Localization in One-dimensional Quasi-periodic Nonlinear Systems

Jiansheng Geng, Jiangong You, Zhiyan Zhao
Department of Mathematics
Nanjing University, Nanjing 210093, P.R.China
Email: jgeng@nju.edu.cn; jyou@nju.edu.cn; zyqiao1985@gmail.com

Abstract

To investigate localization in one-dimensional quasi-periodic nonlinear systems, we consider the Schrödinger equation

$$\mathrm{i}\dot{q}_n + \epsilon(q_{n+1} + q_{n-1}) + V(n\tilde{\alpha} + x)q_n + |q_n|^2 q_n = 0, \quad n \in \mathbb{Z},$$

as a model, with V a nonconstant real-analytic function on $\mathbb{R}/\mathbb{Z}$, and $\tilde{\alpha}$ satisfying a certain Diophantine condition. It is shown that, if ϵ is sufficiently small, then for a.e. $x \in \mathbb{R}/\mathbb{Z}$, dynamical localization is maintained for "typical" solutions in a quasi-periodic time-dependent way.

Contents

1 Introduction **2**

2 Preliminaries **5**
2.1 The quasi-periodic Schrödinger operator 6
2.2 Decay property of matrices 7

3 Abstract KAM theorem **8**
3.1 Function space norms 8
3.2 Statement of the abstract KAM theorem 9
3.3 Application to Equation (1.9) 11

4 KAM step **13**
4.1 Construction of $\mathcal{O}_{\nu+1}$ 14
4.2 Homological equation and its approximate solution 16
4.3 Verification of assumptions after one sub-step 23
4.4 A succession of symplectic transformations 27

5 Proof of the KAM theorem **28**
5.1 Iteration lemma 29
5.2 Convergence 31
5.3 Measure estimate 31

A Appendix **32**
A.1 Outline of the proof of Proposition 1 32
A.2 Hamiltonian vector field and Poisson bracket 36

1 Introduction

Physical motivations. Localization of particles and waves in disordered media is one of the most intriguing phenomena in solid-state physics. This phenomenon was first analyzed by P.W.Anderson[1]. He studied the transport of non-interacting electrons in a crystal lattice, described by a single particle with random on-site energy. In his model, he showed that when the amplitude of the disorder becomes higher than a critical value, the diffusion in the lattice of an initially localized wavepacket is suppressed. An Anderson localized state is characterized by an exponential decay of the amplitude of the wave function.

In many physics experiments, a relatively weak disorder on the structure of the lattice is introduced by a quasi-periodic potential. This kind of system corresponds to an experimental realization of the so-called Aubry-André[2] or Harper[19] model. It is important in the study of Bose-Einstein condensation and nonlinear optics. Anderson localization in such linear systems, especially in the one-dimensional case, has been thoroughly studied[31], and rigorous mathematical results have been established[22]. As a well-known model in mathematical physics, the almost Mathieu operator $H_{x,\lambda,\tilde{\alpha}}$ acting on $\ell^2(\mathbb{Z})$ is defined by

$$(H_{x,\lambda,\tilde{\alpha}}\psi)_n = (\psi_{n+1} + \psi_{n-1}) + \lambda\cos 2\pi(x + n\tilde{\alpha})\psi_n, \quad n \in \mathbb{Z},$$

where n is the primary lattice site index, $\tilde{\alpha}$ is some ratio between the wavenumbers of two lattices, $x \in \mathbb{R}/\mathbb{Z}$ is an arbitrary phase, and ψ_n is a complex variable whose modulus square gives the probability of finding a particle at the lattice site n. With $\tilde{\alpha}$ a fixed Diophantine number, for a.e. x and λ large enough, $H_{x,\lambda,\tilde{\alpha}}$ exhibits *dynamical localization*[16, 17], i.e., for any $\psi \in \ell^2(\mathbb{Z})$ with compact support and arbitrary $d > 0$,

$$\sup_t r^{(d)}(t) := \sup_t \sum_{n\in\mathbb{Z}} n^{2d} |(e^{\mathrm{i}H_{x,\lambda,\tilde{\alpha}}t}\psi)_n|^2 < \infty.$$

In particular, there exists a transparent transition between diffusion and localization for the almost Mathieu operator. From the perspective of spectrum theory, it is shown by Jitomirskaya[22] that, for a.e. x, $H_{x,\lambda,\tilde{\alpha}}$ has

1. $\lambda > 2$: only pure point spectrum with exponentially decaying eigenfunctions;
2. $\lambda = 2$: purely singular-continuous spectrum;
3. $\lambda < 2$: purely absolutely continuous spectrum.

There is a perfect agreement with this conclusion in some experiments(e.g., [20]). For $\tilde{\alpha} = \dfrac{\sqrt{5}-1}{2}$, with an initial $\delta-$function wavepacket, the asymptotic spreading of the wavepacket width $r^{(1)}(t)$ can be parametrized as $r^{(1)}(t) \sim t^\gamma$, and one finds three different regimes

1. $\lambda > 2$: localized regime, $\gamma = 0$;

2. $\lambda = 2$: sub-diffusive, $\gamma \sim 0.5$;

3. $\lambda < 2$: ballistic regime, $\gamma = 1$.

However, the situation is much less clear in the presence of interactions(nonlinearities). It strongly influences the possibility to observe the localization induced by disorder. One can start from the Gross-Pitaevskii(GP) equation[18, 27] in Hartree-Fock theory, and get a generalized Aubry-André model which includes an additional nonlinear term that represents the mean-field interaction. The Hamiltonian is

$$H = \sum_{n\in\mathbb{Z}} \left[(\psi_{n+1}\bar{\psi}_n + \bar{\psi}_{n+1}\psi_n) + \lambda \cos 2\pi(n\tilde{\alpha} + x)|\psi_n|^2 + \frac{1}{2}\beta|\psi_n|^4 \right],$$

and the equation of motion is generated by $\mathrm{i}\dot{\psi}_n = -\frac{\partial H}{\partial \bar{\psi}_n}$, yielding the nonlinear Schrödinger equation

$$\mathrm{i}\dot{\psi}_n + (\psi_{n+1} + \psi_{n-1}) + \lambda \cos 2\pi(n\tilde{\alpha} + x)\psi_n + \beta|\psi_n|^2\psi_n = 0, \quad n \in \mathbb{Z}, \tag{1.1}$$

that can be considered as the GP equation on a discretized lattice. Similar versions of a discretized GP equation have been already used to investigate the dynamics of condensates in different situations(see, for instance, [33]).

It is shown experimentally in [25] that, if the condensate initially occupies a single lattice site, i.e., a $\delta-$function $\psi_n(0) = \delta_{n,0}$, the dynamics of the interacting gas is dominated by self-trapping in a wide range of parameters, even for weak interaction. Conversely, if the diffusion starts from a Gaussian wavepacket of width σ, $\psi_n(0) = ce^{-\frac{n^2}{2\sigma^2}}$, then self-trapping is significantly suppressed and the destruction of localization by interaction is more easily observable.

The aim of the present work is to analyze localization in the quasi-periodic nonlinear dynamical systems, which are modeled by the discrete one-dimensional disordered nonlinear Schrödinger equations of the same form as (1.1). When the disorder is sufficiently large, a rigorous mathematical argument for the maintainability of localization is given in this paper, corresponding to the experimental conclusion in [25].

Related mathematical works. In the theory of mathematical physics, localization in disordered, nonlinear dynamical systems was initiated by Fröhlich-Spencer-Wayne[12](Similar work was also done by Pöschel[28] and Vittot-Bellissard[34]), who constructed infinite-dimensional, compact invariant tori for a large class of non-coupling systems

$$\mathrm{i}\dot{q}_n + V_n q_n + \sum_{m\in\mathbb{Z}} \epsilon_{mn}(q_m + \bar{q}_m)^2 q_n = 0, \quad n \in \mathbb{Z},$$

via the KAM techniques, where $\{V_n\}_{n\in\mathbb{Z}}$ are i.i.d. random variables, ϵ_{mn} are sufficiently small and vanish for $|m-n|$ large enough. Solutions which lie on such tori are localized for all times. Besides the conclusion, they raised the following conjecture in that paper.

