The KAM theorem with a large perturbation and application to the network of Duffing oscillators

Xiaoping Yuan\textsuperscript{1,2,*}, Lu Chen\textsuperscript{2} & Jing Li\textsuperscript{1}

\textsuperscript{1}School of Mathematics and Statistics, Shandong University, Weihai 264209, China; \textsuperscript{2}School of Mathematical Sciences, Fudan University, Shanghai 200433, China
Email: xpyuan@fudan.edu.cn, chenlu@zju.edu.cn, xlijing@sdu.edu.cn

Received March 13, 2022; accepted July 7, 2022; published online November 16, 2022

Abstract  We prove that there is an invariant torus with the given Diophantine frequency vector for a class of Hamiltonian systems defined by an integrable large Hamiltonian function with a large non-autonomous Hamiltonian perturbation. As for application, we prove that a finite network of Duffing oscillators with periodic external forces possesses Lagrange stability for almost all initial data.

Keywords  KAM theorem, Hamiltonian system, invariant torus, Lagrange stability

MSC(2020)  34L15, 34B10, 47E05

Citation:  Yuan X P, Chen L, Li J. The KAM theorem with a large perturbation and application to the network of Duffing oscillators. Sci China Math, 2023, 66: 457–474, https://doi.org/10.1007/s11425-022-2001-4

1 Introduction

It is well known that the forced harmonic oscillator
\begin{equation}
\frac{d^2 x}{dt^2} + \alpha x = k \sin \omega t, \quad \alpha > 0
\end{equation}
has a periodic solution
\[ x(t) = \frac{k}{\alpha - \omega^2} \sin \omega t, \quad \alpha \neq \omega^2. \]

When $\alpha = \omega^2$, resonance happens. Duffing [5] introduced a nonlinear oscillator with a cubic stiffness term to describe the hardening spring effect observed in many mechanical problems
\begin{equation}
\frac{d^2 x}{dt^2} + \alpha x + \beta x^2 + \gamma x^3 = k \sin \omega t, \quad \alpha > 0, \quad \gamma < 0.
\end{equation}

Moon and Holmes [7,16] showed that the equation of the form
\begin{equation}
\frac{d^2 x}{dt^2} + \beta \frac{dx}{dt} - x + x^3 = k \cos \omega t
\end{equation}
The key ingredient in (1.4) is the Moser’s twist theorem (see [18] for more details) which shows the stability of (1.7) with ω = 2π. In order to see this point, let us recall the method used by Dieckerhoff and Zehnder [4] for the case (1.7) with p(t) being of period 2π. An important observation by Dieckerhoff and Zehnder [4] is that the perturbation p_j(t) is smooth enough and are of period, say, 2π (see [11–15, 22, 23] for more details).

In many research fields such as physics, mechanics and mathematical biology and so on arise networks of coupled Duffing oscillators of various forms. For example, the evolution equations for the voltage variables V_1 and V_2 obtained by the Kirchhoff’s voltage law are

\[ R^2C_0\frac{d^2V_1}{dt^2} = -\left(\frac{R^2C}{R_1}\right)\frac{dV_1}{dt} - \left(\frac{R}{R_2}\right)V_1 - \left(\frac{R}{100R_3}\right)V_1^3 + \left(\frac{R}{RC}\right)V_2 + f\sin\omega t, \]

\[ R^2C_0\frac{d^2V_2}{dt^2} = -\left(\frac{R^2C}{R_1}\right)\frac{dV_2}{dt} - \left(\frac{R}{R_2}\right)V_2 - \left(\frac{R}{100R_3}\right)V_2^3 + \left(\frac{R}{RC}\right)V_1, \]

where R, R_1, R_2, R_3, RC and C are the resistors and the capacitor, respectively. This equation can be regarded as one coupled by two Duffing oscillators (see [1–3, 8–10, 20, 21] for more details).

In the present paper, we are concerned with the Lagrange stability for the coupled Hamiltonian system of m Duffing oscillators:

\[ \ddot{x}_i + \sum_{j=0}^{n} p_j(t)x^j = 0, \quad i = 1, 2, \ldots, m, \]

where the polynomial potential

\[ F = F(x, t) = \sum_{\alpha \in \mathbb{N}^m, \|\alpha\| \leq 2n+1} p_\alpha(t)x^\alpha, \quad x \in \mathbb{R}^m \]

with p_\alpha(t) being of period 2π, and n is a given natural number. When m = 1, (1.7) is reduced to (1.5). Note that (1.7) is a Hamiltonian system with the degree of freedom m + 1/2 where time takes 1/2 as the degree of freedom. Generally speaking, the study of the stability of (1.7) with m > 1 is a difficult task. In order to see this point, let us recall the method used by Dieckerhoff and Zehnder [4] for the case m = 1. The key ingredient in [4] is the Moser’s twist theorem (see [18] for more details) which shows that the twist mapping

\[ P_\varepsilon: \begin{cases} \rho = \rho + \varepsilon\rho(\rho, \theta), & \rho \in [a, b], \\ \theta = \theta + \alpha(\rho) + \varepsilon f(\rho, \theta), & \theta \in \mathbb{T} = \mathbb{R}^{2\pi} / 2\pi \end{cases} \]

possesses an invariant curve with the rotational frequency ω = α(ρ_0) ∈ D.C. (Diophantine conditions) provided that \|f\|_{C^{333}}, \|g\|_{C^{333}} ≤ 1, 0 < ε ≪ 1 and P_\varepsilon is symplectic or intersection-preserving. (C^{333} can be relaxed to C^{3+μ} with μ > 0 [19].) Observe that the mapping P_\varepsilon is an integrable one attached by a small perturbation ε(f, g). Although \ddot{x} + \sum_{j=0}^{n} p_j(t)x^j is not small at all in (1.5). An important observation by Dieckerhoff and Zehnder [4] is that the perturbation \sum_{j=0}^{n} p_j(t)x^j is relatively small with respect to the integrable part x^2n+1 in the neighborhood of the infinity. More exactly, one can write (1.5) as a Hamiltonian equation with the Hamiltonian H:

\[ H = H_0(I) + R(I, \theta, t), \quad (I, \theta, t) \in \mathbb{R} \times \mathbb{T} \times \mathbb{T}, \]
where \((I, \theta)\) are action-angle variables, and
\[
H_0(I) \sim I^{\frac{2m+1}{m+2}}, \quad R(I, \theta, t) \sim I^{\frac{2m+1}{m+2}}.
\] (1.9)

Note
\[
\lim_{I \to \infty} (H_0(I))^{-1} R(I, \theta, t) = 0.
\] (1.10)

If the integrable Hamiltonian \(H_0\) and the perturbation \(R\) obey (1.10), we call that \(R\) is relatively small with respect to \(H_0\) in the neighborhood of the infinity. Then by a series of symplectic transformations, the relatively small \(R\) can be changed into a truly small perturbation. The symplectic transformations can be implicitly defined by
\[
\Psi: \begin{cases} 
I = \mu + \frac{\partial S}{\partial \theta}, \\
\theta = \theta + \frac{\partial S}{\partial \mu},
\end{cases}
\] (1.11)

where \(S = S(\mu, \theta, t)\) is the generating function which obeys the homological equation
\[
H_0'(\mu) \frac{\partial S}{\partial \theta} + R(\mu, \theta, t) = [R](\mu, t),
\] (1.12)

where \([R](\mu, t) = \frac{1}{2\pi} \int_0^{2\pi} R(\mu, \theta, t) d\theta\). When \(m = 1\), the homological equation is a scalar equation. Thus,
\[
S = \int_0^\theta \frac{[R](\mu, t) - R(\mu, \theta, t)}{H_0'(\mu)} d\theta.
\] (1.13)

By (1.9), we have
\[
S \sim I^{\frac{n+1}{2n+2}}.
\] (1.14)

Now the perturbation \(R\) is changed into \(R_1\):
\[
R_1 = \frac{\partial S}{\partial \theta} + \int_0^1 (1 - \tau)H_0''(\mu + \tau \frac{\partial S}{\partial \theta}) (\frac{\partial S}{\partial \theta})^2 d\tau \sim O(I^{\frac{n+1}{2n+2}}) + O(I^{\frac{n+1}{2n+2}}) = O(I^{\frac{n+1}{2n+2}}).
\] (1.15)

Repeat the procedure as above \(\nu\) times with \(\nu > 1 + \frac{4}{n} + \log_2 n\). Then \(R\) is changed into
\[
R_\nu = O(I^{\frac{n+1}{2n+2}}), \quad c > 0.
\] (1.16)

Now \(R_\nu\) is small when \(|I| \to +\infty\). By Moser’s twist theorem, one can prove that there are many invariant cylinders of time period \(2\pi\) around \(\infty\) in the extended phase \((x, \dot{x})\). The solution \((x(t), \dot{x}(t))\) with the initial data \((x(0), \dot{x}(0))\) in the cylinder is confined in the cylinder. Thus, the Duffing oscillator (1.5) has the Lagrange stability
\[
\sup_{t \in \mathbb{R}} (|x(t)| + |\dot{x}(t)|) < \infty.
\] (1.17)

When \(m \geq 2\), the equation (1.13) does not hold any more. In order to see this clearly, passing (1.12) to Fourier coefficients of \(S\) and \(R\), we have
\[
\hat{S}(\mu, k, t) = \frac{i}{\langle \partial_{\mu} H_0(\mu), k \rangle} \hat{R}(\mu, k, t), \quad k \in \mathbb{Z}^m \setminus \{0\}, \quad i^2 = -1,
\] (1.18)

where \(\hat{S}(\mu, k, t)\) and \(\hat{R}(\mu, k, t)\) are Fourier coefficients of \(S(\mu, \cdot, t)\) and \(R(\mu, \cdot, t)\), respectively. Note that \(0\) is in the closure of the set \(\{ \langle \partial_{\mu} H_0(\mu), k \rangle : k \in \mathbb{Z}^m \setminus \{0\} \}\), i.e., \(\langle \partial_{\mu} H_0(\mu), k \rangle\)'s are the notorious small divisors. Thus the estimate (1.14) does not hold when \(m \geq 2\).

In the present paper, we directly construct a KAM theorem to deal with the Hamiltonian system with a large perturbation.

