A(\lambda)) = \sigma(A_1) \bigcup \cdots \bigcup \sigma(A_s)$$

is (ρ, ν)-separated on Λ in the following sense.

Definition 2.1 Given an interval Λ, and real number $\rho > 0$, if for any $\lambda \in \Lambda$ we uniformly have a division of eigenvalues, i.e. $\sigma(A(\lambda)) = \sigma_1(\lambda) \bigcup \cdots \bigcup \sigma_s(\lambda) (1 \le s \le m)$, satisfying for any $i \ne j$

$$\mathrm{dist}(\sigma_i(\lambda), \sigma_j(\lambda)) = \inf_{\lambda \in \Lambda; \alpha \in \sigma_i, \beta \in \sigma_j} |\alpha(\lambda) - \beta(\lambda)| \ge \rho,$$

then we say that the division of eigenvalues of $A(\lambda)$ are ρ-separated on Λ. Moreover, if $\exists\, \nu > 0$ such that for each i,

$$\sup_{\lambda \in \Lambda; \alpha_k, \alpha_l \in \sigma_i} |\alpha_k(\lambda) - \alpha_l(\lambda)| \le \nu,$$

then we say that the division of eigenvalues are (ρ, ν)-separated on Λ.

From Appendix A we know that, for $k \ne 0$ the Eq. 2.4 is resolvable if and only if every eigenvalue of the operator $\mathrm{i}\langle k, \omega\rangle Id - [A, \cdot]$ is nonzero. So for the division $\sigma(A(\lambda)) = \sigma_1(\lambda) \bigcup \cdots \bigcup \sigma_s(\lambda)$ we define the non-resonance condition associated with σ_i, σ_j as

[1] As we will see in next section, the equation cannot be solved completely. Some "nice" φ-independent terms will be absorbed into A.

Definition 2.2 σ_i, σ_j is said to be (N, K)-nonresonant if

$$|\mathrm{i}\langle k, \omega\rangle + \alpha - \beta| \geq K^{-1}, \quad K > 0, \quad 0 \neq |k| \leq N, \tag{2.5}$$

holds for all $\alpha \in \sigma_i, \beta \in \sigma_j$.

For given $N, K > 0, 1 \leq i, j \leq s$, we denote the non-resonant parameter set as

$$NR_{ij}(N, K) = \left\{ \lambda \in \Lambda : \begin{array}{c} |\mathrm{i}\langle k, \omega\rangle + \alpha(\lambda) - \beta(\lambda)| \geq K^{-1}, \ 0 < |k| \leq N \\ \forall\, \alpha(\lambda) \in \sigma_i(\lambda), \ \beta(\lambda) \in \sigma_j(\lambda), \ 1 \leq i, j \leq s \end{array} \right\}. \tag{2.6}$$

If $\lambda \in \bigcap_{i,j} NR_{ij}(N, K)$, the Eq. 2.4 can be solved for $0 < |k| \leq N$. However, when $\lambda \notin \bigcap_{i,j} NR_{ij}(N, K)$ the classical KAM iteration is invalid. In this case, we need the following resonance cancellation procedure.

2.2 Removing the Resonance

For convenience, we denote $A(\lambda) \in NR_{ij}(N, K)$ if $\lambda \in NR_{ij}(N, K)$.

In order to make the iteration go on for the system with $\lambda \notin \bigcap_{ij} NR_{ij}(N, K)$, we have to remove the resonance in the eigenvalues of A. The original idea is from [16]. If for the parameter $\lambda \in \Lambda$, $A(\lambda)$ is not in $NR_{ij}(N, K)$, we try to find a transformation such that $\dot{X} = (A+F)X$ conjugates to a new system $\dot{X} = (\tilde{A}+\tilde{F})X$, where $\tilde{A}$ satisfies the non-resonant condition. More precisely, if the grouping of eigenvalues $\sigma(A(\lambda)) = \sigma(A_1) \bigcup \cdots \bigcup \sigma(A_s)$ is (ρ, ν)-separated on Λ and $\lambda \notin NR_{ij}(N, K)$, then we can prove that there exists unique $0 < |k_0| \leq N,\ s.t.$

$$|\mathrm{i}\langle k_0, \omega\rangle + \alpha - \beta| \leq K^{-1},$$

for $\alpha \in \sigma_i, \beta \in \sigma_j$. We set

$$T_{ij} = \exp(Q_{ij}(\varphi)), \quad Q_{ij}(\varphi) = \mathrm{diag}(O_{l_1}, \ldots, O_{l_{j-1}}, \mathrm{i}\langle k_0, \varphi\rangle O_{l_j}, O_{l_{j+1}}, \ldots, O_{l_s}).$$

Thus by (1.13) we have

$$\begin{aligned} \tilde{A} + \tilde{F} &= L_\omega T_{ij} \cdot T_{ij}^{-1} + T_{ij}(A + F)T_{ij}^{-1} \\ &= A + \mathrm{diag}(0, \ldots, 0, \mathrm{i}\langle k_0, \omega\rangle I_{l_j}, 0, \ldots, 0) + T_{ij} F T_{ij}^{-1}, \end{aligned}$$

since A and T_{ij} commute. Let

$$\tilde{A} = A + \mathrm{diag}(0, \ldots, \mathrm{i}\langle k_0, \omega\rangle I_{l_j}, \ldots, 0), \ \ \tilde{F} = T_{ij} F T_{ij}^{-1}.$$

It's easy to see that $\tilde{A}$ has the same block diagonal form as A, moreover, we can prove that for $0 < |k| \leq N^{\tau/\sigma}$, (refnonres) is satisfied for $\sigma(\tilde{A}_i)$ and $\sigma(\tilde{A}_j)$, i.e., $\tilde{A}(\lambda) \in NR_{ij}(N^{\tau/\sigma}, K)$ for suitable K.

Proceeding the above step finitely many times, we can remove all the resonances so that $\tilde{A}(\lambda) \in \bigcap_{i,j} NR_{ij}(N, K)$. After all the resonances are removed, one can do many steps of the KAM iteration. In fact, in the following Sect. 3 we will see that if we fix $\tau/\sigma = 2$, then at the nth step, there exists a number τ_n depending on the iterative step and a positive integer m_0 satisfying $\tau_n^{m_0} < N_n$, such that $\tilde{A}_n \in \bigcap_{ij}(\tau_n^{m_0} N_n, K_n)$. So we can truncate the Fourier series of $F_n(\varphi)$ up to order $\tau_n^{m_0} N_n$ and make the norm of the residual term $|R_n|$ smaller than $\frac{1}{2}\varepsilon_n^{m^{4n}}$. Since there is no resonance for $\tilde{A}_n$ if $|k| < \tau_n^{m_0} N_n$, we can do the classical KAM iteration many times until the new perturbation has the same size as the residual part R_n.

Note that the removing resonance procedure makes the perturbation bigger while the KAM step makes the perturbation be smaller. By careful estimates, can prove that, in total, the new perturbation is smaller after all the above steps have done. This finishes one step of iteration. Note that we do not lose any parameter in above procedure. However, the change of variables may be far from identity if the resonance exists.

We sub-divide the parameter interval Λ so that the resonance occurs for at most one k_0 in each sub-interval. Iterating the above steps will make the perturbation smaller and smaller which converges to zero super-exponentially. The problem is the composition of the change of variables does not converge if the resonance appears infinite many times.

If we stop at the nth step, we get a system

$$\dot{X} = (A_n + F_n)X \tag{2.7}$$

where $|F_n| < \varepsilon_n \to 0$. It is well-known that the measure of the excluding set of parameter depends on the size of the perturbation in KAM theory. Applying KAM theory to this system, we know that (2.7) is reducible for all parameters but a set of small measure $O(\varepsilon_n^c)$, where c is a constant only depending on the dimension of the equation. It follows that, beside a measure zero parameter set, (1.10) is reducible by an analytical change of variables. We remark that the change of variables does not even continuously depend on the parameter and its norm does not have an uniform upper bound with respect to λ.

2.3 A Result of Positive Measure Reducibility

At each iterative step, we need estimate the measure of parameter set on which the system $A_n(\lambda) + F_n(\varphi, \lambda)$ is irreducible by classical KAM method. Jorba and Simó gave a result for the system $\lambda A + F$, where A is diagonalizable and λ is very small. For the purpose of this paper, we need the following generalized result [7].

Theorem 2 *Suppose that ω is fixed and satisfies the Diophantine condition* (1.5)*, Λ is a parameter interval. $A \in C_\delta^\omega(\Lambda, g)$ satisfies non-degeneracy conditions* (1.15) *or the following* (2.9)*, $F \in C_{h,\delta}^\omega(\mathbb{T}^r \times \Lambda, g)$ and $\exists\, M > 0$ such that $|A(\cdot)|_\delta \le M$. Then there exist $\varepsilon_0 > 0,\ h > 0, s.t.$ if $|F(\cdot,\cdot)|_{h,\delta} = \varepsilon_1 \le \varepsilon_0$, the measure of the set of parameter λ for which the system* (1.9) *is non-reducible is no larger than $CL(10\varepsilon_1)^c$, where $C,\ c$, are some positive constants, L is the length of the parameter interval Λ.*

Remark The non-degeneracy conditions (1.15) or the following (2.9) is a little bit weaker than the non-degeneracy conditions used in [7], however it is easy to get the same positive measure result as in [7]. See Sect. 6 for a sketch proof.

2.4 Weaker Non-degeneracy Conditions

At each iterative step, in order to apply Theorem 2. To estimate the measure of parameter set on which the system is reducible by classical KAM method, some kind of non-degeneracy conditions is necessary. Thus we have to prove that the non-degeneracy conditions persist not only the standard KAM iteration but also the procedure of removing the resonance.

Without loss of generality, we suppose that at some iterative step we get the following skew-product systems

$$\dot{X} = (A(\lambda) + F(\varphi, \lambda))X, \quad \dot{\varphi} = \omega.$$

If we have a division $\sigma(A(\lambda)) = \sigma_1 \bigcup \cdots \bigcup \sigma_s$, then we set for $\forall\, u \in \mathbb{R}$

$$g_{ij}(\lambda, u) = \begin{cases} \prod_{\alpha_p \in \sigma_i, \beta_q \in \sigma_j} (\mathrm{i}u + \alpha_p - \beta_q) & i \neq j \\ \prod_{\alpha_p, \alpha_q \in \sigma_i, p \neq q} (\mathrm{i}u + \alpha_p - \alpha_q), & i = j. \end{cases} \tag{2.8}$$

Remark It is easy to see that if the division $\sigma(A) = \sigma_1 \bigcup \cdots \bigcup \sigma_s$ is sufficiently separated and $A \in C^\omega(\Lambda, g)$, then for $\forall\, 1 \leq i, j \leq s$, all of $g_{ij}(\lambda, u)$ are analytic function of λ (See the Appendices A and B for details).

Thus, if we suppose g_{ij} is analytic in λ, then at each iterative step we need to verify the following
Non-degeneracy Conditions. There exist an integer $d \geq 0$ and real number $\chi > 0$ such that for $\forall\, 1 \leq i \leq j \leq s$,

$$\max_{0 \leq l \leq d} \left| \frac{\partial^l}{\partial \lambda^l} g_{ij}(\lambda, u) \right| > \chi, \tag{2.9}$$

uniformly hold for all $\lambda \in \Lambda,\ u \in \mathbb{R}$.

Remark If we take $u = \langle k, \omega \rangle$, the non-degeneracy conditions will guarantee that the small denominator condition holds for "most" parameter λ. Moreover, we only consider those k such that $|\langle k, \omega \rangle| \leq 2(\Delta_0 + 1)$, because when $|\langle k, \omega \rangle|$ is large enough, the matrix $\mathrm{i}\langle k, \omega \rangle I_{l_i l_j} - (I_{l_i} \otimes A_j(\lambda) - A_i^T(\lambda) \otimes I_{l_j})$ is automatically nonsingular and the small denominator condition is obviously satisfied. Especially, when $l_i = 1$, the condition (2.9) is not required for g_{ii}. For systems (1.6) considered by Jorba and Simó where $g_{ij} = \mathrm{i}\langle k, \omega \rangle + \alpha_i(\lambda) - \alpha_j(\lambda)$, it is easy to see (1.7) implies the weaker non-degeneracy (2.9).

We will prove the property that $g_{ij}(\lambda, u)$ depends analytically(or smoothly) on λ can be preserved either under analytic small perturbations or after moving some eigenvalues by a constant, and will discuss the preserving of non-degeneracy conditions in the iterative process.

2.5 Full Measure Reducibility

As the iteration goes on, we get finer and finer partition Π_n of the parameter interval Λ such that on each sub-interval $\Lambda_{n,j}$ the non-degeneracy conditions (2.9) is satisfied by $A_n(\lambda), \lambda \in \Lambda_{n,j}$. Using the result of positive measure reducibility on each sub-interval $\Lambda_{n,j}$, and considering that the speed of increasing of the total number of sub-intervals is relatively slower (see Sect. 7 for details), we can conclude that when $n \to \infty$, the measure of parameter subset $R_n^c = \bigcup_j R_{n,j}^c$ contained in Λ, on which the system $A_n + F_n$ is irreducible, goes to zero since the perturbation converges to zero with the "super-exponential" velocity. Then we know that $\mathrm{meas}(\bigcap_{n=1}^\infty R_n^c) = 0$, that is to say for almost every $\lambda \in \Lambda$, $A_1(\lambda) + F_1(\lambda, \varphi)$ is reducible.