Conjecture.[12] Consider the equation

$$\mathrm{i}\dot{q}_n + \epsilon(q_{n+1} + q_{n-1}) + V_n q_n + \delta|q_n|^2 q_n = 0, \quad n \in \mathbb{Z}, \tag{1.2}$$

with $\{V_n\}_{n\in\mathbb{Z}}$ i.i.d. random variables. If ϵ and δ are small enough, with the equation in a large class, then for "most" initial conditions ("Most", e.g., with respect to the uniform measure on finite-dimensional unit spheres.), $q(0) = (q_n(0))_{n\in\mathbb{Z}}$, of finite support, the solutions $q(t) = (q_n(t))_{n\in\mathbb{Z}}$ of (1.2) satisfy

$$\overline{\lim_{t\to\infty}} t^{-1} \sum_{n\in\mathbb{Z}} n^2 |q_n(t)|^2 = 0.$$

Recently, there are several breakthroughs on such problem. For a large class of equation in (1.2), Bourgain-Wang[7] constructed a quasi-periodic solution when ϵ, δ are sufficiently small. They also considered the slightly tempered equations[8], with the nonlinearity replaced by $\lambda_n|q_n|^2 q_n$, $|\lambda_n| < \epsilon(1+|n|)^{-\tau}$ for some small $\tau > 0$. By constructing symplectic transformations to create energy barriers, they proved that, if ϵ is sufficiently small, then the diffusion bound (i.e., the H^1 norm) of the solution grows polynomially with t almost surely. Moreover, a Nekhoroshev-type result about Equation(1.2) was given by Wang-Zhang [35], who proved the long time Anderson localization for arbitrary ℓ^2 initial data.

By establishing an abstract KAM theorem, Geng-Zhao[15] constructed small-amplitude time quasi-periodic solutions of the lattice Schrödinger equation

$$\mathrm{i}\dot{q}_n + \epsilon(q_{n+1} + q_{n-1}) + \tan \pi(x + n\tilde{\alpha}) q_n + \epsilon|q_n|^2 q_n = 0, \quad n \in \mathbb{Z}, \tag{1.3}$$

for most of $x \in \mathbb{R}/\mathbb{Z}$ if ϵ is sufficiently small and $\tilde{\alpha}$ is Diophantine. This is based on the work by Bellissard-Lima-Scoppola[4], which have studied the linear operator corresponding to Equation (1.3), the well-known Maryland model. The operator, which has dense point spectrum, describes media with no resonance, and this provides convenience for the KAM iteration.

Statement of the main result. Inspired by the conjecture in [12], we try to establish a nonlinear version of "dynamical localization" in the quasi-periodic potential case.

Consider the one-dimensional nonlinear Schrödinger equation

$$\mathrm{i}\dot{q}_n + \epsilon(q_{n+1} + q_{n-1}) + V(n\tilde{\alpha} + x) q_n + |q_n|^2 q_n = 0, \quad n \in \mathbb{Z}, \tag{1.4}$$

for $0 < \epsilon \ll 1$, with V a nonconstant real-analytic function on $\mathbb{R}/\mathbb{Z}$, and $\tilde{\alpha} \in \mathbb{R}$ is a Diophantine number, i.e., there exist $\tilde{\tau} > 1$ and $\tilde{\gamma} > 0$ such that ($|x|_1$ is the absolute value of x modulo 1 defined so that $0 \le |x|_1 \le \frac{1}{2}$.)

$$|n\tilde{\alpha}|_1 \ge \frac{\tilde{\gamma}}{|n|^{\tilde{\tau}}}, \quad n \ne 0. \tag{1.5}$$

The nonconstant real-analytic potential V, as in [11], is a smooth function in the Gevrey class

$$\sup_{x\in\mathbb{R}/\mathbb{Z}} |\partial^m V(x)| \le C L^m (m!)^2, \quad m \ge 0, \tag{1.6}$$

for some C, $L > 0$, and satisfying the transversality condition

$$\max_{0\le m\le \tilde{s}} |\partial_\varphi^m (V(x+\varphi) - V(x))| \ge \tilde{\xi} > 0, \quad \forall x, \forall \varphi, \tag{1.7}$$

$$\max_{0\le m\le \tilde{s}} |\partial_x^m (V(x+\varphi) - V(x))| \ge \tilde{\xi} |\varphi|_1, \quad \forall x, \forall \varphi, \tag{1.8}$$

for some $\tilde{\xi}$, $\tilde{s} > 0$. Clearly, the case $V(x) = \cos 2\pi x$ is included.

Based on an earlier KAM mechanism which was introduced by Eliasson[11], we construct an abstract KAM theorem, and apply this theorem to prove well-localization of Equation (1.4) for typical initial data. From the KAM perspective, the main technical challenges in this work are the following:

i) Unlike the model in [12], we need to tackle with the second order perturbation in the Hamiltonian;

ii) Different from the method in [7], our proof is developed from the traditional KAM method;

iii) Compared with the work in [15], the main difficulty is that the corresponding linear operator has dense point spectrum with *infinitely many resonances.*

The main result can be stated as follows.

Theorem 1 *Consider the one-dimensional nonlinear Schrödinger equation*

$$\mathrm{i}\dot{q}_n + \epsilon(q_{n+1} + q_{n-1}) + V(n\tilde{\alpha} + x)q_n + |q_n|^2 q_n = 0, \quad n \in \mathbb{Z}, \tag{1.9}$$

with V a nonconstant real-analytic function on $\mathbb{R}/\mathbb{Z}$, and $\tilde{\alpha} \in \mathbb{R}$ a Diophantine number. Given an integer $b > 1$, and any $\mathcal{J} = \{n_1, \cdots, n_b\} \subset \mathbb{Z}$, there exists a sufficiently small $\epsilon_* = \epsilon_*(V, \tilde{\alpha}, \mathcal{J})$, such that if $0 < \epsilon < \epsilon_*$, then the following holds for a.e. $x \in \mathbb{R}/\mathbb{Z}$.

There exists a Cantor set $\mathcal{O}_\epsilon = \mathcal{O}_\epsilon(x) \subset [0,1]^b$ with $|[0,1]^b \setminus \mathcal{O}_\epsilon| \to 0$ [1] as $\epsilon \to 0$ such that the solution $q(t) = (q_n(t))_{n\in\mathbb{Z}}$ of Equation (1.9), with initial datum $q(0) \in \mathcal{O}_\epsilon$ supported on $\mathcal{J}$, satisfies, for any fixed $d > 0$,

$$\sup_t \sum_{n\in\mathbb{Z}} n^{2d} |q_n(t)|^2 < \infty.$$

Moreover, for each $n \in \mathbb{Z}$, $q_n(t)$ is quasi-periodic in time.

Remark 1.1 *The quasi-periodic solutions we obtained are not necessarily small-amplitude, since the nonlinearity $|q_n|^2 q_n$ is integrable. Moreover, if the nonlinearity is "diagonal dominant" with some short-range decay, e.g.,*

$$|q_n|^2 q_n + \epsilon \sum_{m\neq n} e^{-\varrho|m-n|} |q_m|^2 q_n,$$

the theorem above can also be obtained.

Remark 1.2 *Smallness assumption on ϵ is necessary, otherwise the result is not true even for the linear problem. This is different from the random potential case.*

2 Preliminaries

From now on, in the formulations and proofs of various assertions, we shall encounter absolute constants depending on the Hamiltonian, the dimension and so on. All such constants will be denoted by $c, c_1, c_2, \cdots$, and sometimes even different constants will be denoted by the same symbol.

[1] Hereafter, we use the symbol $|\mathcal{O}|$ to denote the Lebesgue measure of $\mathcal{O} \subset \mathbb{R}^b$.