In order to state the theorem, we need some notation. Denote by \(C\) (or \(c\)) a universal positive constant which may be different in different places. When those constants \(C\) and \(c\) are necessarily distinguished, denote them by \(C_0, C_1, c_0, c_1\), etc. Let the positive integer \(d\) be the degree of freedom of the Hamiltonian to be considered. Let \([1, 2]^d\) be the product of \(d\) intervals \([1, 2]\) and \(\mathbb{T}^{d+1} = \mathbb{R}^{d+1}/(2\pi \mathbb{Z})^{d+1}\). Denote by \(\varepsilon > 0\) a small constant which measures the size of Hamiltonian functions.
Theorem 1.1. Consider a Hamiltonian $H = \varepsilon^{-a}H_0(I) + \varepsilon^{-b}R(\theta, t, I)$, where $a$ and $b$ are given positive constants with $a > b$, and $H_0$ and $R$ obey the following conditions:

1. $H_0$ and $R(\theta, t, I)$ are real analytic functions in the domain $T^{d+1} \times [1, 2]^d$ and

$$\|H_0\| := \sup_{I \in [1, 2]^d} |H_0(I)| \leq c_1, \quad \|R\| := \sup_{(\theta, t, I) \in T^{d+1} \times [1, 2]^d} |R(\theta, t, I)| \leq c_2.$$

2. $H_0$ is non-degenerate in the Kolmogorov sense:

$$\det(\partial^2_{\theta I} H_0(I)) \geq c_3 > 0, \quad \forall I \in [1, 2]^d.$$

Then there exists an $\varepsilon^* = \varepsilon^*(a, b, d, c_1, c_2, c_3) \ll 1$ such that for any $\varepsilon$ with $|\varepsilon| < \varepsilon^*$, the Hamiltonian system

$$\dot{\theta} = \frac{\partial H(\theta, t, I)}{\partial I}, \quad \dot{I} = -\frac{\partial H(\theta, t, I)}{\partial \theta}$$

possesses a $(d + 1)$-dimensional invariant torus of the rotational frequency vector $(\omega(I_0), 2\pi)$ with $\omega(I) := \frac{\partial H_0(I)}{\partial I}$ for any $I_0 \in [1, 2]^d$ and $\omega(I_0)$ obeying Diophantine conditions:

$$|\varepsilon^{-a}(k, \omega(I_0)) + l| \geq \frac{\varepsilon^{-a} \gamma}{|k|^{d+1}}, \quad \gamma = \left(\log \frac{1}{\varepsilon}\right)^{-\frac{1}{2}c_1}, \quad k \in \mathbb{Z}^d \setminus \{0\},$$

$$l \in \mathbb{Z}, \quad |k| + |l| \leq \left(\log \frac{1}{\varepsilon}\right)^{c_1}$$

and

$$|\varepsilon^{-a}(k, \omega(I_0)) + l| \geq \frac{\gamma}{(1 + |k|)^{d+1}}, \quad (k, l) \in \mathbb{Z}^d \times \mathbb{Z}, \quad |k| + |l| > \left(\log \frac{1}{\varepsilon}\right)^{c_1}.$$

Remark 1.2. We point out that Theorem 1.1 is trivial if the perturbation does not depend on time $t$. However, Theorem 1.1 is not at all trivial when $R$ involves time $t$. This can be seen in the following way. By time rescaling $t = \varepsilon^a \tau$, and introducing the new angle-action variable $(\phi, J) = (\varepsilon^a \tau, J)$, we get an autonomous Hamiltonian

$$H_{\text{new}} = \tilde{H}_0(J, I) + \varepsilon^{-b}R(\theta, \phi, I), \quad \text{where} \quad \tilde{H}_0(J, I) = \varepsilon^a J + H_0(I) \text{ and new time} = \tau.$$

It seems that the invariant tori could be obtained by direct application of Moser’s classical twist theorem. Unfortunately, the unperturbed Hamiltonian $\tilde{H}_0(J, I)$ does not satisfy the twist condition of Moser’s theorem (or Kolmogorov’s non-degenerate condition). In fact, the small divisors for $H_{\text{new}}$ are

$$(*) := \langle k, \omega(I) \rangle + \varepsilon^a l, \quad (k, l) \in \mathbb{Z}^{d+1} \setminus \{0\}.$$

At this time, the divisor $(*)$ is too small for the Newton iteration to work in the proof of the KAM theory. For example, taking $k = 0$ and $l = 1$, we have

$$(*) \ll \varepsilon^a \ll \text{the size of the perturbation} \varepsilon^{-b}R(\theta, \phi, I).$$

This implies that the solution of the homological equation is so large that the symplectic transformation is not well defined in the Newton iteration.

Applying Theorem 1.1 to (1.7), we have the following theorem.

Theorem 1.3. The equation (1.7) is almost Lagrange stable, i.e., for almost all the initial data $(x_j(0), \dot{x}_j(0) : j = 1, \ldots, m)$, the solution $(x_j(t), \dot{x}_j(t) : j = 1, \ldots, m)$ exists globally and obeys

$$\sup_{t \in \mathbb{R}} \sum_{i=1}^{m} (|x_i(t)| + |\dot{x}_i(t)|) < C,$$  \quad (1.19)$$

where $C$ depends on the initial data $(x_j(0), \dot{x}_j(0) : j = 1, \ldots, m)$. 

Remark 1.4. Actually, we can prove that the solutions to (1.7) are time quasi-periodic for almost all large initial data (see Section 5).

Remark 1.5. Theorems 1.1 and 1.3 hold true for the reversible systems, respectively. We do not pursue this end here.

Remark 1.6. Let \( \Theta = \{ I_0 \in [1, 2]^d : \omega(I_0) \) obeys the Diophantine conditions\}. We claim that the Lebesgue measure of \( \Theta \) approaches 1:

\[
\text{Leb} \Theta \geq 1 - C \left( \log \frac{1}{\varepsilon} \right)^{-C} \to 1, \quad \text{as } \varepsilon \to 0.
\]

Let

\[
\tilde{\Theta}_{k,l} = \left\{ \xi \in \omega([1, 2]^d) : |\varepsilon^{-a}(k, \xi) + l| \leq \frac{\varepsilon^{-a\gamma}}{|k|^d + 1}, \quad 0 \neq k \in \mathbb{Z}^d, \quad l \in \mathbb{Z} \text{, } |k| + |l| \leq \left( \log \frac{1}{\varepsilon} \right)^{C_1} \right\}
\]

and

\[
\tilde{\Theta}_{k,l} = \left\{ \xi \in \omega([1, 2]^d) : |\varepsilon^{-a}(k, \xi) + l| \leq \frac{\gamma}{(1 + |k|)^d + 1}, \quad (k, l) \in \mathbb{Z}^d \times \mathbb{Z} \text{, } |k| + |l| > \left( \log \frac{1}{\varepsilon} \right)^{C_1} \right\}.
\]

Note that there exists a direction \( k(\xi) \) such that

\[
\partial_{k(\xi)}(\varepsilon^{-a}(k, \xi) + l) = \varepsilon^{-a}|k|, \quad k = (k_1, \ldots, k_d) \in \mathbb{Z}^d \setminus \{0\}.
\]

Thus,

\[
\text{Leb} \tilde{\Theta}_{k,l} \leq C \gamma / |k|^{d+2}, \quad k \neq 0, \quad |k| + |l| \leq \left( \log \frac{1}{\varepsilon} \right)^{C_1},
\]

\[
\text{Leb} \tilde{\Theta}_{k,l} \leq C \gamma / |k|^{d+2}, \quad k \neq 0, \quad |k| + |l| > \left( \log \frac{1}{\varepsilon} \right)^{C_1}.
\]

Also note that \( \tilde{\Theta}_{k,l} = \emptyset \) when \( k = 0, l \neq 0 \) and when \( |l| > 1 + \varepsilon^{-a}|k| \sup_{I \in [1, 2]^d} |\omega(I)| \). Therefore,

\[
\text{Leb} \left( \bigcup_{(0,0) \neq (k,l) \in \mathbb{Z}^{d+1}} \tilde{\Theta}_{k,l} \right) \leq C \gamma \left( \log \frac{1}{\varepsilon} \right)^{C_1} \leq C \left( \log \frac{1}{\varepsilon} \right)^{-C_1}.
\]

Let \( \Theta = [1, 2]^d \setminus \bigcup_{(0,0) \neq (k,l) \in \mathbb{Z}^{d+1}} \omega^{-1}(\tilde{\Theta}_{k,l}) \). By Kolmogorov’s non-degenerate condition, the map \( \omega : [1, 2]^d \to \omega([1, 2]^n) \) is a diffeomorphism in both directions. Then the proof of the claim is completed by letting \( \Theta = [1, 2]^d \setminus \bigcup_{(k,l) \in \mathbb{Z}^{d+1} \setminus (0,0)} \omega^{-1}(\tilde{\Theta}_{k,l}) \).

Remark 1.7. In 2001, a theorem similar to Theorem 1.3 was given in [24]. The basic method used in [24] comes from [4], where one Duffing oscillator \((m = 1)\) is considered. In terms of notation in [24], let \( H_0(I) = h_1(\lambda) \) and \( m = n \). The proof of [24] depends heavily on the inequalities

\[
|\langle b_1(\lambda), k \rangle| \geq |b_1(\lambda)| |\delta|, \quad \forall \ k \in \mathbb{Z}^n \setminus \{0\}, \quad (1.20)
\]

where \( \delta \) is a positive constant (see [24, p.637]). This inequality (1.20) holds true for \( n = 1 \), which is used to transform the large perturbation into a small one in [4]. When \( n > 1 \), unfortunately, the inequality (1.20) does not hold true for all \( k \in \mathbb{Z}^n \setminus \{0\} \).