2.6 Plan of the Paper

In Sect. 3, we show what we will do at each iterative step. In Sect. 4 we discuss the preserving of non-degeneracy conditions and the regrouping of eigenvalues. Section 5 is the statement and proof of iterative lemma. In Sect. 6, a sketchy proof of the positive measure reducibility result, i.e., Theorem 2 is given. At the end, Sect. 7 shows the proof of the main theorem. Some preliminary knowledge about matrix and polynomial theory can be found in the Appendices.

In the following, for a m-dimensional vector $v = (v_1, \ldots, v_m)^T$ we denote its norm by $\|v\| = \max_{1 \leq j \leq m} |v_j|$. For a $m \times m$ matrix $N = (n_{ij})$, denote by $\|N\|$ its operator norm induced by the vector norm which is equivalent to $m \times \max |n_{ij}|$. And $|N|$ denotes the norm defined in the first section. For $k \in \mathbb{Z}^r$, we put $|k| = |k_1| + \cdots + |k_r|$. If f is a function, $|f|$ denotes its absolute value or modulus. Through the context we use letters C, c to designate positive constants which are independent of the iteration, in order to simplify the computations we allow them to take different values from one line to another when their actual values do not matter.

3 One Step of Iteration

Throughout this paper ω is a fixed Diophantine vector satisfying (1.5). At the n-th step of the KAM scheme, we have a partition of parameter sets Π_n and a system

$$\dot{x} = (A_n + F_n)x, \tag{3.1}$$

defined on $W_{h_n,\delta_n}(\mathbb{T}^r \times \Lambda)$ for each $\Lambda \in \Pi_n$, where A_n is a constant matrix satisfying the non-degeneracy condition (2.9) and F_n is a perturbation with

$$|F_n|_{h_n,\delta_n} < \varepsilon_n, \; |A|_{\delta_n} < \varepsilon_n^{-\frac{1}{4m^2}} \tag{3.2}$$

We then construct a linear transformation P_n, so that the transformed system

$$\dot{x} = (A_{n+1} + F_{n+1})x,$$

satisfies

$$|F_{n+1}|_{h_{n+1},\delta_{n+1}} < \varepsilon_{n+1} \sim \varepsilon_n^{m^{4n}}, \; |A_{n+1}|_{\delta_{n+1}} < \varepsilon_n^{-1}, \tag{3.3}$$

moreover, the constant matrix A_{n+1} satisfies non-degeneracy condition (2.9).

To simplify notations, in what follows, the quantities without subscripts refer to quantities at the nth step, while the quantities with subscripts $+$ denote the corresponding quantities at the $(n+1)$th step. Let us then consider the system

$$\dot{x} = (A + F)x, \tag{3.4}$$

defined in $W_{h_n,\delta_n}(\mathbb{T}^r \times \Lambda)$.

We now define

$$t_n = \left[\log_{\frac{3}{2}} m^{4n}\right] + 2 \tag{3.5}$$

$$\tau_n = 8t_n(3m^{4n} - 2) + 1 \tag{3.6}$$

$$N = \left[\frac{\tau_n - 1}{12(\tau_n^m - 1)} \cdot \frac{\ln \varepsilon^{-1}}{h}\right], \tag{3.7}$$

$$K = 2^{(\sigma+2)} \gamma N^{2\sigma}, \tag{3.8}$$

$$\nu = 4m\rho = K^{-1}, \tag{3.9}$$

$$\left(\frac{M}{\rho}\right)^{m(m+1)} \leq \left(\frac{M}{\rho}\right)^{2m^2} \leq \varepsilon^{-\frac{1}{6}}, \quad M \sim \rho \varepsilon^{-\frac{1}{12m^2}}, \tag{3.10}$$

where t_n, τ_n are positive integers depending on the iteration step, the symbol $[\cdot]$ denotes the integer part of a positive real number. We also define $h_+ = \frac{1}{2}h$, and $\tilde{h}_1, \ldots \tilde{h}_{t_n-1}$ such that

$$h > \tilde{h}_1 > \cdots > \tilde{h}_{t_n-1} > h_+,$$

$$h - \tilde{h}_1 = \tilde{h}_1 - \tilde{h}_2 = \cdots = \tilde{h}_{t_n-1} - h_+.$$

3.1 Finer Division of the Parameter Set Λ

Definition 3.1 The division of parameter set Π_+ is said to be finer than Π if for any $\Lambda_+ \in \Pi_+$ there is a $\Lambda \in \Pi$ such that $\Lambda_+ \subset \Lambda$.

Lemma 3.1 *Suppose $A \in C^\omega_\delta(\Lambda, g)$, $|A|_\delta \le M$, $|\sigma(A)| < \Delta$ on Λ. Then for any given ρ, $(0 < \rho < 1)$, there exist a partition Π of Λ into $\hat{\Lambda}_j$ with equal length, such that on each $\hat{\Lambda}_j$ we have a decomposition of eigenvalues $\sigma = \sigma_1 \bigcup \cdots \bigcup \sigma_s$ which is $(\rho, 4m\rho)$-separated. Moreover, there is a linear transformation S analytically depending on $\lambda \in \Lambda$, bounded by $\|S\|, \|S^{-1}\| \le \left(\frac{\|A\|}{\rho}\right)^{m(m+1)}$, such that*

$$B_1 = S^{-1}AS = \mathrm{diag}(A_1, \ldots, A_s),$$

such that $\sigma(A_i) = \sigma_i$, and on each interval $\Lambda \in \Pi_+$ the system is transformed into $\dot{X} = (B_1 + G_1)X$, where $G_1 = S^{-1}FS$ with $|G_1| \le \varepsilon^{\frac{5}{6}}$. Moreover, the number of intervals in Π is bounded by

$$\#\Pi \le 2^{2m^2+1}(1 + |\Lambda|)(\delta\Delta^{-1}\rho)^{-m^2}. \tag{3.11}$$

Proof We decompose the Λ into $|\Lambda| 2^{2m^2+1}(\delta\Delta^{-1}\rho)^{-m^2}$ intervals with equal length no more than $2^{-(2m^2+1)} \cdot (\delta\Delta^{-1}\rho)^{m^2}$. For any such sub-interval $\hat{\Lambda}_j$, choose a fixed $\lambda_0 \in \hat{\Lambda}_j$, then we can always find a division of eigenvalues, i.e., $\sigma(A(\lambda_0)) = \sigma_1(\lambda_0) \bigcup \cdots \bigcup \sigma_s(\lambda_0)$, such that $\sigma(A(\lambda_0))$ is $(3\rho, 3m\rho)$-separated.

For any eigenvalue $\alpha_q(\lambda)$ of $A(\lambda)$, if $\alpha_q(\lambda)$ is analytic on $\hat{\Lambda}_j$, then we consider the following function

$$D_{10}(\lambda) = \alpha_q(\lambda) - \alpha_q(\lambda_0).$$

Since $|\alpha_q(\lambda) - \alpha_q(\lambda_0)|_\delta \le 2\Delta$, by the Cauchy's estimation

$$\left| \frac{\partial}{\partial \lambda} D_{10}(\lambda) \right|_{\frac{\delta}{2}} \le 2\Delta\left(\frac{\delta}{2}\right)^{-1}$$

so consider the length of each sub-interval, one has

$$|D_{10}|_{\frac{\delta}{2}} \le 4\Delta\delta^{-1}|\lambda - \lambda_0| < \rho^{m^2} < \rho.$$

So $\alpha_q(\lambda)$ must be near some block $\sigma_i(\lambda_0)$ for all $\lambda \in \hat{\Lambda}_j$.

Otherwise, if λ_0 is a singular point for the eigenvalue $\alpha(\lambda)$, then by the perturbation theory for linear operator (see [11]), λ_0 must be an algebraic singular point or in another word, branch point of a global analytic function, i.e., there exists a group of eigenvalues $\{\mu_1(\lambda), \ldots, \mu_l(\lambda)\}$, called a *cycle*, satisfying $\mu_1(\lambda_0) = \cdots = \mu_l(\lambda_0) = \mu_0$, and μ_0 is the center of the cycle, l called the *period* of the cycle. Each μ_i is a branch of an analytic function $\mathcal{U}$, at the branch point μ_0, we have the Puiseux series such as

$$\mu_i(\lambda) = \mu_0 + c_1 E^{(i-1)}(\lambda - \lambda_0)^{1/l} + c_2 E^{2(i-1)}(\lambda - \lambda_0)^{2/l} + \cdots, \quad i = 1, 2, \ldots, l, \tag{3.12}$$

where $E = \exp(2\pi \mathrm{i}/l)$. Since $|E| = 1$, we have

$$|\mu_1(\lambda) - \mu_1(\lambda_0)| \doteq \cdots \doteq |\mu_l(\lambda) - \mu_l(\lambda_0)|.$$

if we omit the higher order small term $O((\lambda - \lambda_0)^{2/l})$. Now we consider the following function

$$D_{20}(\lambda) = \det[I_l \otimes U(\lambda) - U(\lambda_0) \otimes I_l],$$

where the matrix $U(\lambda) = \mathrm{diag}(\mu_1(\lambda) \ldots, \mu_l(\lambda))$. We know that D_{20} is an algebraic function and is analytic on $\hat{\Lambda}_j$, and

$$\left|\frac{\partial}{\partial\lambda} D_{20}\right|_{\frac{\delta}{2}} \leq (2\Delta)^{m^2}\left(\frac{\delta}{2}\right)^{-1},$$

so on $\hat{\Lambda}_j$

$$|D_{20}(\lambda)| = \prod_{1 \leq i \leq l} |\mu_i(\lambda) - \mu_i(\lambda_0)|^l \leq \left(\frac{\rho}{2}\right)^{m^2},$$

thus for each i

$$|\mu_i(\lambda) - \mu_i(\lambda_0)| \leq \frac{\rho}{2} + \epsilon < \rho$$

since $0 < \rho < 1$ and $|\hat{\Lambda}_j| \ll \rho$.

Now, for all $\lambda \in \hat{\Lambda}_j$, we can group eigenvalues as

$$\hat{\sigma}_i(\lambda) = \{\alpha_q(\lambda) : \ \mathrm{dist}(\alpha_q(\lambda), \sigma_i(\lambda_0)) < \rho\} \tag{3.13}$$

and $\sigma(\lambda) = \hat{\sigma}_1(\lambda) \bigcup \cdots \bigcup \hat{\sigma}_s(\lambda)$. Moreover, for $\forall\, \alpha_q \in \hat{\sigma}_i(\lambda),\ \beta_p \in \hat{\sigma}_j(\lambda), \lambda \in \hat{\Lambda}_j$,

$$\begin{aligned} |\alpha_q(\lambda) - \beta_p(\lambda)| &\geq |\alpha_q(\lambda_0) - \beta_p(\lambda_0)| - |\alpha_q(\lambda) - \alpha_q(\lambda_0)| - |\beta_p(\lambda) - \beta_p(\lambda_0)| \\ &\geq 3\rho - \rho - \rho = \rho \end{aligned} \tag{3.14}$$

and

$$\mathrm{diam}(\hat{\sigma}_i(\lambda)) \leq 2\rho + 3m\rho \leq 4m\rho.$$

Moreover, by the Appendix B, we can find a similarity transformation $S \in C^{\omega}_{\delta/2}(\Lambda, GL(m, \mathbb{C}))$ such that

$$B_1 = S^{-1}AS = \mathrm{diag}(A_1, \ldots, A_s)$$

with $|B_1|_{\delta/2} \leq C(\rho^{-1}|A|_\delta)^{2m^2}$, i.e., $\hat{M} \leq C(\rho M)^{2m^2}$, and $|G_1|_{h,\delta/2} \leq \left(\frac{M}{\rho}\right)^{m(m+1)} \varepsilon \leq \varepsilon^{\frac{5}{6}}$. □

In the following we will consider

$$\dot{X} = (B_1 + G_1)X \tag{3.15}$$

on each interval $\Lambda \in \Pi_+$ where $G_1 = S^{-1}FS$.

3.2 Removing the Resonance

In case the resonance exists, we have to find a change of variables which cancels the resonance before solving the homological equations. For any fixed N satisfying (3.7), the following lemma shows that no resonance will occur between eigenvalues of the same group if the diameter of the group is sufficiently small.