2.1 The quasi-periodic Schrödinger operator

Consider the Schrödinger operator $T = T(x) : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ defined as

$$(Tq)_n := \epsilon(q_{n+1} + q_{n-1}) + V(x + n\tilde{\alpha})q_n, \quad n \in \mathbb{Z}, \tag{2.1}$$

with V and $\tilde{\alpha}$ as in Equation (1.9). It is well-known from [11] that if ϵ is sufficiently small, then for a.e. $x \in \mathbb{R}/\mathbb{Z}$, the spectrum of $T(x)$ is pure point. We refer to [6, 9, 13, 21, 23, 32] for other works on the pure point spectrum and Anderson localization of quasi-periodic Schrödinger operators.

Now, we are going to represent the main idea of [11], which is critical for the KAM iteration in this paper. Let us start with some notations for infinite-dimensional matrices. Given an infinite-dimensional matrix D, with $D_{mn} \in \mathbb{R}$ the $(m,n)^{\text{th}}$ entry, for a subset $\Lambda \subset \mathbb{Z}$, we define $\Lambda^\perp := \mathbb{Z} \setminus \Lambda$,

$$\mathbb{R}^\Lambda := \{n \in \mathbb{R}^{\mathbb{Z}} : n_i = 0 \ \text{ if } \ i \notin \Lambda\}, \quad D_\Lambda := \begin{cases} D_{mn}, & m, n \in \Lambda \\ \delta_{mn}, & \text{otherwise} \end{cases}.$$

Then $D_\Lambda : \mathbb{R}^\Lambda + \mathbb{R}^{\Lambda^\perp} \to \mathbb{R}^\Lambda + \mathbb{R}^{\Lambda^\perp}$, acts as $\mathbb{R}^\Lambda \hookrightarrow \mathbb{R}^{\mathbb{Z}} \xrightarrow{D} \mathbb{R}^{\mathbb{Z}} \xrightarrow{\perp \text{proj}} \mathbb{R}^\Lambda$ on the first component and as the identity on the second component. (When there is no risk for confusion, we will use D_Λ also to denote its first component.)

Let $D_0 = \operatorname{diag}\{V(x + n\tilde{\alpha})\}_{n\in\mathbb{Z}}$ and $Z_0 = \epsilon\Delta$ with Δ the discrete Laplacian. With $\varepsilon_0 = \epsilon^{\frac{1}{4}}$, $\sigma_0 = 1$ and any

$$M_0 \geq \max\left\{2^{\tilde{s}+4} C \frac{L^{\tilde{s}+1}((\tilde{s}+1)!)^2}{\tilde{\xi}}, \, 2\tilde{\tau}, \, 8\right\}, \quad N_0 \geq 1, \quad \rho_0 = N_0^{-1},$$

one can define the following sequences as in [11],

$$\begin{aligned} M_{\nu+1} &= M_\nu^{\tilde{s}M_\nu^3}, \quad a_\nu = \frac{1}{\tilde{\tau}} M_\nu^{-3\tilde{s}M_\nu^3}, \quad \varepsilon_{\nu+1} = \varepsilon_\nu^{\frac{1}{2}\varepsilon_\nu^{-a_\nu/2}}, \\ N_{\nu+1} &= \varepsilon_\nu^{-a_\nu}, \qquad \rho_{\nu+1} = \varepsilon_\nu^{a_\nu}, \qquad \sigma_{\nu+1} = \frac{1}{3}\rho_\nu. \end{aligned} \tag{2.2}$$

These sequences of parameters will be applied in the KAM iteration in this paper.

Proposition 1 *There exists a constant $\epsilon_0 = \epsilon_0(C, L, \tilde{\xi}, \tilde{s}, \tilde{\gamma}, \tilde{\tau})$ such that the following holds for the operator (2.1) if $0 < \epsilon < \epsilon_0$.*

Fix any $x \in \mathbb{R}/\mathbb{Z}$. There exists a sequence of orthogonal matrices U_ν, $\nu = 1, 2, \cdots$, with

$$|(U_\nu - I_{\mathbb{Z}})_{mn}| \leq \varepsilon_0^{\frac{1}{2}} e^{-\frac{3}{2}\sigma_\nu |m-n|},$$

such that $U_\nu^*(D_0 + Z_0)U_\nu = D_\nu + Z_\nu$, where Z_ν is a symmetric matrix satisfying

$$|(Z_\nu)_{mn}| \leq \varepsilon_\nu e^{-\rho_\nu |m-n|},$$

and D_ν is a symmetric matrix which can be block-diagonalized via an orthogonal matrix Q_ν with

$$(Q_\nu)_{mn} = 0 \ \text{ if } \ |m-n| > N_\nu.$$

More precisely, there is a disjoint decomposition $\bigcup_j \Lambda_j^\nu = \mathbb{Z}$ *such that*

$$\tilde{D}^\nu = Q_\nu^* D_\nu Q_\nu = \prod_j \tilde{D}^\nu_{\Lambda_j^\nu} \quad \text{with} \quad \sharp\Lambda_j^\nu \le M_\nu, \quad \mathrm{diam}\Lambda_j^\nu \le M_\nu N_\nu, \quad \forall j.^2$$

Moreover, there exists a full-measure subset $\tilde{\mathcal{X}} \subset \mathbb{R}/\mathbb{Z}$ *such that if we fix* $x \in \tilde{\mathcal{X}}$*, then for each* $k \in \mathbb{Z}$*, there is a* $\nu_0(k)$ *such that*

$$\Lambda^{\nu+1}(k) = \Lambda^\nu(k), \quad \forall \nu \ge \nu_0(k).$$

Proposition 1 shows the pure point spectrum of T. In Appendix A.1, we shall give an outline of the proof.

2.2 Decay property of matrices

Lemma 2.1 *Given two matrices* $G = (G_{mn})_{m,n\in\mathbb{Z}}$ *and* $F = (F_{mn})_{m,n\in\mathbb{Z}}$*. Let* $K = GF$*.*

(1) *If* $|G_{mn}| \le c_G e^{-\sigma_G|m-n|}$, $|F_{mn}| \le c_F e^{-\sigma_F|m-n|}$ *for some positive* c_G, c_F, σ_G, $\sigma_F > 0$, *then we have*

$$|K_{mn}| \le c_K e^{-\sigma_K|m-n|}$$

for any $0 < \sigma_K < \min\{\sigma_G,\, \sigma_F\}$ *and* $c_K = c \cdot c_G c_F(\min\{\sigma_G,\, \sigma_F\} - \sigma_K)^{-1}$.

(2) *If* $|G_{mn}| \le c_G e^{-\sigma_G \max\{|m|,\,|n|\}}$, $|F_{mn}| \le c_F e^{-\sigma_F|m-n|}$*, then*

$$|K_{mn}| \le c_K e^{-\sigma_K \max\{|m|,\,|n|\}}.$$

(3) *If* $|G_{mn}| \le c_G e^{-\sigma_G|m-n|}$, $|F_{mn}| \le c_F e^{-\sigma_F \max\{|m|,\,|n|\}}$*, then*

$$|K_{mn}| \le c_K e^{-\sigma_K \max\{|m|,\,|n|\}}.$$

(4) *If* $|G_{mn}| \le c_G e^{-\sigma_G \max\{|m|,\,|n|\}}$, $|F_{mn}| \le c_F e^{-\sigma_F \max\{|m|,\,|n|\}}$*, then*

$$|K_{mn}| \le c_K e^{-\sigma_K \max\{|m|,\,|n|\}}.$$

In particular, if $\sigma_G \ne \sigma_F$*, then the conclusions above hold with* $\sigma_K = \min\{\sigma_G,\, \sigma_F\}$ *and* $c_K = c \cdot c_G c_F |\sigma_G - \sigma_F|^{-1}$*.*

Proof: Since the matrix element of $K = GF$ can be formulated as $K_{mn} = \sum_{l\in\mathbb{Z}} G_{ml}F_{ln}$, we have that, in Case (1), for any $0 < \sigma_K < \min\{\sigma_G, \sigma_F\}$,

$$\begin{aligned}
|(GF)_{mn}| &\le \sum_{l\in\mathbb{Z}} |G_{ml}||F_{ln}| \\
&\le c_G c_F \sum_{l\in\mathbb{Z}} e^{-\sigma_G|m-l|} e^{-\sigma_F|l-n|} \\
&\le c_G c_F e^{-\sigma_K|m-n|} \sum_{l\in\mathbb{Z}} e^{-(\sigma_G-\sigma_K)|m-l|} e^{-(\sigma_F-\sigma_K)|l-n|} \\
&\le c \cdot c_G c_F (\min\{\sigma_G,\, \sigma_F\} - \sigma_K)^{-1} e^{-\sigma_K|m-n|}.
\end{aligned}$$

[2]The disjoint decomposition defines an equivalence relation $m \sim n$ on the integers and, for each $n \in \mathbb{Z}$, we denote its equivalence class by $\Lambda^\nu(n)$.