Remark 1.8. In Theorem 1.3, we prove that for “almost all” initial data, the solutions to (1.7) are stable for all time \( t \in \mathbb{R} \). We are glad to mention the interesting paper [6], where Giorgilli and Zehnder proved that for all initial data, the solutions are stable for “almost all” time (an exponentially long interval of time).
2 The normal form

By the compactness of $T^{d+1} \times [1, 2]^d$, there is a constant $s_0 > 0$ such that the Hamiltonian functions $H_0(I)$ and $R(\theta, t, I)$ are analytic in the complex neighborhood $T^{d+1} \times [1, 2]^d_{2s_0}$ of $T^{d+1} \times [1, 2]^d$ and are real for the real arguments, where

$$T^{d+1} \times [1, 2]^d_{2s_0} = \{ \phi \in \mathbb{C}^{d+1}/(2\pi \mathbb{Z})^{d+1} : |\Im \phi| \leq 2s_0 \} \times \{ z \in \mathbb{C}^d : \text{dist}_\mathbb{C}(z, [1, 2]^d) \leq 2s_0 \}.$$

We call that a function of complex variables is real analytic if it is analytic on the complex variables and real for the real arguments. Thus, $H_0$ and $R$ are real analytic and we can assume

$$\| \partial_{\phi}^a \partial_{\theta}^b H_0(I) \|_{2s_0} = \sup_{I \in [1, 2]^d_{2s_0}} |\partial_{\phi}^a \partial_{\theta}^b H_0(I)| \leq C,$$

$$\| \partial_{\phi}^a \partial_{\theta}^b R(\phi, I) \|_{2s_0} = \sup_{(\phi, I) \in T^{d+1} \times [1, 2]^d_{2s_0}} |\partial_{\phi}^a \partial_{\theta}^b R(\phi, I)| \leq C, \quad |\alpha| + |\beta| \leq 2. \quad (2.1)$$

Now take $I_0 \in [1, 2]^d$ such that $\omega(I_0) = \partial_I H_0(I_0)$ obeys D.C.:

$$|\varepsilon^{-a} \langle k, \omega(I_0) \rangle + l| \geq \frac{\varepsilon^{-a} \gamma}{|k|^{d+1}}, \quad |k| + |l| \leq \left( \frac{\log \frac{1}{\varepsilon}}{C_1} \right), \quad k \in \mathbb{Z}^d \setminus \{0\}, \quad l \in \mathbb{Z}. \quad (2.2)$$

Take a constant $C_2$ with $C_2 > 5(d + 1)C_1$ and define a neighborhood of $I_0$:

$$B_0 = \left\{ I = I_0 + z : z \in \mathbb{C}^d, |z| \leq \left( \frac{\log \frac{1}{\varepsilon}}{C_1} \right)^{-C_2} \right\}.$$

By (2.1) and (2.2), we see that for $|k| + |l| \leq (\log \frac{1}{\varepsilon})^{C_1}$, $k \in \mathbb{Z}^d \setminus \{0\}$, $l \in \mathbb{Z}$ and $I \in B_0$, the frequency $\omega(I)$ obeys

$$|\varepsilon^{-a} \langle k, \omega(I) \rangle + l| \geq |\varepsilon^{-a} \langle k, \omega(I_0) \rangle + l| - \varepsilon^{-a} |\omega(I) - \omega(I_0)|| |k| \geq \frac{\varepsilon^{-a} \gamma}{2|k|^{d+1}}. \quad (2.3)$$

Let $\Omega_0 = \frac{1}{2} \partial_{\phi}^2 H_0(I_0)$. By Kolmogorov’s non-degeneracy, we see that $\Omega_0$ is the invertible matrix and

$$|\Omega_0| \leq C, \quad |\Omega_0^{-1}| \leq C. \quad (2.4)$$

Introduce a truncation operator $\Gamma = \Gamma_K$ depending on $K > 0$ as follows: for any function $f : T^{d+1} \to \mathbb{C}$ (or $\mathbb{C}^l$, $l \geq 1$), write $f(x) = \sum_{k \in \mathbb{Z}^{d+1}} \hat{f}(k)e^{i(k,x)}$ and define the function $(\Gamma_K f)(x)$:

$$(\Gamma_K f)(x) = \sum_{|k| \leq K} \hat{f}(k)e^{i(k,x)}.$$

Now consider $H(\theta, t, I) = \varepsilon^{-a} H_0(I) + \varepsilon^{-b} R(\theta, t, I)$, where $(\theta, t) \in T^{d+1}_{2s_0}$ and $I \in B_0$. Then $H_0(I)$ and $R(\theta, t, I)$ are real analytic on $T^{d+1}_{2s_0} \times B_0$. Take $x = (\theta, t)$. Decompose $R(\theta, t, I) = \Gamma_K R + (1 - \Gamma_K)R$, where

$$(\Gamma_K R)(x, I) = \sum_{|k| + |l| \leq K} \hat{R}(k,l) e^{i(k,\theta) + l t}, \quad (k, l) \in \mathbb{Z}^d \times \mathbb{Z}.$$

Let $A = 200d(a + b)$ and $m_0 = 2 + \lfloor \frac{\log A}{\log 2} \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part of a positive number. Define sequences

- $s_j = 2^{-j} s_0$, $s_j = s_j - \frac{1}{10}(s_j - s_{j+1})$, $l = 0, 1, \ldots, 10$, $j = 0, 1, 2, \ldots, m_0$;
- $K_j = \frac{A}{s_j s_{j+1}} \log 2$, $j = 0, 1, 2, \ldots, m_0$;
- $\tau_j = (\log \frac{1}{\varepsilon})^{-C_2}, \tau_j = 2^{-j} \tau_0$, $\tau_j = \tau_j - \frac{1}{10}(\tau_j - \tau_{j+1})$, $l = 0, 1, \ldots, 10$, $j = 0, 1, \ldots, m_0$;
- $B(\tau_0) = B_0$, $B(\tau_j) = \{ z \in \mathbb{C}^d : |z - I_0| \leq \tau_j \}, \quad j = 0, 1, \ldots, m_0$.
Let $C_1 > 1$. We have
\[ K_j \leq K_{m_0} < \left( \log \frac{1}{\epsilon} \right)^{C_1}, \quad j = 0, 1, 2, \ldots, m_0. \] (2.5)

For a function $f$ defined in $T^{d+1}_s \times B(\tau)$ or $B(\tau)$, define
\[ \|f\|_{s, \tau} = \sup_{(\phi, I) \in T^{d+1}_s \times B(\tau)} |f(\phi, I)| \quad \text{or} \quad \|f\|_{\tau} = \sup_{I \in B(\tau)} |f(I)|. \]

By Cauchy’s formula, we have
\[ \|(1 - \Gamma_{K_0})R\|_{s_0, \tau_0} = \left\| \sum_{|k| + |l| > K_0} \tilde{R}(k, l, I) e^{i((k, \theta)+l, t)} \right\|_{s_0, \tau_0} \]
\[ \leq \left( \sum_{|k| + |l| > K_0} e^{-2(s_0 - \tau_0)(|k| + |l|)} \right) \|R\|_{2s_0, 2\tau_0} \leq \frac{C}{(s_0)^{d+1}} \epsilon^{-s_0 K_0} \leq C \epsilon^{A} \]
and
\[ \|\Gamma_{K_0} R\|_{s_0, \tau_0} \leq C. \]

Let
\[ H^{(0)} = H, \quad R^{(0)} = \epsilon^{-b} \Gamma_{K_0} R(\theta, t, I), \quad R^{(0)}_+ = \epsilon^{-b} (1 - \Gamma_{K_0}) R(\theta, t, I). \]

Then
\[ H^{(0)} = \epsilon^{-a} H_0(I) + R^{(0)}(\theta, t, I) + R^{(0)}_+(\theta, t, I), \quad (\theta, t, I) \in T^{d+1}_s \times B(\tau_0), \] (2.6)
where $R^{(0)}$ and $R^{(0)}_+$ are real analytic in $T^{d+1}_s \times B(\tau_0)$, and
\[ \|R^{(0)}\|_{s_0, \tau_0} \leq C \epsilon^{-b}, \quad \Gamma_{K_0} R^{(0)} \equiv R^{(0)}, \] (2.7)
\[ \|R^{(0)}_+\|_{s_0, \tau_0} \leq C \epsilon^{A - b}. \] (2.8)

For the simplify of notations, the universal constant $C$ in the following lemma may depend on $m_0$, $C_1$ and $C_2$, and thus on $a$, $b$, $A$, $C_1$ and $C_2$.

**Lemma 2.1** (Iterative lemma in finite steps). Assume that we have a Hamiltonian
\[ H^{(j)} = \epsilon^{-a} H_0(I) + h^{(j)}(t, I) + R^{(j)}(\theta, t, I) + R^{(j)}_+(\theta, t, I), \quad (\theta, t, I) \in T^{d+1}_s \times B(\tau_j), \quad 0 \leq j \leq m_0 - 1, \] (2.9)
where
1. the functions $h^{(j)}(t, I)$, $R^{(j)}(\theta, t, I)$ and $R^{(j)}_+(\theta, t, I)$ are real analytic in $T^{d+1}_s \times B(\tau_j)$;
2. $h^{(j)} \equiv 0$, $\|h^{(j)}\|_{s_j, \tau_j} \leq C \epsilon^{-b}$, $0 \leq j \leq m_0 - 1$;
3. $\|R^{(j)}\|_{s_j, \tau_j} \leq C \epsilon^{-b + j(a - b)} \left( \log \frac{1}{\epsilon} \right)^{C}$, $\Gamma_{K_j} R^{(j)} = R^{(j)}$;
4. $\|R^{(j)}_+\|_{s_j, \tau_j} \leq C \epsilon^{-b + A} \left( \log \frac{1}{\epsilon} \right)^{C}$.

Then there exists a symplectic coordinate change
\[ \Psi_j : T^{d+1}_{s_j} \times B(\tau_j) \rightarrow T^{d+1}_{s_{j+1}} \times B(\tau_{j+1}) \] (2.10)
with
\[ \|\Psi_j - \text{id}\|_{s_{j+1}, \tau_{j+1}} \leq C \epsilon^{(j+1)(a-b)} \left( \log \frac{1}{\epsilon} \right)^{C} \] (2.11)
such that
\[ H^{(j+1)} = H^{(j)} \circ \Psi_j \]
\[ = \epsilon^{-a} H_0(I) + h^{(j+1)}(t, I) + R^{(j+1)}(\theta, t, I) + R^{(j+1)}_+(\theta, t, I), \quad (\theta, t, I) \in T^{d+1}_{s_{j+1}} \times B(\tau_{j+1}), \] (2.12)
where the new Hamiltonian functions $h^{(j+1)}(t, I)$, $R^{(j+1)}(\theta, t, I)$ and $R^{(j+1)}_+(\theta, t, I)$ obey the conditions
1. $(j + 1) - 4(j + 1)$. 

Proof. Assume that the change $\Psi_j$ is implicitly defined by

$$
\Psi_j : \begin{cases} 
I = \rho + \frac{\partial S}{\partial \theta}, \\
\phi = \theta + \frac{\partial S}{\partial \rho}, \\
t = t,
\end{cases}
$$

(2.13)

where $S = S(\theta, t, \rho)$ is the generating function, which will be proved to be analytic in a smaller domain $T_{s_j+1}^d \times B(\tau_{j+1})$. By a simple computation, we have

$$
dI \wedge d\theta = d\rho \wedge d\theta + \sum_{i,j=1}^d \frac{\partial^2 S}{\partial \rho_i \partial \theta_j} d\rho_i \wedge d\theta_j = d\rho \wedge d\phi.
$$

Thus the coordinate change $\Psi_j$ is symplectic if it exists. Moreover, we get the changed Hamiltonian

$$
H^{(j+1)} = H^{(j)} \circ \Psi_j
$$

(2.14)

where $\theta = \theta(\phi, t, \rho)$ is implicitly defined by (2.13). We drop the sup-index $(j)$. We replace $h^{(j)}$ by $h$, for example.