Lemma 3.2 *Let ω satisfy (1.5), $B_1(\lambda) = \mathrm{diag}(A_1, \ldots, A_s)$ be (ρ, ν)-separated on Λ, and N, K, ν be given in (3.7), (3.8), (3.9). Then A_i and A_i itself are (N^2, K) non-resonant; i.e., for $\forall \alpha_p, \alpha_q \in \sigma(A_i)$ and $\forall\, k \in \mathbb{Z}^r,\ 0 < |k| \le N^2$*

$$|\mathrm{i}\langle k, \omega\rangle + \alpha_p - \alpha_q| > K^{-1} \tag{3.16}$$

Proof By the assumption (3.9), $|\alpha_p(\lambda) - \alpha_q(\lambda)| < \nu = K^{-1}$ for $\forall\ \lambda \in \Lambda$, together with (3.8), we have

$$\begin{aligned}
&|\mathrm{i}\langle k, \omega\rangle - (\alpha_p(\lambda) - \alpha_q(\lambda))| \\
&\quad \ge |\langle k, \omega\rangle| - |\alpha_p - \alpha_q| \ge |\langle k, \omega\rangle| - \nu \\
&\quad \ge \frac{\gamma^{-1}}{|k|^\sigma} - K^{-1} \\
&\quad > K^{-1}.
\end{aligned} \tag{3.17}$$

□

In the following lemma, we show that if K is sufficiently large, the resonance between A_i and A_j occurs at most once for $0 < |k| \le N^2$ and can be removed by a change of variables.

Lemma 3.3 *Under the assumptions of Lemma 3.2, we have*

(i) *For any given $i \ne j$,*

$$|\mathrm{i}\langle k, \omega\rangle + \alpha - \beta| > K^{-1} \tag{3.18}$$

holds for all $\alpha \in \sigma(A_i)$, $\beta \in \sigma(A_j)$, $|k| \le N^2$, except at most for one $k_0 \in \mathbb{Z}^r$ with $0 < |k_0| \le N^2$;

(ii) *If there is a k_0 such that (3.18) does not hold, then we can find a linear transformation $T_{ij}(\varphi) = \exp(Q_{ij}(\varphi))$ which transforms (3.15) into $\dot{X} = (B_2 + G_2)X$, where*

$$Q_{ij}(\varphi) = \mathrm{diag}(O_{l_1}, \ldots, O_{l_{j-1}}, \mathrm{i}\langle k_0, \varphi\rangle I_{l_j}, O_{l_{j+1}}, \ldots, O_{l_s}),$$

$$B_2 = \mathrm{diag}(\tilde{A}_1, \ldots, \tilde{A}_s)$$

with

$$\begin{cases} \tilde{A}_j &= A_j + \mathrm{i}\langle k_0, \omega\rangle I_{l_j} \\ \tilde{A}_i &= A_i \end{cases} \quad \text{for} \quad i < j;$$

and

$$G_2 = T_{ij} G_1 T_{ij}^{-1}.$$

Moreover, $\tilde{A}_i$ and $\tilde{A}_j$ are (N^2, K)-nonresonant.

Proof (i) Suppose there exist two integers k_0, k_1 with $0 < |k_0|, |k_1| < N^2$, and $\alpha_0, \alpha_1 \in A_i, \beta_0, \beta_1 \in A_j$ such that

$$|\mathrm{i}\langle k_0, \omega\rangle + \alpha_0 - \beta_0| < K^{-1}.$$

$$|\mathrm{i}\langle k_1, \omega\rangle + \alpha_1 - \beta_1| < K^{-1}.$$

It follows that

$$\begin{aligned} |\langle k_0 - k_1, \omega\rangle| &= |\mathrm{i}\langle k_0, \omega\rangle - (\beta_0 - \alpha_0) - (\mathrm{i}\langle k_1, \omega\rangle - (\beta_1 - \alpha_1))| + 2\nu \\ &\le 4K^{-1}. \end{aligned} \tag{3.19}$$

On the other hand, by the diophantine conditions of ω,

$$|\langle k_0 - k_1, \omega\rangle| \ge \frac{\gamma^{-1}}{|k_0 - k_1|^\sigma} \ge \frac{\gamma^{-1}}{(2N^2)^\sigma}. \tag{3.20}$$

It follows that

$$\gamma^{-1}(2N^2)^{-\sigma} \le 4K^{-1}$$

which contradicts with (3.8).

(ii) Suppose that there exist an integer k_0 with $0 < |k_0| < N$, and $\alpha_0 \in A_i, \beta_0 \in A_j$ such that

$$|\mathrm{i}\langle k_0, \omega\rangle + \alpha_0 - \beta_0| < K^{-1}.$$

Let

$$T_{ij} = \exp(Q_{ij}(\varphi)), \quad Q_{ij}(\varphi) = \mathrm{diag}(0_{l_1}, \dots, 0_{l_{j-1}}, \mathrm{i}\langle k_0, \varphi\rangle I_{l_j}, 0_{l_{j+1}}, \dots, 0_{l_s}).$$

Thus by (1.13) we have

$$\begin{aligned} B_2 + G_2 &= L_\omega T_{ij} \cdot T_{ij}^{-1} + T_{ij}(B_1 + G_1)T_{ij}^{-1} \\ &= B_1 + \mathrm{diag}(0, \dots, 0, \mathrm{i}\langle k_0, \omega\rangle I_{l_j}, 0, \dots, 0) + T_{ij}G_1T_{ij}^{-1}, \end{aligned}$$

since B_1 and T_{ij} commute. Let

$$B_2 = B_1 + \mathrm{diag}(0, \dots, \mathrm{i}\langle k_0, \omega\rangle I_{l_j}, \dots, 0), \;\; G_2 = T_{ij}G_1T_{ij}^{-1},$$

it's easy to see that B_2 has the same block diagonal form as B_1. Note that $\sigma(\tilde{A}_j) = \sigma(A_j) + \mathrm{i}\langle k_0, \omega\rangle$. For any $0 < |k| \le N^2, \tilde{\alpha}_q \in \sigma(\tilde{A}_i), \tilde{\beta}_p \in \sigma(\tilde{A}_j)$ we have

$$\begin{aligned} &|\mathrm{i}\langle k, \omega\rangle - (\tilde{\beta}_p - \tilde{\alpha}_q)| \\ &\quad = |\mathrm{i}\langle k, \omega\rangle + \mathrm{i}\langle k_0, \omega\rangle - (\beta_p - \alpha_q)| \\ &\quad \ge |\langle k, \omega\rangle| - |\mathrm{i}\langle k_0, \omega\rangle - (\beta_p - \alpha_q)| \\ &\quad \ge \frac{\gamma^{-1}}{|k|^\sigma} - |\mathrm{i}\langle k_0, \omega\rangle - (\beta_0 - \alpha_0)| - 2\nu \\ &\quad \ge \gamma^{-1}N^{-2\sigma} - 3K^{-1} \ge K^{-1}. \end{aligned} \tag{3.21}$$

□

Lemma 3.3 can be applied to cancel all the resonances. In fact we have the following lemma.

Lemma 3.4 *Suppose that the assumptions of Lemma 3.3 holds. there is a linear transformation $T(\varphi) = \exp(Q(\varphi))$ such that (3.15) is transformed into*

$$\dot{X} = (B + G)X \tag{3.22}$$

where

$$B = (\tilde{A}_1, \ldots, \tilde{A}_s) = (A_1 + \langle k_1, \omega\rangle, \ldots, A_s + \langle k_s, \omega\rangle), |k_i| < \frac{\tau_n^{m_0} - 1}{\tau_n - 1} N, m_0 \le m - 1,$$

$$G = T^{-1} G_1 T,$$

satisfying

$$|\mathrm{i}\langle k, \omega\rangle + \alpha - \beta| \ge K^{-1}, \tag{3.23}$$

for any $\alpha, \beta \in \cup\sigma(\tilde{A}_i)$, $0 < |k| \le \tau_n^{m_0} N$, and

$$|G|_{h,\delta/2} \le \varepsilon^{\frac{2}{3}}; \tag{3.24}$$

moreover, B is non-degenerated.

Proof The procedure is simple. Fix A_1. If A_i is (N, K)-resonant with A_1, we say A_i is in the same group with A_1. We apply Lemma 3.3 to move A_i to $\tilde{A}_i$ so that $\tilde{A}_i$ is not resonant with A_1 for all $0 \ne |k| \le N^2$. In this way, we moved all A_i in the group of A_1 to $\tilde{A}_i$ so that there is no resonance in this group for $0 \ne |k| \le N^2$. Then repeat this procedure for the remained. By applying Lemma 3.3 s' times, we get $\tilde{A}_1, \ldots, \tilde{A}_s$, which can be divided as $s - s'$ groups. Denote by $\mathcal{B} = (\tilde{B}_1, \ldots, \tilde{B}_{s-s'})$ where $\tilde{B}_i$, $i = 1, \ldots, s - s'$ are block diagonalized matrices composed by $\tilde{A}_i$, $i = 1, \ldots, s$ such that there is no resonance for $0 \ne |k| \le N^2$ in each $\tilde{B}_i$. Now we repeat the above procedure for $\tilde{B}_1, \ldots, \tilde{B}_{s-s'}$. If $\tilde{B}_i$ is $(\tau_n N, K)$-resonant with $\tilde{B}_1$, we say $\tilde{B}_i$ is the same group with $\tilde{B}_1$. Group the remained in the same way and apply Lemma 3.3 in each group. Going on this procedure, by applying Lemma 3.3 m_0 times, $m_0 \le m - 1$, we get the conclusion of Lemma 3.4. □

Since T_{ij} is big, the new perturbation G might be much bigger than G_1. In the worst situation, considering (3.7) we have

$$\begin{aligned} |G|_{h,\delta/2} &= |T_{m_0-1}^{-1} \cdots T_1^{-1} G_1 T_1 \cdots T_{m_0-1}|_{h,\delta/2} \\ &< e^{2Nh(1+\cdots+\tau_n^{m_0-1})} |G_1|_{h,\delta/2} = e^{\frac{2}{\tau_n - 1}\left(\tau_n^{m_0} - 1\right)Nh} |G_1|_{h,\delta/2} < \varepsilon^{-\frac{1}{6}} \varepsilon^{\frac{5}{6}} = \varepsilon^{\frac{2}{3}}. \end{aligned} \tag{3.25}$$

The following two subsections are standard KAM iteration.

3.3 Truncation of the Perturbation

Let $T_{\tau_n^m N} G = \sum_{|k| \le \tau_n^m N} G_k e^{\mathrm{i}\langle k, \varphi\rangle}$ be the truncation of the Fourier expansion of G up to $|k| \le \tau_n^m N$. We denote the remained part by $R_{\tau_n^m N} G = G - T_{\tau_n^m N} G$.

Firstly, we state some basic estimates, considering (3.6) and (3.7) the proof is direct.

Proposition 3.1 *For any function $G \in C_h^\omega(\mathbb{T}^r, g)$, $G = \sum_{k \in \mathbb{Z}^r} G_k e^{\mathrm{i}\langle k, \varphi\rangle}$, one has*

$$|G_k| \le |G|_h e^{-|k|h},$$

Moreover, there exist constants $C > 0$, depending on the dimension such that, for any $h' < h$ and $k_0 \in \mathbb{Z}^r$, one has:

$$|T_{\tau_n^m N} G|_{h',\delta'} \leq \frac{C}{(h-h')^r}|G|_{h,\delta} \tag{3.26}$$

$$|R_{\tau_n^m N} G|_{h',\delta'} \leq C\frac{e^{-(\tau_n^m N+1)(h-h')}}{(h-h')^r}|G|_{h,\delta} \leq \varepsilon_n^{m^{4n}} \tag{3.27}$$

Since the non-resonance conditions hold for all $0 \neq |k| < \tau_n^m N$, and $|R_{\tau_n^m N} G| \leq \varepsilon_n^{m^{4n}}$, then we can iterate the standard KAM type step $t_n \geq \log_{\frac{3}{2}} m^{4n} + 1$ times so that the perturbation is of size $\varepsilon_n^{m^{4n}}$.

The following lemma is useful in the next subsection.

Lemma 3.5 *Given an invertible $m \times m$ matrix U and a $m \times m$ matrix V. If their norms satisfy the relation: $|V| \leq \frac{1}{2}|U^{-1}|^{-1}$, then $U + V$ is invertible and one has the estimation*

$$|(U+V)^{-1}| \leq 2|U^{-1}|. \tag{3.28}$$

Proof Since $|U^{-1}V| \leq |U^{-1}| \cdot |V| \leq \frac{1}{2}$, we know that $U + V = U(Id + U^{-1}V)$ is invertible. Thus,

$$\begin{aligned}|(U+V)^{-1}| &\leq |U^{-1}| \cdot |(Id + U^{-1}V)^{-1}| \\ &\leq |U^{-1}|(|Id| + |U^{-1}V| + |U^{-1}V|^2 + \cdots) \\ &\leq |U^{-1}|\sum_{i=0}^{\infty}(\frac{1}{2})^i \leq 2|U^{-1}|.\end{aligned}$$

□

3.4 One KAM Step

In this subsection A is B, F is G, N is $\tau_n^m N$ of last subsection. In fact, in order to make the new perturbation F_+ smaller than ε^{m^4} we shall do the KAM iteration t_n times, for simplicity we only write the process at the first step. And we denote $\tilde{h}_1$ by h'.