Here we have applied the basic triangular inequality $|m-l|+|l-n|\geq|m-n|$.

Moreover, if $\sigma_G\neq\sigma_F$, assume that $0<\sigma_G<\sigma_F$ without loss of generality, then

$$\begin{aligned}|(GF)_{mn}| &\leq c_Gc_F\sum_{l\in\mathbb{Z}}e^{-\sigma_G|m-l|}e^{-\sigma_F|l-n|}\\ &\leq c_Gc_Fe^{-\sigma_G|m-n|}\sum_{l\in\mathbb{Z}}e^{-(\sigma_F-\sigma_G)|l-n|}\\ &\leq c\cdot c_Gc_F(\sigma_F-\sigma_G)^{-1}e^{-\sigma_G|m-n|}.\end{aligned}$$

As for Case (2)$-$(4), the corresponding conclusions can also be obtained by using the trivial facts

$$|m-l|+\max\{|l|,\,|n|\}\geq\max\{|m|,\,|n|\},\quad \max\{|m|,\,|l|\}+\max\{|l|,\,|n|\}\geq\max\{|m|,\,|n|\}.$$

Thus Lemma 2.1 has been proved. ■

Remark 2.1 *If we replace the matrix F satisfying $|F_{mn}|\leq c_Fe^{-\sigma_F\max\{|m|,\,|n|\}}$ with a vector $f=(f_n)_{n\in\mathbb{Z}}$ satisfying $|f_n|\leq c_fe^{-\sigma_f|n|}$ in Case (3) and (4), then for the vector Gf, we can obtain the conclusion that $|(Gf)_n|\leq c_Ke^{-\sigma_K|n|}$, where c_K and σ_K are the same as in the lemma.*

3 Abstract KAM theorem

3.1 Function space norms

Given $d,\rho>0$, let $\ell^1_{d,\rho}(\mathbb{Z})$ be the space of summable complex valued sequences $q=(q_n)_{n\in\mathbb{Z}}$, with the norm

$$\|q\|_{d,\rho}:=\sum_{n\in\mathbb{Z}}|q_n|\langle n\rangle^de^{\rho|n|}<\infty,$$

where $\langle n\rangle:=\sqrt{1+n^2}$. For $r,s>0$, let $\mathcal{D}_{d,\rho}(r,s)$ be the complex b-dimensional neighborhood of $\mathbb{T}^b\times\{0\}\times\{0\}\times\{0\}$ in $\mathbb{T}^b\times\mathbb{R}^b\times\ell^1_{d,\rho}(\mathbb{Z})\times\ell^1_{d,\rho}(\mathbb{Z})$, i.e.,

$$\mathcal{D}_{d,\rho}(r,s):=\{(\theta,I,q,\bar{q}):|\mathrm{Im}\theta|=|\mathrm{Im}(\theta_1,\cdots,\theta_b)|<r,\ |I|<s^2,\ \|q\|_{d,\rho}=\|\bar{q}\|_{d,\rho}<s\},$$

where $|\cdot|$ is the ℓ^1-norm of complex vectors.

Given a real-analytic function $F(\theta,I,q,\bar{q};\xi)$ on $\mathcal{D}=\mathcal{D}_{d,\rho}(r,s)$, C^1_W(i.e., C^1 in the sense of Whitney) parametrized by $\xi\in\mathcal{O}$ [3], a closed region in $\mathbb{R}^b$. We expand F into the Taylor-Fourier series with respect to $\theta,I,q,\bar{q}$:

$$F(\theta,I,q,\bar{q};\xi)=\sum_{\alpha,\beta}F_{\alpha\beta}(\theta,I;\xi)q^\alpha\bar{q}^\beta,$$

where, for multi-indices $\alpha:=\sum_{n\in\mathbb{Z}}\alpha_ne_n$, $\beta:=\sum_{n\in\mathbb{Z}}\beta_ne_n$, $\alpha_n,\beta_n\in\mathbb{N}$, with finitely many non-vanishing components,

$$F_{\alpha\beta}(\theta,I;\xi)=\sum_{k\in\mathbb{Z}^b,\,l\in\mathbb{N}^b}F_{kl\alpha\beta}(\xi)I^le^{\mathrm{i}\langle k,\theta\rangle},\quad q^\alpha\bar{q}^\beta=\prod_{(\alpha_n,\beta_n)\neq(0,0)}q_n^{\alpha_n}\bar{q}_n^{\beta_n}.$$

[3]In this paper, all dependencies on the parameter $\xi\in\mathcal{O}$ are assumed of class C^1_W, thus all derivatives with respective to ξ will be interpreted in this sense.

(Here e_m denotes the vector with the m^{th} component being 1 and the other components being zero.)

Definition 3.1 *For each non-zero multi–index* $(\alpha, \beta) = (\cdots, \alpha_n, \beta_n, \cdots)_{n\in\mathbb{Z}}$, $\alpha_n, \beta_n \in \mathbb{N}$, *with finitely many non-vanishing components, we define*

$$\begin{aligned} n^+_{\alpha\beta} &:= \max\{n \in \mathbb{Z} : (\alpha_n, \beta_n) \neq (0,0)\}, \\ n^-_{\alpha\beta} &:= \min\{n \in \mathbb{Z} : (\alpha_n, \beta_n) \neq (0,0)\}, \\ n^*_{\alpha\beta} &:= \max\{|n^+_{\alpha\beta}|, |n^-_{\alpha\beta}|\}, \end{aligned}$$

and $|\alpha| := \sum_n \alpha_n$, $|\beta| := \sum_n \beta_n$.

In particular, for $|\alpha| = |\beta| = 0$, *we define* $n^+_{\alpha\beta} = n^-_{\alpha\beta} = n^*_{\alpha\beta} = 0$.

With $|\partial_\xi F_{kl\alpha\beta}| := \sum_{i=1}^b |\partial_{\xi_i} F_{kl\alpha\beta}|$ and $|F_{kl\alpha\beta}|_{\mathcal{O}} := \sup_{\xi\in\mathcal{O}} (|F_{kl\alpha\beta}| + |\partial_\xi F_{kl\alpha\beta}|)$, let

$$\|F_{\alpha\beta}\|_{\mathcal{O}} := \sum_{k,l} |F_{kl\alpha\beta}|_{\mathcal{O}}\ |I^l| e^{|k||\mathrm{Im}\theta|}, \quad \|F\|_{\mathcal{O}} := \sum_{k,l,\alpha,\beta} |F_{kl\alpha\beta}|_{\mathcal{O}}\ |I^l| e^{|k||\mathrm{Im}\theta|}\, |q^\alpha||\bar{q}^\beta|.$$

Define the weighted norm of F as

$$\|F\|_{\mathcal{D},\mathcal{O}} := \sup_{\mathcal{D}} \|F\|_{\mathcal{O}}.^4$$

For the Hamiltonian vector field $X_F = (\partial_I F, -\partial_\theta F, (-\mathrm{i}\partial_{q_n} F)_{n\in\mathbb{Z}}, (\mathrm{i}\partial_{\bar{q}_n} F)_{n\in\mathbb{Z}})$ associated F on $\mathcal{D}\times\mathcal{O}$, its norm is defined by

$$\|X_F\|_{\mathcal{D},\mathcal{O}} := \|\partial_I F\|_{\mathcal{D},\mathcal{O}} + \frac{1}{s^2}\|\partial_\theta F\|_{\mathcal{D},\mathcal{O}} + \sup_{\mathcal{D}} \frac{1}{s} \sum_{n\in\mathbb{Z}} (\|\partial_{q_n} F\|_{\mathcal{O}} + \|\partial_{\bar{q}_n} F\|_{\mathcal{O}}) \langle n\rangle^d e^{|n|\rho}.$$

Sometimes, for the sake of notational simplification, we shall not write the subscript $\mathcal{O}$ in the norms defined above if it is obvious enough.