By Taylor's formula, we have

$$
H^{(j+1)} = \varepsilon - a H_0(\rho) + R(\theta, t, \rho) + R_*,
$$

(2.15)

where

$$
R_* = R_+ \left( \theta, t, \rho + \frac{\partial S}{\partial \theta} \right)
$$

(2.16)

and

$$
R_* = \varepsilon - a \int_0^1 (1 - \tau) \partial_\tau^2 H_0 \left( \rho + \tau \frac{\partial S}{\partial \theta} \right) \left( \frac{\partial S}{\partial \theta} \right)^2 d\tau,
$$

(2.17a)

$$
+ \int_0^1 \partial_\tau R \left( \theta, t, \rho + \tau \frac{\partial S}{\partial \theta} \right) \frac{\partial S}{\partial \theta} d\tau,
$$

(2.17b)

$$
+ \int_0^1 \partial_\tau h \left( t, \rho + \tau \frac{\partial S}{\partial \theta} \right) \frac{\partial S}{\partial \theta} d\tau.
$$

(2.17c)

Thus, we derive the homological equation

$$
\frac{\partial S}{\partial \theta} + \varepsilon - a \left( \omega(\rho), \frac{\partial S}{\partial \theta} \right) = \hat{R}(0, t, \rho) - R(\theta, t, \rho), \quad S = S(\theta, t, \rho),
$$

(2.18)

where $\hat{R}(0, t, \rho)$ is the 0-Fourier coefficient of $R(\theta, t, \rho)$ as the function of $\theta$. Recalling 3(j), we have $\Gamma_{K_j} R^{(j)} = R^{(j)}$. Thus we can assume $\Gamma_{K_j} S = S$. For $f \in \{S, R\}$, let

$$
f(\theta, t, \rho) = \sum_{|k| + |l| \leq K_j, k \neq 0} \hat{f}(k, l, \rho) e^{i(k, \theta) + lt}, \quad i^2 = -1.
$$

(2.19)

By passing (2.18) to Fourier coefficients, we have

$$
\widehat{S}(k, l, \rho) = \frac{i}{\varepsilon - a \langle k, \omega(\rho) \rangle + l} \hat{R}(k, l, \rho), \quad |k| + |l| \leq K_j, \quad k \in \mathbb{Z}^d \setminus \{0\}, \quad l \in \mathbb{Z}.
$$

(2.20)
Let \( \rho \in B(\tau_j) \subseteq B(\tau_0) = B_0 \). By (2.3), we have
\[
|\hat{S}(k,l,\rho)| \leq C e^{\frac{|k|}{\gamma}} |\hat{R}(k,l,\rho)|.
\]
Moreover, by 3(j), for \((\theta,t,\rho) \in T^{d+1}_j \times B(\tau_j)\), we get
\[
|S(\theta,t,\rho)| \leq C e^{\frac{|k|}{\gamma}} \| R \|_{s_j,\tau_j} \sum_{|k|+|l| \leq K_j} |k|^{d+1} e^{-(s_j-s^+)} ((k+|l|)) \leq C e^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C,
\]
where \( C = C(m_0, d) \). Then by Cauchy’s estimate, we have
\[
\left\| \frac{\partial^{p+q}}{\partial \theta^p \partial \rho^q} S(\theta,t,\rho) \right\|_{s_j,\tau_j} \leq C e^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C, \quad p + q \leq 2, \quad p \geq 0, \quad q \geq 0.
\]
By (2.13), (2.22) and the implicit function theorem (the analytic version), we see that there are analytic functions \( u = u(\phi,t,\rho) \) and \( v = v(\phi,t,\rho) \) defined on the domain \( T^{d+1}_j \times B(\tau_j) \supset T^{d+1}_{j+1} \times B(\tau_{j+1}) \) with
\[
\frac{\partial S(t,\theta,\rho)}{\partial \theta} = u(\phi,t,\rho), \quad \frac{\partial S(t,\theta,\rho)}{\partial \rho} = -v(\phi,t,\rho)
\]
and
\[
\|u\|_{s_j,\tau_j^3} \leq \varepsilon^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C, \quad \|v\|_{s_j,\tau_j^3} \leq \varepsilon^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C
\]
such that
\[
\Psi_j : \begin{cases} 
\theta = \rho + u(\phi,t,\rho), \\
t = t.
\end{cases}
\]
Note
\[
\varepsilon^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C \ll s_j - s_{j+1}, \quad \varepsilon^{(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C \ll \tau_j - \tau_{j+1}.
\]
Then \( \Psi_j(T^{d+1}_{j+1} \times B(\tau_{j+1}) \subset T^{d+1}_j \times B(\tau_j) \). This leads to (2.10) and (2.11). Then by 4(j), we have
\[
\| R_\ast(\phi,t,\rho) \|_{s_j,\tau_j} \leq \| R_+ \|_{s_j,\tau_j} \leq C \varepsilon^{-b-A} \left( \log \frac{1}{\varepsilon} \right)^C.
\]
By (2.23), (2.24), 2(j) and 3(j), we have
\[
\| R_\ast(\phi,t,\rho) \|_{s_j,\tau_j^3} \leq C \varepsilon^{-a} \| R \|_{s_j,\tau_j} \| u \|_{s_j,\tau_j} \| u \|_{s_j,\tau_j^3} \tau_j^{-1} + \| h \|_{s_j,\tau_j} \| u \|_{s_j,\tau_j^3} \tau_j^{-1} \leq C \varepsilon^{-b+(j+1)(a-b)} \left( \log \frac{1}{\varepsilon} \right)^C.
\]
Let
\[
R^{(j+1)}(\phi,t,\rho) = \Gamma_{K_{j+1}} R_\ast
\]
and
\[
R^{(j+1)}_+(\phi,t,\rho) = (1 - \Gamma_{K_{j+1}}) R_\ast + R_\ast.
\]
By (2.26) and (2.27), we see that \( R^{(j+1)} \) and \( R^{(j+1)}_+ \) satisfy 3(j + 1) and 4(j + 1), respectively. Let
\[
h^{(j+1)}(t,\rho) = h(t,\rho) + \hat{R}(0,t,\rho).
\]
Then by 2(j) and 3(j), we have
\[
\| h^{(j+1)} \|_{s_{j+1},\tau_{j+1}} \leq \| h^j \|_{s_j,\tau_j} + \| R \|_{s_j,\tau_j} \leq C \varepsilon^{-b},
\]
which fulfills $2(j+1)$. Finally, it is obvious that the functions $h^{(j+1)}(t, I)$, $R^{(j+1)}(\theta, t, I)$ and $R_+^{(j+1)}(\theta, t, I)$ are analytic in $T_{s_{m+1}}^{d+1} \times B(\tau_{m+1})$. Since $R$ is real for the real arguments, we have
\[
\hat{R}(-k, -l, \rho) = \overline{\hat{R}(k, l, \rho)}, \quad \rho \in \mathbb{R}^d, \quad (k, l) \neq (0, 0),
\]
where the bar is the complex conjugate. By (2.20), we have that (2.12) holds. This completes the proof of this lemma.

It follows that $S(\theta, t, \rho)$ is real for the real arguments $(\theta, t, \rho)$. Moreover, the functions $h^{(j+1)}(t, I)$, $R^{(j+1)}(\theta, t, I)$ and $R_+^{(j+1)}(\theta, t, I)$ are real for real arguments. By (2.15)–(2.17) and (2.28)–(2.30), we see that (2.12) holds. This completes the proof of this lemma. \(\blacksquare\)

Applying Lemma 2.1 to (2.6) with $j = 0$ and $h^{(0)} \equiv 0$, we get $m_0$ symplectic changes $\Psi_j$ ($j = 0, \ldots, m_0 - 1$) such that
\[
\Psi := \Psi_0 \circ \cdots \circ \Psi_{m_0-1}, \quad \Psi(T_{s_{m_0}}^{d+1} \times B(\tau_{m_0})) \subset T_{s_{m_0}}^{d+1} \times B(\tau_0),
\]
\[
H^{(m_0)} = \varepsilon^{-a} H_0(\rho) + \hat{h}^{(m_0)}(t, \rho) + \hat{R}(\phi, t, \rho), \quad (\phi, t, \rho) \in T_{s_{m_0}}^{d+1} \times B(\tau_{m_0})
\]
with
\[
\hat{R} = \hat{\mathcal{R}}(\phi, t, \rho) = R^{(m_0)}(\phi, t, \rho) + R_+^{(m_0)}(\phi, t, \rho), \quad (\phi, t, \rho) \in T_{s_{m_0}}^{d+1} \times B(\tau_{m_0}),
\]
where $h^{(m_0)}$, $R^{(m_0)}$ and $R_+^{(m_0)}$ obey the conditions 1($j$)–4($j$) with $j = m_0$. Thus,
\[
\|h^{(m_0)}(t, \rho)\|_{s_{m_0}, \tau_{m_0}} \leq C \varepsilon^{-b}, \quad \|\hat{R}(\phi, t, \rho)\|_{s_{m_0}, \tau_{m_0}} \leq C \varepsilon^{-b} \left(\log \frac{1}{\varepsilon}\right)^C.
\]