At first, we solve the Eq. 2.3.

$$L_\omega Y - [A, Y] = -(T_N F - [T_N F]) = -\dot{T}_N F \tag{3.29}$$

where $T_N F = \sum_{|k| \leq N} F_k e^{\mathrm{i}\langle k,\varphi\rangle}$ is the truncation of the Fourier expansion of F up to $|k| \leq N$, $[T_N F]$ is its average. We denote the remained part by $R_N F = F - T_N F$. And the constant matrix $[T_N F]$, which cannot be solved will be added to B so that $B' = B + [T_N F]$.

Since the small divisor conditions holds for $0 \neq |k| \leq N$, (3.29) can be solved by the following lemma.

Lemma 3.6 *Let $\Lambda \in \Pi_+$ be some parameter segment, $A \in C_\delta^\omega(\Lambda, g)$ such that $|A|_\delta \leq M$, and for $\forall\lambda \in \Lambda$ and $0 < |k| \leq N$, the small denominator conditions hold, i.e.,*

$$|\mathrm{i}\langle k, \omega\rangle + \alpha - \beta| \geq K^{-1}, \tag{3.30}$$

holds for all $0 \neq |k| \leq N, \alpha, \beta \in \sigma(A)$. Then there exist $\delta' > 0$ and a map $Y \in C_{h,\delta'}^\omega(\mathbb{T}^r \times \Lambda, g)$, such that Y solves (3.29), and

$$|Y|_{h',\delta'} \leq CM^{m^2}K^{m^2}\frac{|F|_{h,\delta}}{(h-h')^{m\sigma+r}}$$

where $\delta' = CM^{-(m^2+1)}N^{-m\sigma}K^{-m^2}\delta$. *Moreover, the transformation* $x \to (I+Y)x$ *sends* (3.22) *to the following system*

$$\dot{x} = (A + [F] + \tilde{F})x,$$

where

$$\tilde{F} = -L_\omega Y \cdot Y + [Y, F] + AY^2 - YAY + R_N F + O(\|Y\|^3), \tag{3.31}$$

satisfying

$$|\tilde{F}|_{h',\delta'} < |F|_{h,\delta}^{\frac{3}{2}}. \tag{3.32}$$

Proof Expanding Y, F into Fourier series, substituting them into the Eq. 3.29 and comparing the corresponding Fourier coefficient in both sides, one has

$$\mathrm{i}\langle k, \omega\rangle Y_k - (AY_k - Y_k A) = -F_k, \quad 0 \neq |k| \leq N \tag{3.33}$$

In fact, (3.30) implies

$$|\det J(k,\lambda)| = |\det[\mathrm{i}\langle k,\omega\rangle I - (A \otimes I - I \otimes A)]| > (K^{-1})^{m^2-m}(\frac{\gamma^{-1}}{|k|^\sigma})^m.$$

In order to give an estimate we now estimate the inverse of the operator $J(k,\lambda)$ for $k \neq 0$. □

Lemma 3.7 *For all parameters* λ *satisfying small denominator conditions* (3.30) *and* $k \neq 0$, *we have*

$$\|J^{-1}(k,\lambda)\| \leq cM^{m^2}\gamma^m K^{m^2}|k|^{m\sigma}, \tag{3.34}$$

where c is constant.

Proof Since for the nonsingular matrix J, $J^{-1} = \mathrm{ad}J/\det J$, where ad$J$ is the adjoint matrix of J, by the definition of $\|J\|$ and the small denominator conditions (3.30). □

In order to solve Eq. 3.33 in a complex neighborhood, we need the following lemma:

Lemma 3.8 *Under the assumption of lemma* 3.7, *there exists* $\delta' > 0$ *such that for* $\forall z \in W_{\delta'}(\Lambda)$, *we have for* $k \neq 0$

$$\|J^{-1}(k,z)\| \leq cM^{m^2}\gamma^m K^{m^2}|k|^{m\sigma}. \tag{3.35}$$

Proof We consider $z \in W_{\frac{\delta}{2}}(\Lambda)$, by Cauchy's estimate we have $|\partial_z A(z)|_{\frac{\delta}{2}} \leq M\left(\frac{\delta}{2}\right)^{-1}$. Now we firstly choose $\delta' \leq \frac{\delta}{2}$, then for $z \in W_{\delta'}(\Lambda)$, there always exists $\lambda \in \Lambda$ such that $|z-\lambda| \leq \delta'$. Thus, by Taylor's formula one has

$$|A(z) - A(\lambda)|_{\delta'} \leq M\left(\frac{\delta}{2}\right)^{-1}\delta',$$

also

$$|I_{m^2} \otimes (A(z) - A(\lambda)) - (A(z) - A(\lambda))^T \otimes I_{m^2}|_{\delta'} \leq CM\left(\frac{\delta}{2}\right)^{-1}\delta',$$

For the aim of applying the Lemma 3.5, for $k \neq 0$ we let $U = J(k,\lambda)$, $V = I_{m^2} \otimes (A(z) - A(\lambda)) - (A(z) - A(\lambda))^T \otimes I_{m^2}$. Since $|V| \leq CM(\frac{\delta}{2})^{-1}\delta'$, and by (3.34),

$$\frac{1}{2}|U^{-1}|^{-1} \geq cM^{-m^2}\gamma^{-m}K^{-m^2}|k|^{-m\sigma},$$

so if we choose $\delta' = cM^{-m^2}\gamma^{-m}K^{-m^2}|k|^{-m\sigma}\delta$, then by Lemma 3.5, for $\forall\, z \in W_{\delta'}(\Lambda),\ k \neq 0$, we have

$$\|J^{-1}(k,z)\| = |(U+V)^{-1}| \leq 2\|J^{-1}(k,\lambda)\| \leq cM^{m^2}\gamma^m K^{m^2}|k|^{m\sigma}.$$

□

The proof of Lemma 3.6 for $k \neq 0$ solving the Eq. 3.33 is equivalent to solve the following vector equation

$$J(k,z)Y'_k(z) = -F'_k(z)$$

we have

$$Y'_k(z) = -J^{-1}(k,z)F'_k(z) \tag{3.36}$$

and it is easy to see that $\|Y_k\| = \|Y'_k\|$, $\|F_k\| = \|F'_k\|$ (see Appendix A).

Since $F \in C_h^\omega(\mathbb{T}^r \times \Lambda, g)$, by the Proposition 3.1,

$$|F_k| \leq |F|_h e^{-|k|h}.$$

consequently, let $\bar{h} = \frac{h+h'}{2}$ and for any $0 < k \leq N$,

$$\begin{aligned}
|Y_k(z)|_{\delta'} &\leq |J^{-1}(k,z)|_{\delta'} \cdot |F_k(z)|_{\delta'} \\
&\leq CM^{m^2}\gamma^m K^{m^2}|k|^{m\sigma}|F|_{h,\delta'}e^{-|k|h} \\
&= CM^{m^2}\gamma^m K^{m^2}|F|_{h,\delta'}|k|^{m\sigma}e^{-|k|(h-\bar{h})}e^{-|k|\bar{h}}.
\end{aligned} \tag{3.37}$$

Since the function $x^{m\sigma}e^{-x(h-\bar{h})}$ takes its maximal at $x = \frac{m\sigma}{h-\bar{h}}$, one has

$$\begin{aligned}
|Y_k(z)|_{\delta'} &\leq CM^{m^2}\gamma^m K^{m^2}|F|_{h,\delta'}\left(\frac{m\sigma}{(h-\bar{h})e}\right)^{m\sigma} e^{-|k|\bar{h}} \\
&= C(m,\sigma,d)M^{m^2}K^{m^2}|F|_{h,\delta'}\frac{e^{-|k|\bar{h}}}{(h-\bar{h})^{m\sigma}}.
\end{aligned} \tag{3.38}$$

Let

$$Y(\varphi,z) = \sum_{0<|k|\leq N} Y_k(z)e^{\mathrm{i}(k,\varphi)}.$$

Then considering (3.38) we have

$$\begin{aligned}
|Y|_{h',\delta'} &\leq \sum_{0<|k|\leq N}|Y_k|e^{|k|h'} \\
&\leq CM^{m^2}K^{m^2}\frac{|F|_{h,\delta'}}{(h-\bar{h})^{m\sigma}}\sum_{0\neq|k|\leq N}e^{-(\bar{h}-h')|k|} \\
&\leq CM^{m^2}K^{m^2}\frac{|F|_{h,\delta'}}{(h-\bar{h})^{m\sigma}}\left(\frac{2}{\bar{h}-h'}\right)^r \\
&\leq CM^{m^2}K^{m^2}\frac{|F|_{h,\delta'}}{(h-h')^{m\sigma+r}}.
\end{aligned} \tag{3.39}$$

Let $s = m\sigma + r$, we have

$$|Y|_{h',\delta'} \le CM^{m^2}K^{m^2}\frac{|F|_{h,\delta'}}{(h-h')^s}. \tag{3.40}$$

Thus we have solved the linearized equation.

And from (3.40), using Cauchy's estimate we have

$$|L_\omega Y|_{h',\delta'} \le CM^{m^2}K^{m^2}\frac{|F|_{h,\delta'}}{(h-h')^{s+1}} \tag{3.41}$$

$$|L_\omega Y\cdot Y|_{h',\delta'} \le CM^{2m^2}K^{2m^2}\frac{|F|^2_{h,\delta'}}{(h-h')^{2s+1}} \tag{3.42}$$

$$|AY^2|_{h',\delta'}\doteq|YAY|_{h',\delta'} \le CM^{2m^2+1}K^{2m^2}\frac{|F|^2_{h,\delta'}}{(h-h')^{2s}} \tag{3.43}$$

$$|[Y,F]|_{h',\delta'} \le 2|Y|_{h',\delta'}\cdot|F|_{h,\delta'} \le CM^{m^2}K^{m^2}\frac{|F|^2_{h,\delta'}}{(h-h')^{s}} \tag{3.44}$$

and by Proposition 3.1,

$$|R_N F|_{h',\delta'} \le C\frac{e^{-N(h-h')}}{(h-h')^r}|F|_{h,\delta'}. \tag{3.45}$$

So from 3.31 it is easy to get the estimation (3.32). □

Applying Lemma 3.6 t_n times, we have the following result.

Lemma 3.9 *There is a change of variables P_n, so that (3.22) is transformed into a system*

$$\dot{X} = (A_{n+1} + F_{n+1})X$$

so that A_{n+1} and F_{n+1} satisfies

$$|F_{n+1}|_{h_{n+1},\delta_{n+1}} < \varepsilon_{n+1} \sim \varepsilon_n^{m^{4n}}, \quad |A_{n+1}|_{\delta_{n+1}} < \varepsilon_n^{-1}, \tag{3.46}$$

where $h_{n+1} = \frac{h_n}{2}$, and

$$\delta_{n+1} = C(M_n+1)^{-t_n(m^2+1)}N_n^{-t_n m\sigma}K_n^{-t_n m^2}\delta_n > Cm^{-n^4}\varepsilon_n^{\frac{n}{2m^2}}\delta_n. \tag{3.47}$$

Moreover, the transformation matrix is the composition of t_n "near-identity" transformation with form as

$$P_n = (I+Y_1)\circ(I+Y_2)\circ\cdots\circ(I+Y_{t_n})$$

and we have

$$\begin{aligned} |P_n|_{h_{n+1},\delta_{n+1}} &= |I+Y_1|_{h_{n+1},\delta_{n+1}}\cdots|I+Y_{t_n}|_{h_{n+1},\delta_{n+1}} \\ &\le \left(1+|Y_1|_{h_{n+1},\delta_{n+1}}\right)\cdot\left(1+|Y_1|^{\frac{3}{2}}_{h_{n+1},\delta_{n+1}}\right)\cdots\left(1+|Y_1|^{(\frac{3}{2})^{t_n-1}}_{h_{n+1},\delta_{n+1}}\right) \\ &\le 1+C|Y_1|_{h_{n+1},\delta_{n+1}}. \end{aligned} \tag{3.48}$$

4 Persistence of Non-degeneracy Conditions

In this section we discuss the preserving of non-degeneracy conditions corresponding to different conjugation processes in last section, respectively.

4.1 Preparation

Firstly, we introduce the Pyartli function which will be useful for discussing the non-degeneracy and measure estimating.

Definition 4.1 We say that a $C^{\varrho+1}$ function $f\colon (a,b) \to \mathbb{R}$ is (T,t,ϱ)-Pyartli, if the following two inequalities hold:

$$\max_{0\le j\le \varrho+1}\left|\frac{d^j}{dx^j}f(x)\right| \le T, \tag{4.1}$$

$$\max_{0\le j\le \varrho}\left|\frac{d^j}{dx^j}f(x)\right| \ge t > 0 \tag{4.2}$$

for $\forall\, x \in (a,b)$.

It is clear that if $T < T', t > t', \varrho < \varrho'$, then a (T,t,ϱ)-Pyartli function must be a (T',t',ϱ')-Pyartli. And for this Pyartli function, we have the following lemma which has been proven by Krikorian [12,13].