Given two real-analytic functions F and G on $\mathcal{D}$, let $\{\cdot,\cdot\}$ denote the Poisson bracket of such functions, i.e.,

$$\{F, G\} = \langle \partial_I F, \partial_\theta G\rangle - \langle \partial_\theta F, \partial_I G\rangle + \mathrm{i}\sum_{n\in\mathbb{Z}} (\partial_{q_n} F \cdot \partial_{\bar{q}_n} G - \partial_{\bar{q}_n} F \cdot \partial_{q_n} G).$$

Some basic estimates about the Hamiltonian vector field and the Poisson bracket are given in Appendix A.2.

3.2 Statement of the abstract KAM theorem

Now, we consider the perturbed Hamiltonian

$$\begin{aligned} H &= \mathcal{N} + \breve{P} + P \\ &= e(x,\xi) + \langle \omega(x,\xi), I\rangle + \langle \Omega(x,\xi) q, \bar{q}\rangle + \breve{P}(q,\bar{q}; x) + P(\theta, I, q, \bar{q}; x, \xi), \end{aligned} \tag{3.1}$$

defined on the domain $\mathcal{D} = \mathcal{D}_{d,\rho}(r,s)$. Our goal is to prove that, for a.e. $x \in \mathbb{R}/\mathbb{Z}$, the Hamiltonian H admits invariant tori for "most" of the parameter $\xi \in \mathcal{O} = \mathcal{O}(x)$, provided that $\|X_{\breve{P}+P}\|_{\mathcal{D},\mathcal{O}}$ is sufficiently small.

[4] In the case of a vector-valued function $F : \mathcal{D}\times\mathcal{O} \to \mathbb{C}^n (n < \infty)$, the norm is defined as $\|F\|_{\mathcal{D},\mathcal{O}} := \sum_{i=1}^n \|F_i\|_{\mathcal{D},\mathcal{O}}$.

Remark 3.1 *From now on, we shall not report x for convenience if it is irrelevant.*

To this end, we need to impose some conditions on ω, Ω, and the perturbations $\breve{P}+P$.

(A1) *Nondegeneracy of tangential frequencies*: The map $\xi \to \omega$ is a C^1_W diffeomorphism between $\mathcal{O}$ and its image.

(A2) *Regularity of* Ω: $\Omega = T + A + W$.

- T is the symmetric matrix defined in (2.1), independent of ξ. More precisely,

$$T = \mathrm{diag}\{V(x+n\tilde{\alpha})\}_{n\in\mathbb{Z}} + \epsilon\Delta,$$

with V and $\tilde{\alpha}$ as in Equation (1.9).

- A is Hermitian, independent of ξ, satisfying

$$|A_{mn}| \le \begin{cases} c, & |m|,|n| \le \hat{N} \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

for some positive $\hat{N}$.

- W is C^1_W parametrized by $\xi \in \mathcal{O}$, with

$$|W_{mn}|_{\mathcal{O}} \le \begin{cases} pe^{-\sigma \max\{|m|,|n|\}}, & |m|,|n| \le N \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

for some positive $p \ll 1$, $\sigma \gg \rho$ and sufficiently large N.

Moreover, there exists a subset $\mathcal{J} \subset \mathbb{Z}$ such that

$$\Omega_{mn} \equiv 0 \quad \text{if} \quad m \quad \text{or} \quad n \in \mathcal{J}. \tag{3.4}$$

(A3) *Short range of* $\breve{P}$: $\breve{P}(q,\bar{q}) = \sum_{|\alpha|=|\beta|\ge 2} \breve{P}_{\alpha\beta} q^\alpha \bar{q}^\beta$ is real-analytic in $q, \bar{q}$, and independent of ξ, with

$$|\breve{P}_{\alpha\beta}| \le e^{-\rho(n^+_{\alpha\beta} - n^-_{\alpha\beta})}, \quad |\alpha| = |\beta| \ge 2, \tag{3.5}$$

$$\partial_{q_n}\breve{P} = \partial_{\bar{q}_n}\breve{P} \equiv 0, \quad \forall n \in \mathcal{J}. \tag{3.6}$$

(A4) *Decay property of* P: $P = \sum_{\alpha,\beta} P_{\alpha\beta}(\theta, I; \xi) q^\alpha \bar{q}^\beta$ is real-analytic in $\theta, I, q, \bar{q}$, C^1_W parametrized by $\xi \in \mathcal{O}$, and, with $\varepsilon = \epsilon^{\frac{1}{4}}$,

$$\|P_{\alpha\beta}\|_{\mathcal{D},\mathcal{O}} \le \begin{cases} \varepsilon e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta| \le 2 \\ e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta| \ge 3 \end{cases}, \tag{3.7}$$

$$\partial_{q_n} P = \partial_{\bar{q}_n} P \equiv 0, \quad \forall n \in \mathcal{J}. \tag{3.8}$$

(A5) *Gauge invariance of* P: For $P = \sum_{\substack{k\in\mathbb{Z}^b, l\in\mathbb{N}^b \\ \alpha,\beta}} P_{kl\alpha\beta} I^l e^{\mathrm{i}\langle k,\theta\rangle} q^\alpha \bar{q}^\beta$, we have

$$P_{kl\alpha\beta} \equiv 0 \quad \text{if} \quad \sum_{j=1}^{b} k_j + |\alpha| - |\beta| \ne 0.$$

Our abstract KAM theorem can be stated as follows.

Theorem 2 *Consider the Hamiltonian H in (3.1), with* $(\mathbf{A1})-(\mathbf{A5})$ *satisfied. There is a positive constant $\varepsilon_*=\varepsilon_*(\omega,V,\tilde{\alpha},\hat{N},p,\sigma,N,r,s,d,\rho)$ such that if $\|X_{\check{P}+P}\|_{\mathcal{D},\mathcal{O}}\le\varepsilon\le\varepsilon_*$, then for every $x\in\tilde{\mathcal{X}}$, there exists a Cantor set $\mathcal{O}_\varepsilon=\mathcal{O}_\varepsilon(x)\subset\mathcal{O}(x)$ with $|\mathcal{O}\setminus\mathcal{O}_\varepsilon|\to 0$ as $\varepsilon\to 0$, such that the following holds.*

(a) *There exists a C^1_W map $\tilde{\omega}:\mathcal{O}_\varepsilon\to\mathbb{R}^b$, such that $|\tilde{\omega}-\omega|_{\mathcal{O}_\varepsilon}\to 0$ as $\varepsilon\to 0$.*

(b) *There exists a map $\Psi:\mathbb{T}^b\times\mathcal{O}_\varepsilon\to\mathcal{D}_{d,0}(r/2,0)$, real-analytic in $\theta\in\mathbb{T}^b$ and C^1_W parametrized by $\xi\in\mathcal{O}$, such that $\|\Psi-\Psi_0\|_{\mathcal{D}_{d,0}(r/2,0),\mathcal{O}_\varepsilon}\to 0$ as $\varepsilon\to 0$, where Ψ_0 is the trivial embedding: $\mathbb{T}^b\times\mathcal{O}\to\mathbb{T}^b\times\{0\}\times\{0\}\times\{0\}$.*

(c) *For any $\theta\in\mathbb{T}^b$ and $\xi\in\mathcal{O}_\varepsilon$, $\Psi(\theta+\tilde{\omega}(\xi)t,\xi)=(\theta+\tilde{\omega}(\xi)t,I(t),q(t),\bar{q}(t))$ is a b-frequency quasi-periodic solution of equations of motion associated with the Hamiltonian (3.1).*

(d) *For each t, $q(t)=(q_n(t))_{n\in\mathbb{Z}}\in\ell^1_{d,0}(\mathbb{Z})$.*

Remark 3.2 *The statement* (d) *of Theorem 2 implies that*

$$\sup_t\sum_{n\in\mathbb{Z}}n^{2d}|q_n(t)|^2<c\left(\sup_t\sum_{n\in\mathbb{Z}}\langle n\rangle^d|q_n(t)|\right)^2<\infty,$$

which is exactly the conclusion of Theorem 1.