Let $[h^{(m_0)}](I) = \hat{\mathcal{R}}^{(m_0)}(0, I)$ be the 0-Fourier coefficient of $h^{(m_0)}(t, I)$ as the function of $t$. In order to eliminate the dependence of $h^{(m_0)}(t, I)$ on the time-variable $t$, we introduce the following transformation:
\[
\hat{\Psi} : \rho = I, \quad \phi = \theta + \frac{\partial}{\partial t} \hat{S}(t, I), \quad \text{where} \quad \hat{S}(t, I) := \int_0^t ([h^{(m_0)}](I) - h^{(m_0)}(\xi, I)) d\xi.
\]
It is easy to verify $d\rho \wedge d\phi = dI \wedge d\theta$. Thus the transformation $\hat{\Psi}$ is symplectic. Note that the transformation is not small. So $\hat{\Psi}$ is not close to the identity. In order to apply $\hat{\Psi}$ to $H^{(m_0)}$, we introduce a domain
\[
\mathcal{D} := \{t = t_1 + t_2 i \in T_{s_{m_0}} : |t_2| \leq \varepsilon^{2b} \} \times \{I = x + yi \in B(\tau_{m_0}) : |x - I_0| \leq \left(\frac{1}{\varepsilon}\right)^-C, \quad |y| \leq \varepsilon^{2b}\}.
\]
Note that $h^{(m_0)}(t, I)$ is real for real arguments. Thus, for $(t, I) \in \mathcal{D}$, we have
\[
\left|\frac{\partial}{\partial t} \hat{S}(t, I)\right| = \left|\frac{\partial}{\partial t} \hat{S}(t_1 + t_2 i, x + yi) - \frac{\partial}{\partial t} \hat{S}(t_1, x)\right| \leq ||\partial^2 \hat{S}(t, I)||_{\mathcal{D}}(|t_2| + |y|) \leq \varepsilon^{-b} \left(\frac{1}{2}\right)^{s_{m_0}}.
\]
Therefore, \(\hat{\Psi}(T_{s_{m_0}/2}^d \times \mathcal{D}) \subset T_{s_{m_0}}^{d+1} \times B(\tau_{m_0})\)
and
\[
\hat{H} := H^{(m_0)} \circ \hat{\Psi} = \varepsilon^{-a} H_0(I) + [h^{(m_0)}](I) + \hat{R}(\theta, t, I), \quad (\theta, t, I) \in T_{s_{m_0}/2}^{d} \times \mathcal{D},
\]
where
\[
\hat{R}(\theta, t, I) := \hat{R}(\theta + \partial_t \hat{S}(t, I), t, I).
\]
Recall that $\det(\partial^2 H_0(I)) \geq c_3 > 0$. By 2($j$) with $j = m_0$, we have
\[
\varepsilon^a \|\partial^2 [h^{(m_0)}](I)\|_{\tau_{m_0}/2} \leq C \varepsilon^{-b} \left(\frac{1}{\varepsilon}\right)^C \leq c_3.
\]
Solving the equation $\partial_t H_0(I) + \varepsilon^a \partial_t [h^{\text{low}}](I) = \omega(I_0)$ by the Newton iteration, we see that there exists an $I_* \in \mathbb{R}^d \cap B(r_{\text{low}}/2)$ with $|I_* - I_0| \leq C \varepsilon^{-b} (\log \varepsilon)^C$ such that

$$
\partial_t H_0(I_*) + \varepsilon^a \partial_t [h^{\text{low}}](I_*) = \omega(I_0), \quad \text{where } \omega(I_0) = \partial_\theta H_0(I_0).
$$

Let $I = I_* + \rho$ with $\rho \in \mathbb{C}^d$ and $|\rho| < \varepsilon^{2b}$. By Taylor’s formula at $I = I_*$ up to the second order,

$$
\tilde{H} = \varepsilon^{-a} H_0(I_*) + [h^{\text{low}}](I_*) + \varepsilon^{-a} \langle \omega(I_0), \rho \rangle + \varepsilon^{-a} \langle \Omega \rho, \rho \rangle + R_{\text{low}}(\theta, t, \rho) + R_{\text{high}}(\theta, t, \rho), \quad (2.39)
$$

where

$$
\Omega = \frac{1}{2} \partial_\theta^2 (H_0(I) + \varepsilon^a [h^{\text{low}}](I)) |_{I = I_*},
$$

$$
R_{\text{low}}(\theta, t, \rho) = \tilde{R}(\theta, t, I_*) + \langle \partial_\theta \tilde{R}(\theta, t, I_*) \rho, \rho \rangle + \frac{1}{2} \langle \partial_\theta^2 \tilde{R}(\theta, t, I_*) \rho, \rho \rangle, \quad (2.40)
$$

$$
R_{\text{high}}(\theta, t, \rho) = \varepsilon^{-a} \int_0^1 \left( \varepsilon^{-2} \partial_\theta^3 H_0(I_* + x\rho) + \varepsilon^{-a} [h^{\text{low}}](I_* + x\rho) + \varepsilon^a \langle \tilde{R}(\theta, t, I_* + x\rho) \rangle \right) dx. \quad (2.41)
$$

We introduce some notation: with a slight abuse of notation, let $\| \Omega \| = \frac{1}{2} \partial_\theta^2 (H_0(I) + \varepsilon^a [h^{\text{low}}](I)) |_{I = I_*}$, then we write $f = O_\varepsilon (|I|^\alpha)$. If $f$ is independent of $I$, we write $f = O_\varepsilon (|I|^\alpha)$. Note that $|I|^\alpha \lesssim 1$. Let $E = \lambda - 9b - 1$. In view of (2.35), we have

$$
\partial_\theta^p \tilde{R}(\theta, t, I_*) = O_\varepsilon (|I|^\alpha), \quad p = 0, 1, 2, \quad (2.43)
$$

$$
R_{\text{high}}(\theta, t, \rho) = O_\varepsilon (|\rho|^3). \quad (2.44)
$$

By (2.4), (2.38) and (2.40), we have

$$
\| \Omega \| \leq C, \quad \| \Omega^{-1} \| \leq C. \quad (2.45)
$$

### 3 Extension of Kolmogorov’s theorem

By a slight abuse of notation, reset $s_0 := \varepsilon^{2b}$ and $r_0 := \varepsilon^{2b}$.

**Theorem 3.1.** Consider a Hamiltonian

$$
H = C + N(I) + R_{\text{low}}(\theta, t, I) + R_{\text{high}}(\theta, t, I), \quad (\theta, t, I) \in T_{s_0}^{d+1} \times B(r_0)
$$

with the symplectic structure $dI \wedge d\theta$ satisfying

1. $N(I) = \varepsilon^{-a} \langle \omega, I \rangle + \varepsilon^{-a} \langle \Omega, I \rangle$;
2. $\omega = \omega(I_0)$ is D.C., i.e.,

$$
\varepsilon^{-a} (k, \omega) + I \geq \frac{\gamma}{(1 + |k|)^{d+1}}, \quad (k, l) \in \mathbb{Z}^d \times \mathbb{Z} \setminus \{(0, 0)\};
$$

3. $\| \Omega \| \leq C$, $\| \Omega^{-1} \| \leq C$;
4. $N(I)$, $R_{\text{low}}(\theta, t, I)$ and $R_{\text{high}}(\theta, t, I)$ are real analytic in $T_{s_0}^{d+1} \times B(r_0)$;
5. $R_{\text{low}} = R_0(\theta, t) + (R_1(\theta, t), I) + (R_2(\theta, t) I) I$ with $R_j = O_\varepsilon (|I|^j)$, $j = 0, 1, 2$;
6. $R_{\text{high}} = O_{s_0, r_0} (|\rho|^{3})$.

Then there exists a symplectic coordinate change

$$
\Psi : I = \rho + u(\phi, t, \rho), \quad \theta = \phi + v(\phi, t, \rho), \quad t = t,
$$
Then there exists a symplectic coordinate where $u = O_{s_0/2,r_0/2}(\varepsilon^{\frac{d}{2}})$ and $v = O_{s_0/2,r_0/2}(\varepsilon^{\frac{d}{2}})$ such that $\Psi(\mathbb{T}^{d+1}_{s_0/2} \times B(\frac{r_0}{2})) \subset \mathbb{T}^{d+1} \times B(r_0)$ and the Hamiltonian $H$ is changed by $\Psi$ into

$$H^{\infty}(\phi, t, \rho) = H \circ \Psi(\phi, t, \rho) = C + \varepsilon^{-a}(\Omega^{\infty}(\rho) + O_{s_0/2,r_0/2}(\varepsilon^{-a-6b}|\rho|^{3}), \ (\phi, t, \rho) \in \mathbb{T}^{d+1}_{s_0/2} \times B\left(\frac{r_0}{2}\right),$$

where

$$|\Omega^{\infty} - \Omega| \leq C \varepsilon^{\frac{d}{2}}.$$

**Proposition 3.2.** The Hamiltonian system defined by the Hamiltonian $H$ has an invariant torus with the rotational frequency $(\varepsilon^{-a}\omega(I_0),2\pi)$ in the extended phase space $\mathbb{T}^{d+1} \times \mathbb{R}^d$.

**Remark 3.3.** The frequency function $\omega(I) = \partial_t H_0(I)$ is independent of $\varepsilon$. The choice of $I_0$ indeed depends on $\varepsilon$, since the Diophantine condition involves $\varepsilon$.

The proof is finished by the following iterative lemma. To this end, we introduce some iterative constants, iterative parameters and iterative domains:

- $\epsilon_0 = \epsilon^{\frac{d}{2}}, \epsilon_j = \epsilon^{\frac{d}{2}(4/3)^j}, \ j = 1, 2, \ldots, \epsilon_0 > \epsilon_1 > \cdots > \epsilon_j \downarrow 0$;
- $\epsilon_j = \frac{\sum_{l=1}^{j} \epsilon_{l-1}^{-2}}{100 \sum_{l=1}^{j} \epsilon_{l-1}^{-2}}$;
- $s_j = s_0(1 - \epsilon_j)$ (so $s_j > s_0/2$ for all $j = 1, 2, \ldots$);
- $r_j = r_0(1 - \epsilon_j)$ (so $r_j > r_0/2$ for all $j = 1, 2, \ldots$);
- $s_i = (1 - \frac{1}{10})s_j + \frac{1}{10}s_{j+1}$ (so $s_{j+1} < s_j, s_j < s_i, l = 1, 2, \ldots, 10$);
- $r_i = (1 - \frac{1}{10})r_j + \frac{1}{10}r_{j+1}$ (so $r_{j+1} < r_j, l = 1, 2, \ldots, 10$);
- for any $\alpha > 0$, write $\epsilon_j^{\alpha} = j^C(\log \frac{1}{\epsilon_j})^{C} \epsilon_j^{\alpha}$.

**Lemma 3.4 (Iterative lemma).** Assume that we have had $m$ Hamiltonian functions for $j = 0, 1, \ldots, m - 1$ satisfying

$$H^{(j)} = N^{(j)} + R^{(j)}_{\text{low}} + R^{(j)}_{\text{high}}, \quad (3.1)$$

$$N^{(j)} = \varepsilon^{-a}(\omega, I) + \varepsilon^{-a}(\Omega^{(j)}(I), I), \quad ||\Omega^{(j)} - \Omega^{(j-1)}|| \leq C\epsilon_{j-1}^{-1} \varepsilon^{-b}, \quad j \geq 1, \quad (3.2)$$

$$\Omega^{(0)} = \Omega, \quad b^{'} := 12d(a + b),$$

$$R^{(j)}_{\text{low}}(\theta, t, I) \quad \text{and} \quad R^{(j)}_{\text{high}}(\theta, t, I) \quad \text{are real analytic in} \quad \mathbb{T}^{d+1}_{s_j} \times B(r_j). \quad (3.3)$$

$$R^{(j)}_{\text{low}} = R^{(j)}_{\text{low}}(\theta, t) + \langle R^{(j)}_{1}(\theta, t), I \rangle + \langle R^{(j)}_{2}(\theta, t), I \rangle, \quad (3.4)$$

$$R^{(j)}_{p}(\theta, t, I) = O_{s_j}(\epsilon_j), \quad p = 0, 1, 2, \quad (3.5)$$

$$R^{(j)}_{\text{high}}(\theta, t, I) = O_{s_j}(\epsilon_j^{(a+6b)}|I|^{3}). \quad (3.6)$$

Then there exists a symplectic coordinate

$$\Psi_m : I = \nu + \rho + u_m(\phi, t, \rho), \quad \theta = \phi + v_m(\phi, t, \rho), \quad t = t, \quad (3.7)$$

where $\nu$ is a constant depending on $\varepsilon$, $(\phi, t, \rho) \in \mathbb{T}^{d+1}_{s_m} \times B(r_m)$ and

$$u_m = O_{s_m,r_m}(\varepsilon^{-3b'} \epsilon_{m-1}), \quad v_m = O_{s_m,r_m}(\varepsilon^{-3b'} \epsilon_{m-1}), \quad (3.8)$$

$$\Psi_m(\mathbb{T}^{d+1}_{s_m} \times B(r_m)) \subset \mathbb{T}^{d+1}_{s_{m-1}} \times B(r_{m-1}) \quad (3.9)$$

such that the changed Hamiltonian $H^{(m)} = H^{(m-1)} \circ \Psi_m$ satisfies (3.1)–(3.6) with $j$ replaced by $m$.