Lemma 4.1 *If f is (T,t,ϱ)-pyartli on $(a,b)(\varrho \ge 1)$ then for all $0 < \mu < t$, the set $\{\lambda|\ \lambda \in (a,b),\ |f(\lambda)| > \mu\}$ has at most $2\varrho([(b-a)T/t]+1)$ connected components and the Lebesgue measure of $\{\lambda|\ \lambda \in (a,b),\ |f(\lambda)| \le \mu\}$ is smaller than $2\varrho([(b-a)T/t]+1)2^{\varrho+2}(\frac{\mu}{t})^{\frac{1}{\varrho}}$.*

For the convenience of discussing the non-degeneracy conditions, we need the following two lemmas, we refer the reader to [5] and [13] for proof of some similar results.

Lemma 4.2 *Given n analytic functions defined on $W_\delta(\Lambda)$: $f_1,\ldots,f_n$. Suppose that each f_i is a (T_i,t_i,ϱ_i)-Pyartli, $1 \le i \le n$, let $T = \max_{1\le i\le n} T_i$, $t = \min_{1\le i\le n} t_i$, $\varrho = \varrho_1+\cdots+\varrho_n$. Then $F = f_1\cdots f_n$ is (T',t',ϱ')-Pyartli with*

$$T' = T^n,\ t' = ((\varrho T)^{-n\varrho}t)^{n^{\varrho+1}},\ \varrho' = \varrho.$$

Lemma 4.3 *Suppose all n analytic functions $f_1,\ldots,f_n$ defined on $W_\delta(\Lambda)$ satisfy $|f_i|_\delta \le T$, and the product $F = f_1\cdots f_n$ is (T,t,ϱ)-Pyartli, then each f_i is a (T,t',ϱ)-Pyartli function with $t' = (nT)^{-n\varrho}t$.*

Since for $\forall\,\lambda \in W_\delta(\Lambda)$, each $|iu+\alpha_q(\lambda)-\beta_p(\lambda)| \le 2\Delta + u$, so $|g_{ij}(\lambda,u)|$ is less than $(2\Delta+|u|)^{l_il_j}$. Thus on $W_{\frac{\delta}{2}}(\Lambda)$,

$$\max_{0\le l\le d+1}\left|\frac{\partial^l g_{ij}(\lambda,u)}{\partial\lambda^l}\right| \le (2\Delta+|u|)^{l_il_j}/(\frac{\delta}{2})^{d+1}. \tag{4.3}$$

So if the non-degeneracy conditions (2.9) hold, then for all $1 \le i, j \le s$, $g_{ij}(\lambda,u)$ is (T,t,ϱ)-Pyartli function on Λ with

$$T = (2\Delta+|u|)^{m^2}/\left(\frac{\delta}{2}\right)^{d+1},\quad t = \chi,\quad \varrho = d. \tag{4.4}$$

4.2 Non-degeneracy After Block-diagonalization

Corresponding the Lemma 3.1 we have the following result.

Lemma 4.4 *Suppose* $A \in C^{\omega}_{\delta}(\Lambda, g)$, $|A|_{\delta} \leq M$, $|\sigma(A)| < \Delta$ *on* Λ. *Also non-degeneracy conditions* (1.15) *are satisfied. Then for the block diagonal matrix* B_1 *whose eigenvalues are* $(\rho, 4m\rho)$*-separated, and on each new sub-interval* $\hat{\Lambda}_j$ *obtained in the Lemma 3.1, there exist constants* $\hat{d} \geq 1$, $\hat{\chi} > 0$, *s.t. for* $\forall\ 1 \leq i \leq j \leq \hat{s}$,

$$\max_{0 \leq l \leq \hat{d}} \left| \frac{\partial^l}{\partial \lambda^l} g_{ij}(\lambda, u) \right| > \hat{\chi},$$

uniformly hold for $u \in \mathbb{R}$, *where*

$$\hat{d} = d, \quad \hat{\chi} = (C(1 + |u|)\Delta\delta^{-1})^{-10m^5 d^2} \chi, \quad \hat{M} \leq C(\rho^{-1} M)^{2m^2}. \tag{4.5}$$

Proof We notice that under the finer decomposition,

$$\sigma(B_1) = \sigma(A_1) \bigcup \cdots \bigcup \sigma(A_s),$$

so we have

$$g(\lambda, u) = \prod_{1 \leq i, j \leq s} g_{ij}(\lambda, u),$$

where $u \in \mathbb{R}$. Since $g(\lambda, u)$ is (T, t, ϱ)-Pyartli with

$$T = (2\Delta + |u|)^{m^2} / \left(\frac{\delta}{2}\right)^{d+1}, \quad t = \chi, \quad \varrho = d,$$

so by Lemma 4.3, each g_{ij} is a $(\hat{T}, \hat{t}, \hat{\varrho})$-Pyartli with

$$\hat{T} = T, \quad \hat{t} = \hat{\chi} \geq (m^2 T)^{-m^2 d} t \geq (C(1 + |u|)\Delta\delta^{-1})^{-10m^5 d^2} \chi, \quad \hat{\varrho} = \varrho = \hat{d}. \tag{4.6}$$

Thus the non-degeneracy conditions are preserved. □

4.3 Non-degeneracy After Removing Resonance

Corresponding the Sect. 3.2, we consider the preserving of non-degeneracy after removing resonance between eigenvalues. We give the following lemma

Lemma 4.5 *Given a parameter interval* Λ, *and the system* $B_1 + G_1$, *we assume that* $B_1(\lambda) = \mathrm{diag}(A_1, \ldots, A_s)$ *is* (ρ, ν)*-separated with* $\nu = 4m\rho \leq K^{-1}$, *and the non-degeneracy conditions* (2.9) *are satisfied,* $|\sigma(B_1)| < \Delta$. *Then after all the operator in Sect. 3.2 we have* $B_1 + G_1 \equiv B + G$, *where* $|B|_{\delta'} \leq \tilde{M}$, $|\sigma(B)| < \tilde{\Delta}$, *and* $\exists\ \tilde{d} \in \mathbb{Z}^+$, $\tilde{\chi} > 0$ *such that non-degeneracy conditions* (1.15) *are preserved, i.e.*

$$\max_{0 \leq l \leq \tilde{d}} \left| \frac{\partial^l}{\partial \lambda^l} g(\lambda, u) \right| > \tilde{\chi}$$

uniformly hold for all $u \in \mathbb{R}$, $\lambda \in \Lambda$. *Moreover, we have*

$$\tilde{d} = m^2 d, \quad \tilde{\chi} = ((C(1 + |u|)\Delta d)^{-1} \delta \chi)^{d(m^2)^{d+2}}. \tag{4.7}$$

$$\tilde{M} = M + m^2 N, \quad \tilde{\Delta} = \Delta + 2mK^{-1}. \tag{4.8}$$

Proof From Lemma 3.4 we know that after removing the resonance, in general we get the matrix $B(\lambda) \in \bigcap_{ij} NR_{ij}(\tau_n^{m_0}N, K)$, $m_0 \leq m$. We assume that $\sigma(B) = \sigma(\bar{A}_1) \bigcup \cdots \bigcup \sigma(\bar{A}_s)$, by the process of translation in the Sect. 3.2 we know that for each $1 \leq i \leq s$, there exists a constant $y_i \in [0, 2\Delta]$ such that

$$\sigma(\bar{A}_i) = \sigma(A_i) + \mathrm{i}y_i.$$

So for all $i \leq j$,

$$\bar{\alpha}_q - \bar{\beta}_p = \alpha_q - \beta_p + \mathrm{i}(y_i - y_j),$$

thus for $\forall\, \bar{u} \in \mathbb{R}$,

$$\mathrm{i}\langle k, \omega\rangle + (\bar{\alpha}_q - \bar{\beta}_p) - \mathrm{i}\bar{u} = \mathrm{i}\langle k, \omega\rangle + (\alpha_q - \beta_p) - \mathrm{i}(\bar{u} + y_j - y_i).$$

Now we see that for all $u = \bar{u} + y_j - y_i \in \mathbb{R}$, we have

$$g(\sigma(\bar{A}_i), \sigma(\bar{A}_j), \bar{u}) = g(\sigma(A_i), \sigma(A_j), u),$$

so by the non-degeneracy conditions (2.9), we get

$$\max_{1\leq l\leq d} \left| \frac{\partial^l}{\partial \lambda^l} g(\sigma(\bar{A}_i), \sigma(\bar{A}_j), \bar{u}) \right| > \chi. \tag{4.9}$$

Then it is easy to see that

$$g(\lambda, u) = \prod_{1\leq i,j\leq s} g(\sigma(\bar{A}_i), \sigma(\bar{A}_j), u). \tag{4.10}$$

So by the Lemma 4.2 we know that there exist $\tilde{d} = m^2 d$, $\tilde{\chi} > 0$ such that for $\forall \lambda \in \Lambda$

$$\max_{0\leq j\leq \tilde{d}} \left| \frac{\partial^j}{\partial \lambda^j} g(\lambda, u) \right| > \tilde{\chi} \tag{4.11}$$

uniformly hold for $u \in \mathbb{R}$, where $\tilde{\chi} \geq ((C(1 + |u|)\Delta d)^{-1} \delta \chi)^{d(m^2)^{d+2}}$, since each g_{ij} is $(T = (2\Delta + |u|)^{m^2} \delta^{-(d+1)},\ t = \chi,\ \varrho = d)$-Pyartli.

It is easy to estimate

$$\tilde{M} \leq M + \min\{m^2 N, 2\Delta + 1\}, \quad \tilde{\Delta} \leq \Delta + 2mK^{-1}$$

The proof is completed. □

4.4 Non-degeneracy After Perturbation

Considering only the influence of perturbation on the non-degeneracy conditions after one step of KAM iteration, we have the following lemma which is corresponding to Lemma 3.6.

Lemma 4.6 *Given a parameter interval* Λ, *suppose* $B(\lambda) \in \bigcap_{ij} NR_{ij}(\tau_n^m N, K)$ *for all* $\lambda \in \Lambda$. *And we assume that the non-degeneracy conditions* (1.15) *are satisfied, and on* $W_\delta(\Lambda)$, $|\sigma(B(\lambda))| \leq \Delta$. *If there exists* $\varepsilon > 0$ *such that* $|F|_{h,\delta} = \varepsilon$ *with*

$$\varepsilon \leq C(M\Delta\rho^{-1})^{-2m^2},$$

then after one step of iteration for the new system $B' + G'$, *where* $B' \in \bigcap_{ij} NR_{ij}\left(\tau_n^m N, K - \frac{K}{2t_n}\right)$, *the non-degeneracy conditions is preserved, i.e.,* $\exists\ d' = d$ *and* $\chi' > 0$

such that

$$\max_{0\le l\le d'}\left|\frac{\partial^l}{\partial\lambda^l}g(\lambda,u)\right|\ge\chi', \tag{4.12}$$

uniformly hold for $u\in\mathbb{R}$, $\lambda\in\Lambda$ *with*

$$\chi'=\chi-C(M\Delta\rho^{-1})^{10m^5}\varepsilon^{\frac{2}{3}}/\left(\frac{\delta}{2}\right)^d, \tag{4.13}$$

$$\Delta'\le\Delta+CM\Delta^6\rho^{-2}\varepsilon^{\frac{2}{3m}},\quad M'\le M+\varepsilon^{\frac{2}{3}}. \tag{4.14}$$

Proof We consider the character polynomials of $\sigma(B)$ and $\sigma(B')$, then we have

$$P_{\sigma(B')}(\lambda,X)=P_{\sigma(B)}(\lambda,X)+\eta(\lambda,X). \tag{4.15}$$

When $|u|\le 2(\Delta+1)$, by a simple calculation, note that $|G|_{h,\delta}=\varepsilon^{\frac{2}{3}}$, we have

$$|\eta|_{\delta'}\le M^m\varepsilon^{\frac{2}{3}}. \tag{4.16}$$

Since $\varepsilon\le C(M\Delta\rho^{-1})^{-2m^2}$, and $|P_{\sigma(B)}|_\delta\le C\Delta^m$, so it is easy to verify that

$$|\eta|_{\delta'}\le C(\Delta^{3m+1}\rho^{-2m}(1+|P_\sigma|_\delta))^{-1}$$

then by the Appendix C, one has

$$|g(\sigma_i',\sigma_j',u)|_{\delta'}\le|g(\sigma_i,\sigma_j,u)|_\delta+C(\Delta\rho^{-1}(1+|P_\sigma|_\delta))^{10m^3}M^m\varepsilon^{\frac{2}{3}}. \tag{4.17}$$

Considering the non-degeneracy conditions (2.9), we get the estimate (4.12). (4.13) and (4.14) is a direct result from Lemma C.