Remark 3.3 *In case that H satisfies* $(\mathbf{A1})-(\mathbf{A5})$ *at the first step, all assumptions hold for the Hamiltonian at each KAM step (with suitable parameters).*

3.3 Application to Equation (1.9)

The Hamiltonian associated with Equation (1.9) is

$$H=\sum_{n\in\mathbb{Z}}V(x+n\tilde{\alpha})q_n\bar{q}_n+\epsilon\sum_{n\in\mathbb{Z}}\bar{q}_n(q_{n+1}+q_{n-1})+\frac{1}{2}\sum_{n\in\mathbb{Z}}|q_n|^4. \tag{3.9}$$

Fix $\mathcal{J}=\{n_1,\cdots,n_b\}\subset\mathbb{Z}$, and $\mathbb{Z}_1=\mathbb{Z}\setminus\mathcal{J}$. Let $\varepsilon=\epsilon^{\frac{1}{4}}$, with ϵ sufficiently small such that

$$|n_i|\le|\ln\varepsilon|=\frac{1}{4}|\ln\epsilon|,\quad i=1,\cdots,b.$$

We introduce action-angle variables and amplitude parameters to the Hamiltonian (3.9),

$$q_n=\sqrt{I_n+\xi_n}e^{\mathrm{i}\theta_n},\quad \bar{q}_n=\sqrt{I_n+\xi_n}e^{-\mathrm{i}\theta_n},\quad n\in\mathcal{J},$$

where $(I,\theta)=(I_{n_1},\cdots,I_{n_b},\theta_{n_1},\cdots,\theta_{n_b})$ are the standard action-angle variables in the $(q_n,\bar{q}_n)_{n\in\mathcal{J}}$-space around ξ, with $\xi=(\xi_{n_1},\cdots,\xi_{n_b})\in\mathcal{O}=[\epsilon^{\frac{1}{12}},1]\subset[0,1]^b$ a parameter, and $(q,\bar q)=(q_n,\bar q_n)_{n\in\mathbb{Z}_1}$. Then the Hamiltonian (3.9) becomes $H=\mathcal{N}(\theta,I,q,\bar q;x,\xi)+\breve P(q,\bar q)+P(\theta,I,q,\bar q;\xi)$, with

$$\begin{aligned}\mathcal{N}(\theta,I,q,\bar q;x,\xi) &:= \sum_{n\in\mathcal{J}}(V(x+n\tilde\alpha)\xi_n+\frac{1}{2}\xi_n^2)+\sum_{n\in\mathcal{J}}(V(x+n\tilde\alpha)+\xi_n)I_n\\ &\quad+\sum_{n\in\mathbb{Z}_1}V(x+n\tilde\alpha)|q_n|^2+\epsilon\sum_{\substack{n\notin\mathcal{J}\\ n+1\notin\mathcal{J},}}(\bar q_nq_{n+1}+q_n\bar q_{n+1}),\\ \breve P(q,\bar q) &:= \frac{1}{2}\sum_{n\in\mathbb{Z}_1}|q_n|^4,\\ P(\theta,I,q,\bar q;\xi) &:= \frac{1}{2}\sum_{n\in\mathcal{J}}I_n^2+\epsilon\sum_{\substack{m\in\mathcal{J},\,n\notin\mathcal{J}\\ |m-n|=1}}\sqrt{I_m+\xi_m}(e^{-\mathrm{i}\theta_m}q_n+e^{\mathrm{i}\theta_m}\bar q_n)\\ &\quad+\epsilon\sum_{\substack{m,\,n\in\mathcal{J}\\ |m-n|=1}}\sqrt{I_m+\xi_m}\sqrt{I_n+\xi_n}(e^{-\mathrm{i}(\theta_m-\theta_n)}+e^{\mathrm{i}(\theta_m-\theta_n)}).\end{aligned}$$

After introducing the action-angle variables, we find that the structure of the linear operator T in (3.9) has been destroyed. To overcome this disadvantage, we need to add b variables $q'_{n_1},\cdots,q'_{n_b}$ and the corresponding conjugates $\bar q'_{n_1},\cdots,\bar q'_{n_b}$ into this system. For convenience, omit the prime of the newly-added variables and still use q to denote $(q_n)_{n\in\mathbb{Z}}$, since there is no confusion. We then rewrite $\mathcal{N}$ as

$$\begin{aligned}\mathcal{N} &= \sum_{n\in\mathcal{J}}(V(x+n\tilde\alpha)\xi_n+\frac{1}{2}\xi_n^2)+\sum_{n\in\mathcal{J}}(V(x+n\tilde\alpha)+\xi_n)I_n\\ &\quad+\left[\sum_{n\in\mathbb{Z}_1}V(x+n\tilde\alpha)|q_n|^2+\sum_{n\in\mathcal{J}}V(x+n\tilde\alpha)|q_n|^2+\epsilon\sum_{n\in\mathbb{Z}}(\bar q_nq_{n+1}+q_n\bar q_{n+1})\right]\\ &\quad-\sum_{n\in\mathcal{J}}V(x+n\tilde\alpha)|q_n|^2-\epsilon\sum_{n\text{ or }n+1\in\mathcal{J}}(\bar q_nq_{n+1}+q_n\bar q_{n+1})\\ &= e(x,\xi)+\langle\omega(x,\xi),I\rangle+\langle T(x)q,\bar q\rangle+\langle A(x)q,\bar q\rangle,\end{aligned}$$

with $e(x,\xi):=\sum_{n\in\mathcal{J}}V(x+n\tilde\alpha)\xi_n+\frac{1}{2}\sum_{n\in\mathcal{J}}\xi_n^2$,

$$\omega(x,\xi) := (V(x+n_1\tilde\alpha)+\xi_{n_1},\,\cdots,\,V(x+n_b\tilde\alpha)+\xi_{n_b}), \tag{3.10}$$

$$T_{mn}(x) := \begin{cases} V(x+m\tilde\alpha), & m=n\\ \epsilon, & m-n=\pm1\\ 0, & \text{otherwise}\end{cases}, \tag{3.11}$$

$$A_{mn}(x) := \begin{cases} -V(x+m\tilde\alpha), & m=n,\ \ m\in\mathcal{J}\\ -\epsilon, & m-n=\pm1,\ \ m\ \text{ or }\ n\in\mathcal{J}\\ 0, & \text{otherwise}\end{cases}. \tag{3.12}$$

Now, on some domain $\mathcal{D}_{d,\rho}(r,s)$, the regularity of $\breve P+P$ holds true:

Lemma 3.1 *For $\epsilon>0$ sufficiently small and $s=\frac{1}{8}\epsilon^{\frac14}$, if $|I|<s^2$ and $\|q\|_{d,\rho}<s$, then*

$$\|X_{\breve P+P}\|_{\mathcal{D}_{d,\rho}(r,s),\mathcal{O}}\le\epsilon^{\frac14}=\varepsilon.$$

We need to show that the Hamiltonian $H = \mathcal{N} + \check{P} + P$ satisfies the assumptions (**A1**) – (**A5**) of the KAM theorem, in which **(A3)** and **(A5)** are obviously satisfied.

(**A1**)**:** Since $\{V(x+n\tilde{\alpha})\}_{n\in\mathbb{Z}}$ is independent of ξ, we have that $\frac{\partial\omega}{\partial\xi} \equiv I_{\mathcal{J}}$ in view of (3.10). Thus (**A1**) holds.

(**A2**)**:** Here, $W \equiv 0$. Then it is evident that (**A2**) holds with $\hat{N} = \frac{1}{4}|\ln\epsilon|$, by (3.12).