**Proof.** Assume that the coordinate change $\Psi = \Psi_m$ can be implicitly defined by

$$\Psi : I = \nu + \rho + \frac{\partial S}{\partial \theta}, \quad \phi = \theta + \frac{\partial S}{\partial \rho}, \quad t = t, \quad (3.10)$$

where the function $S = S(\theta, t, \rho)$ and constant vector $\nu$ will be specified by homological equations later. Then $\Psi$ is symplectic if it is well defined. Let

$$S = S_0(\theta, t) + \langle S_1(\theta, t), \rho \rangle + \langle S_2(\theta, t), \rho \rangle. \quad (3.11)$$
Then
\[
H^{(m)} = H^{(m-1)} \circ \Psi_m
\]
\[
= N^{(m-1)} \left( v + \rho + \frac{\partial S}{\partial \theta} \right) + R^{(m-1)}_{\text{low}} \left( \theta, t, v + \rho + \frac{\partial S}{\partial \theta} \right) + R^{(m-1)}_{\text{high}} \left( \theta, t, v + \rho + \frac{\partial S}{\partial \theta} \right) + \frac{\partial S(\theta, t, \rho)}{\partial t}
\]
\[
= \varepsilon^{-a} \langle \omega, v + \rho + \partial_t S \rangle + \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (v + \rho + \partial_t S, v + \rho + \partial_t S) + R^{(m-1)}_0(\theta, t) + \frac{\partial S}{\partial t}
\]
\[
+ \left( R^{(m-1)}_1(\theta, t), v + \rho + \partial_t S \right) + \langle R^{(m-1)}_2(\theta, t)(v + \rho + \partial_t S) \rangle, v + \rho + \partial_t S
\]
\[
+ R_{\text{high}}(\theta, t, v + \rho + \partial_t S).
\]
Let \( \omega \cdot \partial_t S = \langle \omega, \partial_t S \rangle \). By a simple computation, we have
\[
H^{(m)} = \varepsilon^{-a} \langle \omega, v \rangle + \varepsilon^{-a} \langle \omega, \rho \rangle + \varepsilon^{-a} \langle \omega \cdot \partial_t S_0 + \langle \omega \cdot \partial_t S_1, \rho \rangle + \langle \omega \cdot \partial_t S_2, \rho \rangle \rangle
\]
\[
+ \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (v, v) + 2 \langle \Omega^{(m-1)} \rangle (v, \rho) + \langle \Omega^{(m-1)} \rangle (v, \partial_t S) \rangle + 2 \langle \Omega^{(m-1)} \rangle (\partial_t S, \partial_t S, \rho) \rangle
\]
\[
+ \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (\partial_t S_0, \partial_t S_0) + R^{(m-1)}_0(\theta, t) + \langle R^{(m-1)}_1(\theta, t), v + \partial_t S \rangle + \langle R^{(m-1)}_2(\theta, t), v + \partial_t S \rangle + \langle R^{(m-1)}_2(\theta, t), v + \partial_t S \rangle
\]
\[
+ \frac{1}{2} \langle \partial_t R_{\text{high}}(\theta, t, 0) \partial_t S_0(\rho, \rho) + \partial_t S \rangle
\]
\[
+ R_{\text{high}}(\theta, t, v + \partial_t S) - \frac{1}{2} \langle \partial_t R_{\text{high}}(\theta, t, 0) \partial_t S_0(\rho, \rho) \rangle.
\]

In the following, we omit the term \( \varepsilon^{-a} \langle \langle \omega, v \rangle + \langle \Omega^{(m-1)} \rangle (v, v) \rangle \) which does not affect the dynamics. Let
\[
R_+ (\theta, t) = 2 \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (v, \partial_t S_0(\theta, t)) + R^{(m-1)}_0(\theta, t),
\]
\[
R_{++} (\theta, t) = \frac{1}{2} \partial_t R_{\text{high}}(\theta, t, 0) \partial_t S_0(\theta, t) + 2 \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (\partial_t S_0(\theta, t)) + R^{(m-1)}_2(\theta, t).
\]

From the changed \( H^{(m)} \), we derive the homological equations
\[
\varepsilon^{-a} \omega \cdot \partial_t S_0(\theta, t) + \partial_t S_0(\theta, t) + R^{(m-1)}_0(\theta, t) = R^{(m-1)}_0(0, 0),
\]
\[
\varepsilon^{-a} \omega \cdot \partial_t S_1(\theta, t) + \partial_t S_1(\theta, t) + R_+ (\theta, t) = R_+(0, 0),
\]
\[
2 \varepsilon^{-a} \Omega^{(m-1)}(v) = - R_+(0, 0),
\]
\[
\varepsilon^{-a} \omega \cdot \partial_t S_2(\theta, t) + \partial_t S_2(\theta, t) + R_{++}(\theta, t) = R_{++}(0, 0).
\]

Let
\[
\Omega^{(m)} = \Omega^{(m-1)} + \varepsilon^a R_{++}(0, 0)
\]
and
\[
R_+ = \varepsilon^{-a} \langle \langle \Omega^{(m-1)} \rangle (\partial_t S_0, \partial_t S) + 2 \langle \Omega^{(m-1)} \rangle (v, \partial_t S) \rangle
\]
\[
+ \langle R^{(m-1)}_1(\theta, t), v + \partial_t S \rangle
\]
\[
+ \langle R^{(m-1)}_2(\theta, t), v + \partial_t S \rangle + 2 \langle R^{(m-1)}_2(\theta, t), v + \partial_t S \rangle
\]
\[
+ R_{\text{high}}(\theta, t, v + \partial_t S) - \frac{1}{2} \langle \partial_t R_{\text{high}}(\theta, t, 0) \partial_t S_0(\rho, \rho) \rangle
\]
\[
+ 2 \varepsilon^{-a} \langle \Omega^{(m-1)} \rangle (\partial_t S_2, \partial_t S_2 \partial_t S_2, \partial_t S_2)
\]

Then
\[
H^{(m)} = C + \varepsilon^{-a} \langle \langle \omega, \rho \rangle + \langle \Omega^{(m)} \rangle (\rho, \rho) \rangle + R_+.
\]

We are now in a position to investigate the homological equations.
(1) The solution to (3.15): passing (3.15) to Fourier coefficients, we have
\[ S_0(\theta, t) = \sum_{(0,0) \neq (k,l) \in \mathbb{Z}^d \times \mathbb{Z}} \frac{i}{\varepsilon^{-a} \langle k, \omega \rangle + \lambda} R_0^{(m-1)}(k, l) e^{i(k, \theta) + i\lambda}. \]

By \( \omega \in \text{D.C.} \),
\[ \|S_0\|_{s_{m-1}} \leq C \left( \frac{1}{s_{m-1}^2 - s_{m-1}^0} \right)^{1/2 \gamma} \|R_0\|_{s_{m-1}} \leq \varepsilon_{m-1}^{-b(4d+4)}. \]  

(2) The solution to (3.16): by (3.5), (3.13) and (3.22) with \( j = m - 1 \), and using Cauchy’s estimate, we have
\[ R_* = O_{s^2}(e^{-b(4d+6) - a}). \]

By (3.16), we have
\[ S_1(\theta, t) = \sum_{(0,0) \neq (k,l) \in \mathbb{Z}^d \times \mathbb{Z}} \frac{i}{\varepsilon^{-a} \langle k, \omega \rangle + \lambda} \tilde{R}_*(k, l) e^{i(k, \theta) + i\lambda}. \]

Therefore,
\[ \|S_1\|_{s_{m-1}} \leq C \left( \frac{1}{s_{m-1}^2 - s_{m-1}^0} \right)^{1/2 \gamma} \|R_*\|_{s_{m-1}} \leq \varepsilon_{m-1}^{-b(8d+10) - a}. \]  

(3) The solution to (3.17): by (3.17) and (2.3),
\[ |\nu| = |(2\varepsilon^{-a} \Omega^{(m-1)})^{-1} \tilde{R}_*(0, 0)| \leq C \varepsilon_{m-1}^{-b(4d+6)}. \]

By (3.14), (3.22), (3.24) and Cauchy’s estimate, we have
\[ \|R_* + \lambda\|_{s_{m-1}} \leq C \varepsilon_{m-1}^{-b(8d+12) - 2a}. \]

Passing (3.18) to Fourier coefficients, we obtain
\[ S_2(\theta, t) = \sum_{(0,0) \neq (k,l) \in \mathbb{Z}^d \times \mathbb{Z}} \frac{i}{\varepsilon^{-a} \langle k, \omega \rangle + \lambda} \tilde{R}_*(k, l) e^{i(k, \theta) + i\lambda}. \]

It follows that
\[ \|S_2\|_{s_{m-1}} \leq C \left( \frac{1}{s_{m-1}^2 - s_{m-1}^0} \right)^{1/2 \gamma} \|\tilde{R}_*\|_{s_{m-1}} \leq \varepsilon_{m-1}^{-b(12d+16) - 2a}. \]  

(3) The solution to (3.17): by (3.17) and (3.23),
\[ |\nu| = |(2\varepsilon^{-a} \Omega^{(m-1)})^{-1} \tilde{R}_*(0, 0)| \leq C \varepsilon_{m-1}^{-b}. \]

This verifies (2.3) in the iterative lemma. By (3.22), (3.24) and (3.26), we have
\[ \|S\|_{s_{m-1}, r_{m-1}} = \|S_0 + \langle S_1, \rho \rangle + \langle S_2(\rho, \rho) \rangle\|_{s_{m-1}, r_{m-1}} \leq C \varepsilon^{-\frac{3\varepsilon}{\varepsilon} \varepsilon_{m-1}}. \]

By Cauchy’s estimate,
\[ \left\| \sum_{|\alpha| + |\beta| \leq 2} \frac{\partial^\alpha \theta \partial^\beta \theta}{\|S\|_{s_{m-1}, r_{m-1}}} \right\|_{s_{m-1}, r_{m-1}} \leq C \varepsilon^{-2b} \varepsilon_{m-1}. \]

By using the implicit theorem, we see that there exist \( u = u_m(\phi, t, \rho) \) and \( v = v_m(\phi, t, \rho) \) with
\[ \|u\|_{s_{m-1}, r_{m-1}}, \|v\|_{s_{m-1}, r_{m-1}} \leq C \varepsilon^{-\frac{3\varepsilon}{\varepsilon} \varepsilon_{m-1}} \]
such that \( \Psi = \Psi_m \),
\[ \Psi_m: \begin{cases} I = v + \rho + u(\phi, t, \rho), \\ \theta = \phi + v(\phi, t, \rho), \\ \phi, t, \rho \in T^d_{s_{m-1}} \times B(r_{m-1}) \end{cases} \]  

Finally, we have
\[ \Psi = \Psi_m. \]
is well defined and
\[ \Psi_m(\mathbb{S}_{s_m}^{d+1} \times B(r_m)) \subset \mathbb{T}_{s_{m-1}}^{d+1} \times B(r_{m-1}). \]
This verifies (3.8) and (3.9).