When $|u|>2\Delta+1$, it is easy to see $|g(\lambda,u)|>1>\chi$, and g is a $((2\Delta+|u|)^{m^2},1,0)$-Pyartli, the non-degeneracy conditions (4.12) still hold. □

5 Iteration

Firstly, we introduce the following sequences:

$$h_n=\frac{h_1}{2^{n-1}}, \tag{5.1}$$

$$\varepsilon_n=\varepsilon_1^{(m^4)^{1+2+\cdots+(n-1)}}=\varepsilon_1^{m^{2n(n-1)}}, \tag{5.2}$$

$$N_n=\left[\frac{\tau_n-1}{12(\tau_n^m-1)}\cdot\frac{\ln\varepsilon_n^{-1}}{h_n}\right], \tag{5.3}$$

$$K_n=2^{(\sigma+2)}\gamma N_n^{2\sigma}, \tag{5.4}$$

$$\nu_n=4m\rho_n=K_n^{-1}. \tag{5.5}$$

Now we state our iterative lemma:

Iterative lemma *Given a parameter interval* Λ *and a positive real number* δ_1*, let* $A_1 \in C^{\omega}_{\delta_1}(\Lambda, g)$, $|A_1|_{\delta_1} < M_1$, $|\sigma(A_1)|_{\delta_1} \le \Delta_1$. *If for* A_1 *the non-degeneracy conditions* (1.15) *are satisfied with constant* (χ_1, d_1)*, then there exist positive real numbers* $\varepsilon_1,\ h_1$*, such that for any* $F_1 \in C^{\omega}_{h_1,\delta_1}(\mathbb{T}^r \times \Lambda, g)$*, if*

$$|F_1|_{h_1,\delta_1} = \varepsilon_1 \le C M_1^{-100m^2} h_1^{100m^2\sigma}, \tag{5.6}$$

then for any positive integer n, there exist a partition Π_n *of the parameter interval* Λ *into* $\tilde{\Lambda}_{n,j}$*, and some positive real numbers* $\delta_n,\ M_n,\ \Delta_n,\ \chi_n,\ d_n \in \mathbb{Z}^+$*, and a new system* $A_n + F_n$*, where* $A_n \in C^{\omega}_{\delta_n}(\tilde{\Lambda}_{n,j}, g)$, $F_n \in C^{\omega}_{h_n,\delta_n}(\mathbb{T}^r \times \tilde{\Lambda}_{n,j}, g)$ *satisfying the following conclusions:*

1. *For each* $\tilde{\Lambda}_{n,j} \in \Pi_n$ *and* $\forall\, \lambda \in \tilde{\Lambda}_{n,j},\ \varphi \in \mathbb{T}^r$,

$$A_1(\lambda) + F_1(\varphi, \lambda) \equiv A_n(\lambda) + F_n(\varphi, \lambda)$$

$$|A_n|_{\delta_n} \le M_n, \quad |F_n|_{h_n,\delta_n} = \varepsilon_n.$$

2. *On each* $\tilde{\Lambda}_{n,j}$*, the non-degeneracy conditions* (1.15) *are satisfied:*

$$\max_{0\le l\le d_n} \left|\frac{\partial^l}{\partial\lambda^l} g^n(\lambda, u)\right| > \chi_n,$$

uniformly for $\lambda \in \tilde{\Lambda}_{n,j},\ u \in \mathbb{R}$.

3. *For the number of all sub-intervals* $\tilde{\Lambda}_{n,j}$ *we have the estimation:*

$$\#\Pi_n \le C e^{n^3\mathcal{K}} z_n^{-n\mathcal{K}}. \tag{5.7}$$

where $\mathcal{K}$ *is a constant only depending on dimension m and*

$$z_n \gg e^{-\alpha m^{n^2}}, \quad \alpha \in (0, 1). \tag{5.8}$$

4. *We have the following relations,*

$$M_n \le \varepsilon_n^{-\frac{1}{6m^4}}, \tag{5.9}$$

$$\Delta_n \le \Delta_0 + 1, \tag{5.10}$$

$$\min\{\chi_n,\ \delta_n\} \ge z_n, \tag{5.11}$$

$$1 \le d_n \le m^{2(n-1)} d_1. \tag{5.12}$$

Proof Suppose that all conclusions hold for all positive integers from 1 to n. By induction, we should show all these conclusions still hold at the $(n+1)$th step.

The conclusion 1. is a direct result from the Lemma 3.9. In fact, if (5.6) holds, then it is easy to verify $(M_1/\rho_1)^{m(m+1)} \le \varepsilon_1^{-\frac{1}{6}}$, using the relations (5.1)–(5.5) we have for $\forall n$, $\left(\frac{M_n}{\rho_n}\right)^{m(m+1)} \le \varepsilon_n^{-\frac{1}{6}}$, so we always have $|F_n|_{h_n,\delta_n} \le \varepsilon_n$.

For 2. we apply Lemmas 4.4, 4.5 and 4.6, the non-degeneracy conditions are preserved. And we get

$$\tilde{\chi}_n = \left(C(1+|u|)\Delta_n d_n \delta_n^{-1}\right)^{-11 d_n^3 (m^2)^{d_n+5}} \chi_n^{d_n (m^2)^{d_n+2}}$$

and

$$\begin{aligned}\chi_{n+1} &\geq \tilde{\chi} - Ct_n(M_n\Delta_n\rho_n^{-1})^{11m^5}\varepsilon_n^{\frac{2}{3}} / \left(\frac{\delta_{n+1}}{2}\right)^d \\ &\geq [(CM_n\Delta_nN_nK_n)^{-1}\delta_n\chi_n]^{d_n^2\mathcal{K}^{d_n}} - [(CM_n\Delta_nN_nK_n)^{-1}\delta_n]^{\mathcal{K}}\varepsilon_n, \qquad (5.13)\end{aligned}$$

where $\mathcal{K} = \mathcal{K}(m) > 1$ is a constant only depending on the dimension of the system.

Now we consider finer division of $\sigma(A_{n+1})$, by Lemma 3.1 for each $\tilde{\Lambda}_{n+1,j}$ there exists a partition $\Pi_{\tilde{\Lambda}_{n+1,j}}$ which divides each $\tilde{\Lambda}_{n+1}$ into $\hat{\Lambda}_{n+1,j}$, such that on each $\hat{\Lambda}_{n+1,j}$ there exists a decomposition of eigenvalues: $\sigma(A_{n+1}) = \sigma_1 \bigcup \ldots \bigcup \sigma_{\hat{s}_{n+1}}$, which is (ρ_{n+1}, ν_{n+1})-separated and well-ordered, where

$$\#\hat{\Pi}_{n+1} \leq 2^{2m^2+1}(1 + |\tilde{\Lambda}'_{n,j}|)(\delta'_n(\Delta_{n+1})^{-1}\rho'_n)^{-m^2}. \qquad (5.14)$$

Moreover, consider (5.14), one has the following inductive formula for the number of new sub-intervals in $\Lambda_{\tilde{\Lambda}_{n+1,j}}$:

$$\begin{aligned}\#\Pi_{n+1} &\leq 2^{2m^2+1}(1 + |\tilde{\Lambda}_{n+1,j}|)(\delta'_n(\Delta_{n+1})^{-1}\rho'_n)^{-m^2} \cdot \#\Pi_n \\ &\leq (CM_n\Delta_nN_nK_n\delta_n^{-1}\rho_n^{-1})^{\mathcal{K}} \cdot \#\Pi_n. \qquad (5.15)\end{aligned}$$

- Estimate for Δ_n, $\#\Pi_n$.

From Sect. 4 and the iterative sequences (5.1), (5.3), (5.4) we have

$$\begin{aligned}\Delta_{n+1} &\leq \Delta_n + 2m^2((2m^4+1)K_n^{-1})^{\frac{1}{m^2}} + (C\tilde{M}_n\tilde{\Delta}_n\rho_n^{-1})^8\tilde{\varepsilon}_n^{1/m} \\ &\leq \Delta_1 + C\sum_{i=1}^n K_i^{\frac{-1}{m^2}} + \sum_{i=1}^n (C\tilde{M}_i\tilde{\Delta}_i\rho_i^{-1})^{i\mathcal{K}}\tilde{\varepsilon}_i^{1/m} \\ &\leq \Delta_1 + C\sum_{i=1}^n \left(\frac{2^{i-1}[\theta_1(1+\kappa)^i + \theta_2]}{h_1}\right)^{-\tau} \\ &\quad + \sum_{i=1}^n (C\tilde{M}_i\tilde{\Delta}_i\rho_i^{-1})^{i\mathcal{K}}(\eta e^{-\alpha(1+\kappa)^i})^{\frac{1}{m}}, \qquad (5.16)\end{aligned}$$

where the constant $\alpha = 1 - 2m^2\theta_1 > 0$. It is obvious that both last series are convergent, and if we choose h_1 and η which are sufficiently small, we can get for $\forall\, n \geq 1$

$$\Delta_n \leq \Delta_0 + 1.$$

Now, we consider the total number of all sub-intervals at each step. Recall (5.15),

$$\#\Pi_{n+1} \leq (CM_n\Delta_nN_nK_n\delta_n^{-1}\rho_n^{-1})^{\mathcal{K}} \cdot \#\Pi_n.$$

Since

$$N_n = 2^{n-1}h_1^{-1}(\theta_1(1+\kappa)^n + \theta_2), \quad K_n = CN_n^{\tau m^2},$$

$$\rho_n = \frac{1}{4m}K_n^{-1}, \;\; \Delta_n \leq \Delta_0 + 1,$$

$$\delta_{n+1} = C(M_nN_nK_n)^{-\mathcal{K}}\delta_n,$$

$$M_n \leq 2M + (n-1)m^2N_n \leq C(n-1)N_n,$$

we have

$$\begin{aligned}\#\Pi_{n+1} &\le (C(n-1)m^{2n}N_n^{3\tau m^2}\delta_n^{-1})^{\mathcal{K}}\cdot\#\Pi_n\\ &\le (C(n-1)[2m(1+\kappa)]^{3n\tau m^2}\delta_n^{-1})^{\mathcal{K}}\cdot\#\Pi_n.\end{aligned} \tag{5.17}$$

It is easy to verify that

$$\min(\rho_n, \chi_n, \delta_n) \ge z_n \gg e^{-m^{n^2}}$$

so from (5.17),

$$\#\Pi_{n+1} \le Ce^{n^3\mathcal{K}} z_n^{-n\mathcal{K}}.$$

Thus, we complete the proof of the iterative lemma. □

6 Positive Measure Result

At each step, we will use a classical positive measure reducibility result (i.e., Theorem 2), by which we can show that on any one sub-interval $\Lambda_{n,j}$ of length $L_{n,j}$, the measure of parameter set, on which the system $A_n(\lambda) + F_n(\varphi, \lambda)$ is reducible, is relatively large. The way to prove Theorem 2 is the classical KAM method, and the iterative process goes along the same line as the preceding iterative lemma but is much simpler, since for positive measure result we need n't consider the resonant parameter set and need not truncate the Fourier series, also we need n't regroup eigenvalues at each step. So in this section we only show the perturbation decreases very quickly, and estimate the measure of the removed parameter set.

Firstly, we give two iterative sequences:

$$h_n = \left(\frac{1}{2} + \frac{1}{2^n}\right) h_1, \tag{6.1}$$

$$K_n = \left(\frac{(\frac{6}{5})^n + \frac{1}{\eta}}{h_{n-1} - h_n}\right)^{\bar\gamma} = (h_1)^{-\bar\gamma} 2^{n\bar\gamma}\left((\frac{6}{5})^n + \frac{1}{\eta}\right)^{\bar\gamma} \tag{6.2}$$

where the constant $\bar\gamma \ge d$, and η will be taken in the following lemma

Lemma 6.1 *There exist positive constants $\eta < 1$, b such that, if ε_1 is sufficiently small, then for all $n \ge 1$,*

$$\begin{aligned}\varepsilon_n &\le \eta^b e^{-(\frac{6}{5})^n}.\\ M_n &\le 2^{n-1} M_1\end{aligned}$$

Proof Supposing that we have applied the precedent method until to nth step, and we have got

$$|F_n|_{h_n} \le \varepsilon_n \le \eta^b e^{-(\frac{6}{5})^n}$$

and

$$M_n \le M_{n-1} + \varepsilon_{n-1} \le 2^{n-1} M_1$$

by induction, we want to show that

$$|F_{n+1}|_{h_{n+1}} \le \eta^b e^{-(\frac{6}{5})^{n+1}} \tag{6.3}$$

$$M_{n+1} \le 2^n M_1 \tag{6.4}$$

In fact, (6.4) is satisfied since

$$M_{n+1} \le M_n + \eta^b e^{-(\frac{6}{5})^n} \le M_n + 1 \le 2M_n.$$

And from (3.40) we can easily get

$$\varepsilon_{n+1} \le CM_n^{2m^2+1} K_n^{2m^2} \frac{\varepsilon_n^2}{(h_n - h_{n+1})^{2s+1}}$$