(**A4**)**:** We focus on the formulation of P. Note that terms of P merely correspond to the normal variables q_n, $\bar{q}_n$, $n \notin \mathcal{J}$, $n-1$ or $n+1 \in \mathcal{J}$, with the coefficients no more than ϵ, and $\mathcal{J} \subset [-\hat{N}, \hat{N}] = [-\frac{1}{4}|\ln\epsilon|, \frac{1}{4}|\ln\epsilon|]$. Then, with $\rho \le \frac{1}{6}\hat{N}^{-1}$, (3.7) is verified since

$$c\epsilon^{1-\frac{1}{24}} \le \epsilon^{\frac{1}{4}} e^{-\rho\hat{N}}.$$

Hence, Theorem 1 is a corollary of Theorem 2.

4 KAM step

To start the KAM iteration for the Hamiltonian (3.1), let $\mathcal{D}_0 = \mathcal{D}_{d,\rho_0}(r_0, s_0)$, $\mathcal{O}_0$, $\mathcal{N}_0$(including e_0, ω_0, W_0, p_0, σ_0, N_0), P_0, $\varepsilon_0 = \epsilon^{\frac{1}{4}}$ denote the initial quantities given in the assumptions (**A1**) – (**A5**) respectively, and require that ϵ smaller than the ϵ_0 given in Proposition 1.

Suppose we have arrived at the ν^{th} step of the KAM iteration, $\nu = 0, 1, 2, \cdots$, recalling that several sequences have been given in (2.2). We consider the Hamiltonian on $\mathcal{D}_\nu := \mathcal{D}_{d,\rho_\nu}(r_\nu, s_\nu)$ and $\mathcal{O}_\nu$,

$$\begin{aligned} H_\nu &= \mathcal{N}_\nu + \check{P} + P_\nu \\ &= e_\nu + \langle\omega_\nu, I\rangle + \langle\Omega_\nu q, \bar{q}\rangle + \check{P} + P_\nu, \end{aligned} \tag{4.1}$$

where $\Omega_\nu = T + A + W_\nu$, and (**A1**) – (**A5**) are satisfied, including (3.2), (3.5), (3.6) and

$$(\Omega_\nu)_{mn} \equiv 0, \quad m \text{ or } n \in \mathcal{J}, \tag{4.2}$$

$$|(W_\nu)_{mn}|_{\mathcal{O}_\nu} \le \begin{cases} p_\nu e^{-\sigma_\nu \max\{|m|,|n|\}}, & |m|, |n| \le N_\nu \\ 0, & \text{otherwise} \end{cases}, \tag{4.3}$$

$$\|(P_\nu)_{\alpha\beta}\|_{\mathcal{D}_\nu,\mathcal{O}_\nu} \le \begin{cases} \varepsilon_\nu e^{-\rho_\nu n^*_{\alpha\beta}}, & |\alpha| + |\beta| \le 2 \\ e^{-\rho_\nu n^*_{\alpha\beta}}, & |\alpha| + |\beta| \ge 3 \end{cases}, \tag{4.4}$$

$$\partial_{q_n} P_\nu = \partial_{\bar{q}_n} P_\nu \equiv 0, \quad n \in \mathcal{J}. \tag{4.5}$$

Moreover, $\|X_{\check{P}+P_\nu}\|_{\mathcal{D}_\nu,\mathcal{O}_\nu} \le \varepsilon_\nu$.

Choose some $r_{\nu+1}$ such that $0 < r_{\nu+1} < r_\nu$, and let $J_\nu := \left[\frac{5}{2}\varepsilon_\nu^{-\frac{a_\nu}{2}}\right]$. For $j = 0, 1, \cdots, J_\nu$, we define the quantities at each KAM sub-step as

$$\rho_\nu^{(j)} = (1 - \frac{j}{2J_\nu})\rho_\nu, \quad r_\nu^{(j)} = r_\nu - \frac{j(r_\nu - r_{\nu+1})}{J_\nu}, \quad s_\nu^{(j)} = 2^{-3j}\varepsilon_\nu^{\frac{j}{5}} s_\nu,$$

and $\mathcal{D}_\nu^{(j)} = \mathcal{D}_{d,\rho_{\nu+1}}(r_\nu^{(j)}, s_\nu^{(j)})$, $\varepsilon_\nu^{(j)} = \varepsilon_\nu^{\frac{j}{5}+1}$. Our goal is to construct a set $\mathcal{O}_{\nu+1} \subset \mathcal{O}_\nu$ and a finite sequence of maps

$$\Phi_\nu^{(j)} : \mathcal{D}_\nu^{(j)} \to \mathcal{D}_\nu^{(j-1)}, \quad j = 1, 2, \cdots, J_\nu,$$

so that the Hamiltonian transformed into the $(\nu+1)^{\text{th}}$ KAM cycle

$$\begin{aligned} H_{\nu+1} &= H_\nu \circ \Phi_\nu^{(1)} \circ \cdots \circ \Phi_\nu^{(J_\nu)} \\ &= \mathcal{N}_{\nu+1} + \breve{P} + P_{\nu+1} \\ &= e_{\nu+1} + \langle \omega_{\nu+1}, I \rangle + \langle \Omega_{\nu+1} q, \bar{q} \rangle + \breve{P} + P_{\nu+1} \end{aligned}$$

satisfies all the above iterative assumptions (**A1**) $-$ (**A5**) on $\mathcal{D}_{\nu+1} = \mathcal{D}_\nu^{(J_\nu)}$ and C_W^1 parametrized by $\xi \in \mathcal{O}_{\nu+1}$, with new suitable parameters. Moreover,

$$\|X_{\breve{P}+P_{\nu+1}}\|_{\mathcal{D}_{\nu+1}, \mathcal{O}_{\nu+1}} \le \varepsilon_\nu^{(J_\nu)} \le \varepsilon_\nu^{\frac{1}{2}\varepsilon_\nu^{-a_\nu/2}} = \varepsilon_{\nu+1}.$$

In the remaining part of this paper, all constants labeled with c, c_0, c_1, $\cdots$ are positive and independent of the iteration step.

4.1 Construction of $\mathcal{O}_{\nu+1}$

As described in Proposition 1, there exists an orthogonal matrix U_ν with

$$|(U_\nu - I_{\mathbb{Z}})_{mn}| \le \varepsilon_0^{\frac{1}{2}} e^{-\frac{3}{2}\sigma_\nu |m-n|}, \tag{4.6}$$

such that $U_\nu^* T U_\nu = D_\nu + Z_\nu$, where Z_ν is a symmetric matrix satisfying

$$|(Z_\nu)_{mn}| \le \varepsilon_\nu e^{-\rho_\nu |m-n|}, \tag{4.7}$$

and D_ν is a symmetric matrix which can be block-diagonalized via an orthogonal matrix Q_ν with

$$(Q_\nu)_{mn} = 0 \ \text{ if } \ |m-n| > N_\nu. \tag{4.8}$$

More precisely,

$$\tilde{D}_\nu = Q_\nu^* D_\nu Q_\nu = \prod_j \tilde{D}^\nu_{\Lambda_j^\nu} \ \text{ with } \ \sharp\Lambda_j^\nu \le M_\nu, \ \ \text{diam}\Lambda_j^\nu \le M_\nu N_\nu, \ \ \forall j.$$

To describe $U_\nu^* \Omega_\nu U_\nu$, we need furthermore to consider $U_\nu^* A U_\nu$ and $U_\nu^* W_\nu U_\nu$. In view of (3.2), (4.3) and (4.6), there exists a constant $c_1 > 0$ such that

$$|\,(U_\nu^*(A + W_\nu)U_\nu)_{mn}\,|_{\mathcal{O}_\nu} \le c_1 \max\{\hat{N}^2 e^{3\sigma_\nu \hat{N}},\, p_\nu \sigma_\nu^{-2}\} \cdot e^{-\sigma_\nu \cdot \max\{|m|, |n|\}},$$

by a simple application of Lemma 2.1. Define the truncation $\hat{A}_\nu$ as

$$(\hat{A}_\nu)_{mn} := \begin{cases} (U_\nu^*(A + W_\nu)U_\nu)_{mn}, & |m|, |n| \le N_\nu \\ 0, & \text{otherwise} \end{cases}. \tag{4.9}$$

It follows that

$$\left|\left(U_\nu^*(A + W_\nu)U_\nu - \hat{A}_\nu\right)_{mn}\right|_{\mathcal{O}_\nu} \le \varepsilon_\nu e^{-\rho_\nu \max\{|m|, |n|\}} \tag{4.10}$$

under the assumption

(C1) $c_1 \max\{\hat{N}^2 e^{3\sigma_\nu \hat{N}},\, p_\nu \sigma_\nu^{-2}\} \cdot e^{-(\sigma_\nu - \rho_\nu)N_\nu} \le \varepsilon_\nu$.