We are now in a position to estimate the new perturbation \( R_+ \). By (3.25) and (3.29)–(3.31), we have
\[ \|(3.20a)\|_{s_{m-1},r_{m-1}^7} \leq C e^{-\alpha}(\varepsilon^{-3}\varepsilon_{m-1}^7)^2 \leq C e^{-\alpha}(\varepsilon_{m-1}^7)^2 \left( \log \frac{1}{\varepsilon} \right)^C \leq \varepsilon_{m-1}^{4\varepsilon}. \]
Similarly,
\[ \|(3.20b)\|_{s_{m-1},r_{m-1}^7} \leq \varepsilon_{m-1}^{4\varepsilon}, \]
\[ \|(3.20c)\|_{s_{m-1},r_{m-1}^7} \leq \varepsilon_{m-1}^{4\varepsilon}. \]
Clearly,
\[ (3.20e) = O_{s_{m-1},r_{m-1}^7}(\varepsilon^{-3}\varepsilon_{m-1}^7 |\rho|^3) = O_{s_{m-1},r_{m-1}^7}(|\rho|^3). \]
Recall \( R_{\text{high}} = O_{s_{m-1},r_{m-1}^7}(\varepsilon^{-3} |I|^3) \). Then by Taylor’s formula and in view of (3.28), we have
\[ (3.20d) = R_{\text{high}}(\theta, t, \rho, \varepsilon_S) - \langle (\partial_x^2 R_{\text{high}}(\theta, t, \rho, \varepsilon_S) \rangle \rho, \rho \rangle = \sum_{\alpha \in \mathbb{N}^3, \alpha < \varepsilon_S} \frac{1}{\alpha!} ((\rho + \varepsilon_S) \cdot \varepsilon_t)^{\alpha} R_{\text{high}}(\theta, t, \rho, \varepsilon_S) - \frac{1}{2} \langle (\partial_x^2 R_{\text{high}}(\theta, t, \rho, \varepsilon_S) \rangle \rho, \rho \rangle = O_{s_{m-1},r_{m-1}^7}(\varepsilon^{-3} \varepsilon_{m-1}^7 |\rho|^3) + O_{s_{m-1},r_{m-1}^7}(\varepsilon_{m-1}^{4\varepsilon}). \]
Consequently, by (3.32)–(3.36), we have
\[ R_+(\phi, t, \rho) = O_{s_{m-1},r_{m-1}^7}(\varepsilon^{-3} |\rho|^3) + O_{s_{m-1},r_{m-1}^7}(\varepsilon_{m-1}^{4\varepsilon}). \]
By developing \( R_+(\phi, t, \rho) \) into Taylor’s formula of order 3 in \( \rho \), we can rewrite
\[ R_+(\phi, t, \rho) = R_{\text{low}}(\phi, t, \rho) + R_{\text{high}}(\phi, t, \rho), \]
where
\[ R_{\text{low}}(\phi, t) = \sum_{\alpha \in \mathbb{N}^3, \alpha < \varepsilon_S} \frac{1}{\alpha!} ((\rho + \varepsilon_S) \cdot \varepsilon_t)^{\alpha} R_{\text{low}}(\phi, t, \rho, \varepsilon_S) - \frac{1}{2} \langle (\partial_x^2 R_{\text{low}}(\theta, t, \rho, \varepsilon_S) \rangle \rho, \rho \rangle = O_{s_{m-1},r_{m-1}^7}(\varepsilon^{-3} |\rho|^3). \]
This verifies (3.4)–(3.6) with \( j = m \). As for the functions \( R_{\text{low}}(\phi, t, \rho) \) and \( R_{\text{high}}(\phi, t, \rho) \) are real for the real argument \( (\phi, t, \rho, \varepsilon_{s_m, r_{m-1}}) \) in \( \mathbb{T}^{d+1}_{s_0} \times B(r_m) \), the proof is the same as that of Lemma 2.1. This completes the proof of the iterative lemma.

4 Proof of the extended Kolmogorov’s theorem

Let \( \Psi(x, I) = \lim_{m \to \infty} \Psi_1 \circ \cdots \circ \Psi_m(x, I) \) and \( (x, I) \in \mathbb{T}^{d+1}_{s_{0}/2} \times B(r_0/2). \)
Recall that in the iterative lemma, we have proved
\[ \text{(i) } \mathbb{T}^{d+1}_{s_0} \times B(r_0) \supset \cdots \supset \mathbb{T}^{d+1}_{s_{j}} \times B(r_{j}) \supset \cdots \supset \mathbb{T}^{d+1}_{s_{0}/2} \times B(r_0/2); \]
\[ \text{(ii) } \Psi_j : \mathbb{T}^{d+1}_{s_{j}} \times B(r_j) \to \mathbb{T}^{d+1}_{s_{j-1}} \times B(r_{j-1}) \text{ is real analytic.} \]
Then \( \Psi^m := \Psi_1 \circ \cdots \circ \Psi_m : \mathbb{T}^{d+1}_{s_{j}} \times B(r_j) \subset \mathbb{T}^{d+1}_{s_{0}/2} \times B(r_0) \to \mathbb{T}^{d+1}_{s_{0}/2} \times B(r_0) \) is well defined and analytic in \( \mathbb{T}^{d+1}_{s_{0}/2} \times B(r_m). \) Also by (3.30) and (3.31) in the iterative lemma, we have
\[ \|\langle \partial \Psi_j(x, I) - 1 \rangle(z)\|_{s_{j},r_{j}} \leq e^{-\varepsilon_{j-1}} \varepsilon_{j-1} \|z\| \text{ for any } z \in T_{x,I} \mathbb{T}^{d+1}_{s_{j}} \times B(r_j). \]
for any \( z \in T_{x,I} \mathbb{T}^{d+1}_{s_{j}} \times B(r_j) \). Observe that \( \partial \Psi^m = (\partial \Psi_1 \circ \Psi_2 \circ \cdots \circ \Psi_m)(\partial \Psi_2 \circ \Psi_3 \circ \cdots \circ \Psi_m) \cdots (\partial \Psi_m). \)
Then by (4.1),
\[ \| \partial \Psi^{(m)}(x, t) \|_{L^{s_{0}, r_{0}}} \leq \prod_{j=1}^{m} \| (1 + \epsilon_{j-1}e^{-4b_{j}}) \|_{L^{s_{0}, r_{0}}} \leq \prod_{j=1}^{m} \left( 1 + \frac{1}{2j} \right) \| z \| \leq 2 \| z \|, \quad \forall z \in \mathbb{C}^{d+1} \times \mathbb{C}^{d}. \] (4.3)

Thus, for \( w = (x, t) \in T_{s_{0}/2}^{d+1} \times B(r_{0}/2) \),
\[ |\Psi^{(m+1)}(w) - \Psi^{(m)}(w)| = |\Psi^{(m)}(\Psi_{m+1}(w)) - \Psi^{(m)}(w)| \leq \| \partial \Psi^{(m)}(\Psi_{m+1}(w)) \|_{s_{0}/2, r_{0}/2} \sup_{w \in T_{s_{0}/2}^{d+1} \times B(r_{0}/2)} \| \Psi_{m+1}(w) - w \| \leq 2e^{-4b_{j}} \leq \frac{1}{2}, \] (4.4)

where we have used (3.8), (3.25) and (4.3) in (4.4). Write \( \Psi^{(m)} = \Psi^{(1)} + \sum_{l=2}^{m} (\Psi^{(l)} - \Psi^{(l-1)}) \), where \( \Psi^{(1)} := \Psi_{1} \). Then
\[ \| \Psi(w) - w \|_{s_{0}/2, r_{0}/2} \leq \| \Psi_{1}(w) - w \|_{s_{0}/2, r_{0}/2} + \sum_{l=2}^{\infty} \| \Psi^{(l)}(w) - \Psi^{(l-1)}(w) \|_{s_{0}/2, r_{0}/2} \leq \sum_{l=0}^{\infty} \frac{1}{2l} \leq 1. \]

This completes the proof of this theorem.

5 Application to coupled Duffing oscillators

Consider the coupled Duffing oscillators
\[ \ddot{x}_{j} + x_{j}^{2n+1} + \frac{\partial F(x, t)}{\partial x_{j}} = 0, \quad j = 1, 2, \ldots, m, \quad x = (x_{1}, \ldots, x_{m}), \] (5.1)
where \( m, n > 0 \) are fixed integers,
\[ F(x, t) = \sum_{0 \leq |\alpha| \leq 2n+1} \frac{P_{\alpha}(t)x^{\alpha}}{\alpha_{0}! \cdots \alpha_{m}!} \sum_{0 \leq \alpha_{1} + \cdots + \alpha_{m} \leq 2n+1} P_{\alpha_{1} \cdots \alpha_{m}}(t)x_{1}^{\alpha_{1}} \cdots x_{m}^{\alpha_{m}}, \quad \alpha_{j} \in \mathbb{Z}_{+}, \] (5.2)
and \( P_{\alpha}(t) : T \to \mathbb{R} \) is analytic (in this section, \( T := \mathbb{R}/\mathbb{Z} \)).