To prove (6.3) we need

$$CM_n^{2m^2+1} K_n^{2m^2} \frac{\eta^{2b} e^{-(\frac{6}{5})^{2n}}}{(h_n - h_{n+1})^{2s+1}} \le \eta^b e^{-(\frac{6}{5})^{n+1}}$$

then using (6.1), (6.2) and (6.4), we should have

$$CM_1^{2m^2+1} h_1^{-(2s+1)} 2^{n(2m^2+1)+(n+1)(2s+1)} K_n^{2m^2} \eta^b e^{-(4/5)(\frac{6}{5})^n} \le 1 \tag{6.5}$$

let $Q_n(\eta) = K_n^{2m^2} \eta^{b-1}$, if we choose

$$b > 2m^2\bar{\gamma} + 1, \tag{6.6}$$

then by (6.2) we know that when η turns smaller, Q_n turns smaller, too. Now, we firstly take an $\eta = \eta_0 < 1$. Since the sequence

$$2^{n(2m^4+1)+(n+1)(2s+1)+nd} Q_n(\eta_0) e^{-(4/5)(\frac{6}{5})^n}$$

is bounded from above, we denote its maximum by β. In order that the (6.5) be satisfied, it is enough to choose η such that

$$Ch_1^{-(2s+1)} M_1^{2m^4+1} \beta\eta \le 1$$

so, we choose

$$\eta \le \min\{Ch_1^{2s+1} M_1^{-(2m^4+1)} \beta^{-1}, \eta_0\},$$

and (6.5) is obtained. If we take $\eta = (10\varepsilon_1)^{1/b}$, then it is enough to choose

$$\varepsilon_1 \le \min\left\{ \frac{Ch_1^{b(2s+1)} M_1^{-b(2m^4+1)}}{10\beta^b}, \eta^b e^{-\frac{6}{5}} \right\}, \tag{6.7}$$

the lemma is proved. □

From (3.40) we can see that the sequence $|Y_n|_{h_n}$ also converges to zero with "super-exponential" velocity, then by (2.1) $P_n = I + Y_n \to I$, and so the composition of transformations $P_n \circ \cdots \circ P_1$ will converge, too. On the other hand, by Sect. 7 we know that $\chi_n \ge \chi_{n-1} - CM_n\varepsilon_n$, so

$$\chi_n \ge \chi - C \sum_{1\le i\le n-1} M_i\varepsilon_i \ge \frac{\chi}{2} \tag{6.8}$$

if ε_1 is small enough. Thus, the non-degeneracy conditions are preserved. By the way, we also have the estimate for sufficiently small ε_1:

$$\rho_n > \rho - c \sum_{1\le i\le n-1} \varepsilon_i^{\frac{1}{m}} \ge \frac{\rho}{2}. \tag{6.9}$$

• *Measure of the removed set*

Now, we estimate the measure of removed parameter set. At the nth step, for $\forall\, i,\, j,\ 1 \leq i,\, j \leq s$, we denote the removed set (corresponding to resonance of $\langle k, \omega\rangle$ and A_i, A_j) by

$$R^n_{kij} = \left\{\lambda :\ |g^n_{ij}(k,\lambda)| \leq \frac{K_n^{-1}}{|k|^\tau}\right\}. \tag{6.10}$$

and let

$$R^n_k = \bigcup_{1\leq i,j\leq s} R^n_{kij}$$
$$R^n = \bigcup_{0\neq k\in\mathbb{Z}^d} R^n_k$$

in order to estimate the measure of R^n_{kij}, we recall the Lemma 4.1. Since the function g^1_{ij} by hypothesis is (T, t, ϱ)-Pyartli, then according to the above iterative process in this section, we know that for $\forall\, n \geq 1$, g^n_{ij} is $(2T, \frac{t}{2}, \varrho)$-Pyartli. Thus let L be the length of the parameter interval Λ, by Lemma 4.1 we have

$$|R^n_{kij}| \leq 2\varrho([4(b-a)T/t]+1)2^{\varrho+2}\left(\frac{2u}{t}\right)^{\frac{1}{\varrho}}$$

since $T = (6\Delta+4)^{m^2}/(\frac{\delta}{2})^{d+1},\ t = \chi,\ \varrho = d,\ u = \frac{K_n^{-1}}{|k|^\tau}$, we write

$$|R^n_{kij}| \leq Cd2^d\Delta^{m^2}\delta^{-(d+1)}\chi^{-2}L\left(\frac{K_n^{-1}}{|k|^\tau}\right)^{\frac{1}{d}} \tag{6.11}$$
$$\leq CB(d,\delta,\Delta,\chi)L\left(\frac{K_n^{-1}}{|k|^\tau}\right)^{1/d} \tag{6.12}$$

Thus,

$$|R^n| \leq Cm^2BLK_n^{-\frac{1}{d}}\sum_{0\neq k\in\mathbb{Z}^r}|k|^{\frac{-\tau}{d}}$$
$$\leq Cm^2BLK_n^{-\frac{1}{d}}r\sum_{N\geq 1}N^{r-1}N^{\frac{-\tau}{d}}$$
$$\leq C(m,r)BLK_n^{-\frac{1}{d}} \tag{6.13}$$

if $\tau > (r-3)d$, where r denotes the dimension of frequency vector ω. By (6.2), $K_n > \frac{2^{n\bar\gamma}}{\eta^{\bar\gamma}}$, so

$$K_n^{-\frac{1}{d}} \leq \eta^{\frac{\bar\gamma}{d}}\cdot\frac{1}{2^{\frac{n\bar\gamma}{d}}}.$$

Therefore, notice that $\eta = (10\varepsilon_1)^{1/b}$ and $\bar\gamma \geq d$, one has

$$\left|\bigcup_{n=1}^{\infty}R^n\right| \leq CBL\eta^{\frac{\bar\gamma}{d}}\sum_{n=1}^{\infty}2^{-\frac{n\bar\gamma}{d}} \leq CBL\eta^{\frac{\bar\gamma}{d}}$$
$$\leq C(m,r,\gamma,\sigma,\tau)B(d,\delta,\Delta,\chi)L(10\varepsilon_1)^{\frac{\bar\gamma}{bd}} \tag{6.14}$$

so the proof of the Theorem 2 is finished.

7 Proof of the Main Theorem

From the iterative lemma we know that at each step we have a partition of the parameter interval Λ into at most $\#\Pi_n$ intervals, i.e., $\Lambda_{n,j} \in \Pi_n$, such that on each $\Lambda_{n,j}$ we have a conjugation

$$\lambda A + F(\varphi) \equiv A_n(\lambda) + F_n(\varphi, \lambda).$$

Since A_n, F_n satisfy the assumptions of the classical result of positive measure reducibility, then Theorem 2 told us for each system $A_n + F_n$, there exists a subset, denoted by $R_{n,j} \subset \Lambda_{n,j}$, such that for any $\lambda \in R_{n,j}$, the system is reducible, and by (6.11) (6.14) the measure of $\Lambda_{n,j} - R_{n,j}$ is no larger than $Cd_n 2^{d_n} \Delta_n^{m^2} \delta_n^{-(d_n+1)} \chi_n^{-2} L_{n,j} \varepsilon_n^{\frac{1}{\mathcal{K}d_n}}$ if we put $\bar{\gamma}_n = d_n,\ b_n = \mathcal{K}\bar{\gamma}_n$ in the iterative process of positive measure result. By the estimations in last section, we have

$$\text{meas}(\Lambda_{n,j} - R_{n,j}) \le CL_{n,j} d_n (2z_n^{-2})^{d_n} \varepsilon_n^{\frac{1}{\mathcal{K}d_n}} \tag{7.1}$$

Therefore at nth step, the total measure of irreducible parameter set in Λ is bounded by

$$\text{meas}(\Lambda - R_n) \le \text{meas}\left(\Lambda - \bigcup_{1\le j\le \#\Pi_n} R_{n,j}\right) \le \#\Pi_n \cdot CL_{n,j} d_n (2z_n^{-2})^{d_n} \varepsilon_n^{\frac{1}{\mathcal{K}d_n}} \tag{7.2}$$

where $R_n = \bigcup_{1\le j\le \#\Pi_n} R_{n,j}$. Since

$$\varepsilon_n \le \eta \varepsilon_1^{m^{(n^2)}}, \quad d_n \le m^{2(n-1)} d_1, \quad \text{and} \quad \#\Pi_n \le Ce^{n^3\mathcal{K}} z_n^{-n\mathcal{K}},$$

it is easy to see

$$\lim_{n\to\infty} \text{meas}(\Lambda - R_n) = 0. \tag{7.3}$$

So we know that

$$\text{meas}\left(\bigcap_{n=1}^{\infty} (\Lambda - R_n)\right) = 0. \tag{7.4}$$

i.e., for almost every $\lambda \in \Lambda, \exists\ n_\lambda \ge 1$, such that $\lambda \in R_{n_\lambda}$, in other words, the system $A(\lambda) + F(\varphi, \lambda)$ is reducible.

By the way, for $a.e.\ \lambda \in \Lambda$, we only need finitely many times the transformation T_n, by which we removed the resonance, and the similarity transformation S_n, by which we reduced the matrix A_n into (ρ_n, ν_n)-separated block diagonal form. So using (5.9), for almost every $\lambda \in \Lambda$ there exists a positive integer N_λ, such that

$$M_\infty \le C\varepsilon_{N_\lambda}^{-\frac{1}{6m^4}}, \tag{7.5}$$

Thus the composition of conjugation transformations

$$P_n \circ T_n \circ S_n \circ P_{n-1} \circ T_{n-1} \circ S_{n-1} \circ \cdots \circ P_1 \circ T_1 \circ S_1$$

is convergent, since P_n converges to identity with "super-exponential" velocity.

However, since we have chosen the iterative sequence $h_n \to 0$, the composition B_λ of infinitely many times conjugation transformations can only be in $C^\infty(\mathbb{T}^r, G)$.

This completes the main theorem.

Acknowledgments The work was partially supported by National Basic Research Program of China (973 Program)(2007CB814800), NNSF of China (Grant 10531050) and the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20070284004).

Appendices

Appendix 1

In this appendix, we introduce some results in matrix theory(see [15,19] for proof).

Definition A.1 The tensor product of two matrices $A_{m\times n}$, $B_{k\times l}$ is a $mk\times nl$ matrix defined by

$$A\otimes B=\begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \vdots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$$

Remark By the definition of tensor product, it is obvious that if A is block diagonal matrix as $\mathrm{diag}(A_1, A_2, \ldots, A_s)$, the tensor product can be expressed as

$$A\otimes B=\begin{pmatrix} A_1\otimes B & & & \\ & A_2\otimes B & & \\ & & \ddots & \\ & & & A_s\otimes B \end{pmatrix}$$

Now, let $A,\ B,\ C$ be $n\times n,\ m\times m,\ n\times m$ matrices, respectively, and X be a $n\times m$ unknown matrix. Suppose the eigenvalues of $A,\ B$ are $\lambda_1,\ldots,\lambda_n$ and $\mu_1,\ldots,\mu_m$. One has the following lemma

Lemma A.1 *The eigenvalues of* $I_m\otimes A+B\otimes I_n$ *are*

$$\alpha_{ij}=\lambda_i+\mu_j,\quad i=1,\ldots,n;\ j=1,\ldots,m.$$

Especially, the eigenvalues of $I_n\otimes A-A\otimes I_n$ *are* $(\lambda_i-\lambda_j)$ *and* 0, $1\le i,j\le n, i\ne j$.

For any $n\times m$ matrix $C=(C_1,\ldots,C_m)$, where the C_j represent the entries in the j^{th} column, we denote by $C'=(C_1^T,\ldots,C_m^T)^T$ the corresponding nm-vector, and we consider the following two equations

$$AX+XB=C \tag{A.1}$$

$$(I_m\otimes A+B^T\otimes I_n)X'=C' \tag{A.2}$$

the following lemma shows that the two equations are equivalent.

Lemma A.2 *The matrix equation* (A.1) *is solvable if and only if the vector equation* (A.2) *is solvable.*

Proof Rewrite (A.1) as

$$A(X_1,\ldots,X_m)+XB=(C_1,\ldots,C_m),$$

comparing the corresponding columns in both sides, one has

$$AX_j + \sum_{i=1}^{m} X_i b_{ij} = C_j, \quad j = 1, \ldots, m.$$

Rewrite tightly these equations into a vector equation, one gets (A.2). □

From the two lemmas, one has

Corollary 1 *If $A(\lambda)$ is a one parameter $m \times m$ matrix defined on Λ, and $\alpha_j(\lambda)$, $j = 1, \ldots, m$, is the eigenvalue of A, let $J(k,\lambda) = \mathrm{i}\langle k,\omega\rangle I_{m^2} - (I_m \otimes A(\lambda) - A^T(\lambda) \otimes I_m)$, then the eigenvalues of $J(k,\lambda)$ are $\mathrm{i}\langle k,\omega\rangle - (\alpha_i(\lambda) - \alpha_j(\lambda))$, $1 \le i, j \le m$. Therefore, the matrix equation $\mathrm{i}\langle k,\omega\rangle X - (A(\lambda)X - XA(\lambda)) = C$ is solvable for all parameter $\lambda \in \Lambda$, at which $J(k,\lambda)$ is nonsingular.*

Corollary 2 *If A is a $m \times m$ quasi-diagonal matrix* $\mathrm{diag}(A_1, A_2, \ldots, A_s)$, *where each block $A_i (1 \le i \le s)$ is a $l_i \times l_i$ matrix with $\sum_{i=1}^{s} l_i = m$, then the matrix $(I_m \otimes A - A^T \otimes I_m)$ is also quasi-diagonal with form as*

$$\begin{pmatrix} I_{l_1} \otimes A - A_1^T \otimes I_m & & & \\ & I_{l_2} \otimes A - A_2^T \otimes I_m & & \\ & & \ddots & \\ & & & I_{l_s} \otimes A - A_s^T \otimes I_m \end{pmatrix}$$

where $I_{l_i} (1 \le i \le s)$ is a $l_i \times l_i$ identity matrix. Thus, $J(k,\lambda)$ is nonsingular if and only if each $ml_i \times ml_i$ block $\mathrm{i}(k,\omega)I_{ml_i} - (I_{l_i} \otimes A - A_i^T \otimes I_m)$ *is nonsingular.*

From the Lemma A.1 and last corollary, the following conclusion is obvious.

Corollary 3 *The eigenvalues of $(I_{l_i} \otimes A - A_i^T \otimes I_m)$ are just the union of eigenvalues of all the matrices $(I_{l_i} \otimes A_j - A_i^T \otimes I_{l_j})$, $j = 1, 2, \ldots, s$. So, $J(k,\lambda)$ is nonsingular if and only if for all $i, j (1 \le i, j \le s)$,* $\mathrm{i}(k,\omega)I_{l_i l_j} - (I_{l_i} \otimes A_j(\lambda) - A_i^T(\lambda) \otimes I_{l_j})$ *is nonsingular.*

Remark We denote $J_{ij}(k,\lambda) = \mathrm{i}(k,\omega)I_{l_i l_j} - (I_{l_i} \otimes A_j(\lambda) - A_i^T(\lambda) \otimes I_{l_j})$. In fact, if $s = 1$, $J_{ij}(k,\lambda) = J(k,\lambda)$; if $s = m$, $J_{ij}(k,\lambda)$ is just the eigenvalue of $J(k,\lambda)$, i.e., $\mathrm{i}(k,\omega) - (\alpha_i(\lambda) - \alpha_j(\lambda)), i \ne j$.

Remark It is easy to see that $\det[J(k,\lambda)] = \langle k, \mathrm{i}\omega\rangle^m g(\lambda, \langle k,\omega\rangle)$. In general, by the definition (1.14) we have

$$g(\lambda, u) = \det[\mathrm{i}uI_{m^2} - (I_m \otimes A(\lambda) - A^T(\lambda) \otimes I_m)]/(\mathrm{i}u)^m. \tag{A.3}$$

So it is obvious that g is analytic in λ. And for block diagonal matrix A=diag$(A_1, \ldots, A_s)$, by the definition (2.8), the analyticity of g_{ij} is also easy to verify. In fact, in next Appendix we will see that if the spectrum of $A(\lambda)$ is sufficiently separated on a parameter segment, then A can always be block diagonalized.

Appendix 2

The following lemma states that any analytic matrix function can be reduced into a block-diagonal form by an analytic similarity transformation.

Lemma B *Suppose* $A \in C_\delta^\omega(\Lambda, gl(m, \mathbb{C}))$*, and the division of spectrum* $\sigma(A) = \sigma_1 \bigcup \cdots \bigcup \sigma_s$ *is* ρ*-separated on* Λ*. Then we can find a similarity transformation* $S \in C_\delta^\omega(\Lambda, GL(m, \mathbb{C}))$*, such that*

$$A(\lambda) = S(\lambda)\mathrm{diag}(A_1(\lambda), \ldots, A_s(\lambda))S^{-1}(\lambda)$$

where $A_i \in C_\delta^\omega(\Lambda, gl(l_i, \mathbb{C}))$ *and the spectrum of* A_i *is* $\sigma_i,\ 1 \le i \le s$*. Moreover, if we choose* S *suitably such that* $|S|_\delta \le 1$*, then we have*

$$|S^{-1}|_\delta \le \max\left\{ const. \left(\frac{|A|_\delta}{\rho}\right)^{m(m+1)},\ 1\right\}. \tag{A.4}$$

where $|\cdot|_\delta = \sup_{\lambda \in W_\delta(\Lambda)} \|\cdot\|$*, and* $\|\cdot\|$ *is the operator norm.*

Proof Consider the m-dimensional vector space V. We can decompose V into s invariant subspaces $\mathbb{V}_i (1 \le i \le s)$ by choosing suitable projection $P_i (1 \le i \le s)$, i.e., $V = \mathbb{V}_1 \bigoplus \cdots \bigoplus \mathbb{V}_s$. Since the division of eigenvalues of $A(\lambda)$ is ρ-separated on Λ, we can always take s loops in the complex plane $\gamma_1(\lambda), \ldots, \gamma_s(\lambda)$, such that each loop $\gamma_i(\lambda)$ encloses all eigenvalues in σ_i and is at least ρ-separated with other loops $\gamma_j(\lambda),\ i \ne j$. Thus we can write the projection as:

$$P_i(\lambda) = -\frac{1}{2\pi \mathrm{i}} \int_{\gamma_i(\lambda)} R(\lambda, z)dz, \tag{A.5}$$

where $R(\lambda, z) = (A(\lambda) - z \cdot Id)^{-1}$ is the resolvent of A.

Using projections and choosing a suitable basis of V, we can obtain the similarity transformation $S(\lambda)$, and the analyticity is obvious. The final estimation of $|S^{-1}|$ is a conclusion by Eliasson (see [6] Appendix). □

Appendix 3

In this appendix, we give some estimations for polynomial surviving perturbation, for the proof we refer the reader to [13].

We first consider a polynomial in $\mathbb{C}[X]$ with the form as $\chi = X^n + a_1 X^{n-1} + \cdots + a_n$, and we denote the space of polynomials with degree no larger than q by $\mathbb{C}_q[X]$, which is a linear normed space with norm: $|\chi| = \sup_{1 \le i \le q} |a_i|$ (If χ analytically depends on one parameter defined in $W_\delta(\Lambda)$, we can define the norm as $|\chi|_\delta = \sup_{1 \le i \le q} |a_i|_\delta$).

Now given m complex numbers $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$, we denote their union by $\sigma = \{\alpha_1, \ldots, \alpha_m\}$. In general, we suppose σ can be divided into s subsets, i.e., $\sigma = \sigma_1 \bigcup \cdots \bigcup \sigma_s$. We say that the division of σ is ρ-separated if for all $1 \le i \ne j \le s$ and $\forall\, \alpha_p \in \sigma_i,\ \beta_p \in \sigma_j$, $|\alpha_q - \beta_p| > \rho$. Now for any $u \in \mathbb{C}$ and $1 \le i \ne j \le s$ we introduce

$$E_{ij}(u) = E(\sigma_i, \sigma_j, u) = \begin{cases} \prod_{\alpha_q \in \sigma_i, \beta_p \in \sigma_j} (\alpha_q - \beta_p - u), & i \ne j; \\ \prod_{\alpha_q, \alpha_p \in \sigma_i;\, p \ne q} (\alpha_q - \alpha_p - u), & i = j. \end{cases} \tag{A.6}$$

and especially write

$$E_{ii} = E(\sigma_i) = \prod_{\alpha_p, \alpha_q \in \sigma_i;\, p \ne q} (\alpha_p - \alpha_q). \tag{A.7}$$

We consider a polynomial $P_\sigma \in \mathbb{C}_m[X]$, whose roots are exactly $\alpha_1, \ldots, \alpha_m$, called *character polynomial* of σ, i.e., $P_\sigma(X) = \prod_{i=1}^m (X - \alpha_i)$, and a polynomial $\eta \in \mathbb{C}_m[X]$ which is very small and is regarded as a perturbation. Then we denote the set of roots of the following new polynomial by σ':

$$P_{\sigma'}(X) = P_\sigma(X) + \eta(X), \tag{A.8}$$

and we have the following estimations.

Lemma C *If there exists* $\Delta > 0$*, such that* $|\sigma| \le \Delta$*, and* $\sigma = \sigma_1 \bigcup \cdots \bigcup \sigma_s$ *is* ρ*-separated with* $0 < \rho < 1$*, and*

$$|\eta| \le C(\Delta^{3m+1}\rho^{-2m}(1 + |P_\sigma|))^{-1}$$

(*where* C *is a constant only depending on* m*), then there exists a division* $\sigma' = \sigma'_1 \bigcup \cdots \bigcup \sigma'_s$ *such that*

(i) *For all* $1 \le i \le s$,

$$|P_{\sigma'_i} - P_{\sigma_i}| \le C(1 + |P_\sigma|)^m \Delta^{4m^2} \rho^{-2m^2} |\eta|; \tag{A.9}$$

(ii) *For all* $\alpha'_p \in \sigma'_i$,

$$\mathrm{dist}(\alpha'_p, \sigma_i) \le C(1 + |P_\sigma|)\Delta^5 \rho^{-2} |\eta|^{\frac{1}{m}}; \tag{A.10}$$

(iii) $\sigma' = \sigma'_1 \bigcup \cdots \bigcup \sigma'_s$ *is* ρ'*-separated with*

$$\rho' = \rho - C(1 + |P_\sigma|)\Delta^5 \rho^{-2} |\eta|^{\frac{1}{m}}; \tag{A.11}$$

(iv) *For each* σ_i*, its character polynomial satisfies:*

$$|P_{\sigma'_i}| \le C\Delta^{3m^2} \rho^{-m^2} (1 + |P_\sigma|)^m; \tag{A.12}$$

(v) *If* $|u| \le 2\Delta + 1$*, then for all* $1 \le i, j \le s$,

$$|E(\sigma'_i) - E(\sigma_i)| \le C(\Delta\rho^{-1}(1 + |P_\sigma|))^{10m^3} |\eta|; \tag{A.13}$$

$$|E(\sigma'_i, \sigma'_j, u) - E(\sigma_i, \sigma_j, u)| \le C(\Delta\rho^{-1}(1 + |P_\sigma|))^{10m^3} |\eta|. \tag{A.14}$$

References

1. Bogoljubov, N.N., Mitropolski, Ju.A., Samoilenko, A.M.: Methods of Accelerated Convergence in Nonlinear Mechanics. Springer-Verlag, New York (1976)
2. Coppel, W.A.: Pseudo-autonomous linear equation. Bull. Austral. Math. Soc. **16**, 61–65 (1977)
3. Dinaburg, E.I., Sinai, Y.G.: The one dimensional Schrödinger equation with a quasi-periodic potential. Funkt. Anal. i. Priloz **9**, 8–21 (1975)
4. Eliasson, L.H.: Floquet solutions for the one-dimensional quasi-periodic Schrödinger equation. Commun. Math. Phys. **146**, 447–482 (1992)
5. Eliasson, L.H.: Discrete one-dimensional quasi-periodic Schrödinger operators with pure point spectrum. Acta Math. **179**, 153–196 (1997)
6. Eliasson, L.H.: Almost reducibility of linear quasi-periodic systems. In: Smooth Ergodic Theory and its Applications (Seattle, WA, 1999). Proc. Sympos. Pure Math., vol. 69, pp. 679–705. Amer. Math. Soc., Providence, RI (2001)
7. He, H.L., You, J.G.: An improved result for positive measure reducibility of quasi-periodic linear systems. Acta Math. Sin. (Engl. Ser.) **22**(1), 77–86 (2006)
8. Johnson, R.A., Moser, J.: The rotation number for almost periodic potentials. Commun. Math. Phys. **84**, 403–438 (1982)

9. Johnson, R.A., Sell, G.R.: Smoothness of spectral subbundles and reducibility of quasi-periodic linear differential systems. J. Diff. Eqn. **41**, 262–288 (1981)
10. Jorba, A., Simó, C.: On the reducibility of linear differntial equations with quasi-periodic coefficients. J. Diff. Eq. **98**, 111–124 (1992)
11. Kato, T.: Perturbation Theory for Linear Operator. Springer-Verlag, New York Inc (1966)
12. Krikorian, R.: Réductibilité des systèmes produits-croisés à valeurs dans des groupes compacts (French) [Reducibility of compact-group-valued skew-product systems]. Astérisque **259**, 1–216 (1999)
13. Krikorian, R.: Réductibilité presque partout des flots fibrés quasi-périodiques á valeurs dans des groups compacts. Ann. Sci. École. Norm. Sup. **32**, 187–240 (1999)
14. Krikorian, R.: Global density of reducible quasi-periodic cocycles on $\mathbb{T}^1 \times SU(2)$. Ann. Math. **154**, 269–326 (2001)
15. Lancaster, P.: Theory of Matrices. Academic Press, NewYork and London (1969)
16. Moser, J., Pöschel, J.: An extension of a result by Dinaburg and Sinai on quasi-periodic potentials. Comment. Math. Helvetici. **59**, 39–85 (1984)
17. Rüssmann, H.: On the one dimensional Schrödinger equation with a quasi-periodic potential. Annal. NY Acad. Sci. **357**, 90–107 (1980)
18. Rychlik, M.: Renormalization of cocycles and linaer ODE with almost-periodic coefficients. Invent. Math. **110**(1), 173–206 (1992)
19. You, J.: Perturbations of lower dimensional tori for Hamiltonian systems. J. Diff. Eq. **152**, 1–29 (1999)