Let \( \tilde{A} \) be a large constant. Replacing \( x_{j} \) by \( \tilde{A}x_{j} \) in (5.1), we get
\[ \tilde{A}\ddot{x}_{j} + \tilde{A}^{2n+1}x_{j}^{2n+1} + \tilde{A}^{-1} \frac{\partial F(\tilde{A}x, t)}{\partial x_{j}} = 0, \]
i.e.,
\[ \ddot{x}_{j} + \tilde{A}^{2n}x_{j}^{2n+1} + \tilde{A}^{-2} \frac{\partial F(\tilde{A}x, t)}{\partial x_{j}} = 0, \quad j = 1, 2, \ldots, m. \] (5.3)

Let
\[ y_{j} = \tilde{A}^{-n}x_{j}, \quad j = 1, 2, \ldots, m. \]

Then
\[ \ddot{y}_{j} = \tilde{A}^{-n}\ddot{x}_{j} = -\tilde{A}^{n}x_{j}^{2n+1} - \tilde{A}^{-2n} \frac{\partial F(\tilde{A}x, t)}{\partial x_{j}}. \]

Thus, (5.1) can be written as a Hamiltonian system
\[ \dot{x}_{j} = \frac{\partial H}{\partial y_{j}}, \quad \dot{y}_{j} = -\frac{\partial H}{\partial x_{j}}, \quad j = 1, 2, \ldots, m, \] (5.4)
where
\[ H = \tilde{A}^{n} \sum_{j=1}^{m} \left( \frac{1}{2}y_{j}^{2} + \frac{1}{2(n+1)} x_{j}^{2n+2} \right) + \tilde{A}^{-(n+2)} F(\tilde{A}x, t). \] (5.5)
Consider an auxiliary Hamiltonian system which is autonomous:
\[
\dot{x} = \frac{\partial H_0}{\partial y}, \quad \dot{y} = -\frac{\partial H_0}{\partial x}, \quad H_0 = \frac{1}{2} y^2 + \frac{1}{2(n+1)} x^{2n+2}, \quad (x,y) \in \mathbb{R}^2.
\]  
(5.6)

Let \((x,y) = (u_0(t), v_0(t))\) be the solution to (5.6) with initial values \((u_0(0), v_0(0)) = (1, 0)\). Then this solution \((u_0, v_0)\) is clearly periodic. Let \(T_0\) be its minimal period. By energy conservation, we have
\[
(n+1)u_0^2(t) + u_0^{2n+2}(t) = 1, \quad \forall t \in \mathbb{R}.
\]  
(5.7)

Let
\[
\begin{align*}
\begin{cases}
  x = e^{\alpha t} I^a u_0(\theta T_0), \\
y = e^{\beta t} I^b v_0(\theta T_0),
\end{cases} \quad (\theta, I) \in T \times \mathbb{R}_+,
\end{align*}
\]

where \(\alpha = \frac{1}{n^2}, \beta = 1 - \alpha\) and \(c = \frac{1}{n^2}\). It is easy to check \(\text{det} \frac{\partial(x,y)}{\partial(\theta I)} = 1\), i.e., \(dx \wedge dy = d\theta \wedge dI\). Thus
\[
\Phi : \begin{cases}
x_j = e^{\alpha t} I_j^a u_0(\theta_j T_0), \\
y_j = e^{\beta t} I_j^b v_0(\theta_j T_0),
\end{cases} \quad j = 1, 2, \ldots, m
\]

is well defined and \(\Phi : (\theta, I) \in \mathbb{T}^m \times \mathbb{R}^m_+ \rightarrow \mathbb{R}^m \times \mathbb{R}^m\) is a symplectic transformation. Moreover,
\[
H^{(1)} := H \circ \Phi = \tilde{H}_0(I) + \tilde{R}(\theta, t, I),
\]

where
\[
\tilde{H}_0(I) = \frac{\epsilon^{2a}}{2(n+1)} \tilde{A}^n \sum_{j=1}^m I_j^{(2n+1)},
\]
\[
\tilde{R}(\theta, t, I) = \tilde{A}^{-(n+2)} F(\tilde{A}^a I_j^a u_0(\theta_j T_0), \ldots, \tilde{A}^a I_m^a u_0(\theta_m T_0), t).
\]

Note that \(F\) is a polynomial in \(x \in \mathbb{R}^m\) with degree \(2n+1\). So \(\tilde{R}(\theta, t, I)\) is a polynomial of \(\tilde{A}\) of degree \(n-1\) with its coefficients analytically depending on \((\theta, t, I) \in \mathbb{T}^{m+1} \times [1, 2]\). Let \(\epsilon = \tilde{A}^{-1}, a = n, b = n-1\) and
\[
H_0(I) = \frac{\epsilon^{2a}}{2(n+1)} \sum_{j=1}^m I_j^{(2n+1)}, \quad R(\theta, t, I) = \epsilon^b \tilde{R}(\theta, t, I).
\]

Then
\[
H^{(1)} = \epsilon^{-a} H_0(I) + \epsilon^{-b} R(\theta, t, I),
\]

where \(R(\theta, t, I)\) depends on \(\epsilon\). Note that it is harmless not to write explicitly the dependence of \(R\) on \(\epsilon\). Applying Theorem 1.1 and Remark 1.6 to \(H^{(1)}\), we see that there exists a subset \(\tilde{J}_0(\tilde{A}) \subset [1, 2]^m\) with \(\text{Leb} \tilde{J}_0(\tilde{A}) \geq 1 - (\log \tilde{A})^{-c_0}\) (some \(c_0 > 0\)) such that any solution to (5.4) starting from \(I(0) = (I_1(0), \ldots, I_m(0)) \in \tilde{J}_0(\tilde{A})\) is quasi-periodic with the rotational frequency \((\omega(I(0)), 1)\), where \(\omega = (\omega_j : j = 1, \ldots, m)\) and \(\omega_j = \epsilon^{-a} \frac{\partial H_0(I(0))}{\partial I_j}\). Returning to the coordinates \((x_i, \dot{x}_i : i = 1, \ldots, m)\), we see that for any large \(\tilde{A}\), there exist sets \(\Theta_{\tilde{A}}\) and \(\Theta_{\tilde{A}}\) with \(\Theta_{\tilde{A}} \subset \Theta_{\tilde{A}} \subset [C_1 \tilde{A}, C_2 \tilde{A}]^{2m}\), \(\text{Leb} \Theta_{\tilde{A}} = 1\) and \(\text{Leb} \Theta_{\tilde{A}} \geq 1 - (\log \tilde{A})^{-C_0}\) such that any solution to (5.1) starting from \((x_j(0), \dot{x}_j(0) : j = 1, \ldots, m) \in \Theta_{\tilde{A}}\) is quasi-periodic with the rotational frequency \((\omega(I(0)), 1)\). Note that any quasi-periodic solution is bounded. Also note that
\[
\lim_{\tilde{A} \rightarrow \infty} \frac{\Theta_{\tilde{A}}}{\Theta_{\tilde{A}}} = 1.
\]

Thus, we see that (5.1) is almost Lagrange stable. This completes the proof of Theorem 1.3.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 12071254). The authors are grateful to the referees for their helpful comments, which greatly improve the original manuscript of this paper.
References
1. Chatterjee A K, Kundu A, Kulkarni M. Spatiotemporal spread of perturbations in a driven dissipative Duffing chain: An out-of-time-ordered correlator approach. Phys Rev E, 2020, 102: 052103
2. Clerc M G, Coulibaly S, Ferré M A, et al. Chimera states in a Duffing oscillators chain coupled to nearest neighbors. Chaos, 2018, 28: 083126
3. Deka J P, Sarma A K, Govindarajan A, et al. Multifaceted nonlinear dynamics in PT-symmetric coupled Liénard oscillators. Nonlinear Dynam, 2020, 100: 1629–1640
4. Dieckhoff R, Zehnder E. Boundedness of solutions via twist theorem. Ann Sc Norm Super Pisa Cl Sci (5), 1987, 14: 79–95
5. Duffing G. Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung. Braunschweig: F. Vieweg & Sohn, 1918
6. Giorgilli A, Zehnder E. Exponential stability for time dependent potentials. Z Angew Math Phys, 1992, 43: 827–855
7. Holmes P J, Moon F C. Addendum: A magnetoelastic strange attractor. J Sound Vib, 1980, 69: 339
8. Jothimurugan R, Thamilmaran K, Rajasekar S, et al. Multiple resonance and anti-resonance in coupled Duffing oscillators. Nonlinear Dynam, 2016, 83: 1803–1814
9. Kiss I Z, Zhai Y M, Hudson J L. Resonance clustering in globally coupled electrochemical oscillators with external forcing. Phys Rev E, 2008, 77: 046204
10. Kovaleva A. Capture into resonance of coupled Duffing oscillators. Phys Rev E, 2015, 92: 022909
11. Laederich S, Levi M. Invariant curves and time-dependent potentials. Ergodic Theory Dynam Systems, 1991, 11: 365–378
12. Levi M. KAM theory for particles in periodic potentials. Ergodic Theory Dynam Systems, 1990, 10: 777–785
13. Levi M. Quasi-periodic motions in superquadratic time-periodic potentials. Comm Math Phys, 1991, 143: 43–83
14. Liu B. Boundedness for solutions of nonlinear Hill’s equations with periodic forcing terms via Moser’s twist theorem. J Differential Equations, 1989, 79: 304–315
15. Liu B. Boundedness for solutions of nonlinear periodic differential equations via Moser’s twist theorem. Acta Math Sin (Engl Ser), 1992, 8: 91–98
16. Moon F C, Holmes P J. A magnetoelastic strange attractor. J Sound Vib, 1979, 65: 275–296
17. Morris G R. A case of boundedness in Littlewood’s problem on oscillatory differential equations. Bull Aust Math Soc, 1976, 14: 71–93
18. Moser J. On invariant curves of area-preserving mapping of an annulus. Matematika, 1962, 6: 51–68
19. Rüssmann H. On the existence of curves of twist mappings of an annulus. In: Geometric Dynamics. Lecture Notes in Mathematics, vol. 1007. Berlin-Heidelberg: Springer, 1983, 677–718
20. Sarkar P, Ray D S. Vibrational antiresonance in nonlinear coupled systems. Phys Rev E, 2019, 99: 052221
21. Shena J, Lazarides N, Hizanidis J. Multi-branched resonances, chaos through quasiperiodicity, and asymmetric states in a superconducting dimer. Chaos, 2020, 30: 123127
22. Yuan X P. Invariant tori of Duffing-type equations. Adv Math (China), 1995, 24: 375–376
23. Yuan X P. Invariant tori of Duffing-type equations. J Differential Equations, 1998, 142: 231–262
24. Zhang F B. Quasiperiodic solutions of higher dimensional Duffing’s equations via the KAM theorem. Sci China Ser A, 2001, 44: 631–644