Let $K_{\nu+1} := N_{\nu+1} - (M_\nu + 1)N_\nu$ with the sequences M_ν, N_ν, $\nu = 0, 1, \cdots$, defined in (2.2) and

$$\tilde{D}^\nu_{\Lambda^\nu} := \prod_{\Lambda^\nu_j \subset \Lambda^\nu} \tilde{D}^\nu_{\Lambda^\nu_j}, \quad \tilde{A}_\nu := Q^*_\nu \hat{A}_\nu Q_\nu, \tag{4.11}$$

where $\Lambda^\nu := \bigcup\{\Lambda^\nu_j : \Lambda^\nu_j \cap [-(K_{\nu+1} + N_\nu),\ K_{\nu+1} + N_\nu] \ne \emptyset\} \subset [-N_{\nu+1},\ N_{\nu+1}]$. In view of (4.8) and (4.9), we have

$$(\tilde{A}_\nu)_{mn} \equiv 0 \ \text{ if } \ |m| \ \text{ or } \ |n| > 2N_\nu.$$

Since both of $\tilde{D}^\nu_{\Lambda^\nu}$ and $\tilde{A}_\nu$ are Hermitian, there is an orthogonal matrix O_ν such that

$$O^*_\nu (\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu) O_\nu = \mathrm{diag}\{\mu^\nu_j\}_{j \in \Lambda^\nu},$$

where $\{\mu^\nu_j\}_{j\in\Lambda^\nu}$ are eigenvalues of $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$. Due to the block-diagonal structure of $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$, we also have

$$(O_\nu)_{mn} \equiv 0 \ \text{ if } \ |m - n| > 2(M_\nu + 2)N_\nu. \tag{4.12}$$

Indeed, $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$ can be expressed as

$$\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu = (\tilde{D}^\nu_{\Lambda'_\nu} + \tilde{A}_\nu) \cdot \prod_{\Lambda^\nu_j \cap [-2N_\nu, 2N_\nu] = \emptyset} \tilde{D}^\nu_{\Lambda^\nu_j}$$

where $\Lambda'_\nu := \bigcup\{\Lambda^\nu_j : \Lambda^\nu_j \cap [-2N_\nu, 2N_\nu] \ne \emptyset,\ \Lambda^\nu_j \subset \Lambda^\nu\}$ with $\mathrm{diam}\Lambda'_\nu \le 2(M_\nu + 2)N_\nu$. The diagonalization of $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$ is just the diagonalization of blocks $(\tilde{D}^\nu_{\Lambda'_\nu} + \tilde{A}_\nu)$ and $\tilde{D}^\nu_{\Lambda^\nu_j}$.

As for the eigenvalues of $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$, it is well-known that $\{\mu^\nu_n\}_{n\in\Lambda^\nu}$ C^1_W-smoothly depend on ξ and there exist orthonormal eigenvectors ψ^ν_n corresponding to μ^ν_n, C^1_W-smoothly depending on ξ (see e.g. [10]). In fact, $\mu^\nu_n = \langle (\tilde{D}_{\Lambda^\nu} + \tilde{A}_\nu)\psi^\nu_n, \bar{\psi}^\nu_n \rangle$ and

$$\partial_{\xi_j} \mu^\nu_n = \langle (\partial_{\xi_j}(\tilde{D}_{\Lambda^\nu} + \tilde{A}_\nu))\psi^\nu_n, \bar{\psi}^\nu_n \rangle, \quad j = 1, \cdots, b.$$

By the construction of $\tilde{A}_\nu$, we have $\partial_{\xi_j} \tilde{A}_\nu = Q^*_\nu (\partial_{\xi_j} \hat{A}_\nu) Q_\nu$, with $\hat{A}_\nu$ the truncation of $U^*_\nu (A + W_\nu(\xi)) U_\nu$. Since D_ν, A, U_ν and Q_ν are all independent of ξ,

$$\sup_{\xi \in \mathcal{O}_\nu} |\partial_\xi \mu^\nu_n| \le c \sup_{\substack{\xi \in \mathcal{O}_\nu \\ m,n}} |\partial_\xi (W_\nu)_{mn}| \le c p_\nu. \tag{4.13}$$

Now we defined the new parameter set $\mathcal{O}_{\nu+1} \subset \mathcal{O}_\nu$ as

$$\mathcal{O}_{\nu+1} := \left\{ \xi \in \mathcal{O}_\nu : \begin{array}{c} |\langle k, \omega_\nu \rangle| > \frac{\gamma_\nu}{|k|^\tau}, \quad k \ne 0, \\ |\langle k, \omega_\nu \rangle + \mu^\nu_n| > \frac{\gamma_\nu}{|k|^\tau N^2_{\nu+1}}, \quad k \ne 0, \ \ n \in \Lambda^\nu, \\ |\langle k, \omega_\nu \rangle + \mu^\nu_m \pm \mu^\nu_n| > \frac{\gamma_\nu}{|k|^\tau N^4_{\nu+1}}, \quad k \ne 0, \ \ m, n \in \Lambda^\nu. \end{array} \right\} \tag{4.14}$$

for some $0 < \gamma_\nu \ll 1$, $\tau \ge b$. These inequalities are famous small-divisor conditions for controlling the solutions of the linearized equations.

From now on, to simplify notations, the subscripts (or superscripts) "ν" of quantities at the ν^{th} step are neglected, and the corresponding quantities at the $(\nu + 1)^{\text{th}}$ step are labeled with "+". In addition, we still use the superscript (j) to distinguish quantities at various sub-steps.

4.2 Homological equation and its approximate solution

For $P = \sum_{k,l,\alpha,\beta} P_{kl\alpha\beta}(\xi) I^l e^{\mathrm{i}\langle k,\theta\rangle} q^\alpha \bar{q}^\beta$, according to (4.4) and the definition of norm in subsection 3.1, we have

$$|P_{kl\alpha\beta}|_{\mathcal{O}} \le \varepsilon e^{-\rho n^*_{\alpha\beta}} e^{-|k|r}, \quad 2|l|+|\alpha|+|\beta| \le 2, \ \ \forall k \in \mathbb{Z}^b. \tag{4.15}$$

Decompose $P = R + (P - R)$ with

$$R := \sum_{\substack{k \\ 2|l|+|\alpha|+|\beta|\le 2}} P_{kl\alpha\beta} e^{\mathrm{i}\langle k,\theta\rangle} I^l q^\alpha \bar{q}^\beta, \quad P - R = \sum_{\substack{k \\ 2|l|+|\alpha|+|\beta|\ge 3}} P_{kl\alpha\beta} e^{\mathrm{i}\langle k,\theta\rangle} I^l q^\alpha \bar{q}^\beta.$$

It follows $\|X_R\|_{\mathcal{D},\mathcal{O}} \le \|X_P\|_{\mathcal{D},\mathcal{O}} \le \varepsilon$. Recalling that $\breve{P}(q,\bar{q})$ is a sum of high-order terms, there exists a constant $c_2 > 0$ such that

$$\|X_{\breve{P}}\|_{\mathcal{D}_{d,\rho}(r,\eta s),\mathcal{O}}, \ \|X_{P-R}\|_{\mathcal{D}_{d,\rho}(r,\eta s),\mathcal{O}} \le c_2 \eta s \le \frac{1}{8}\varepsilon^{\frac{6}{5}}, \tag{4.16}$$

with $\eta := \varepsilon^{\frac{1}{5}}$, provided

(C2) $c_2 s \le \frac{1}{8}\varepsilon$.

Let $e' := P_{0000}$ and $\omega' := \displaystyle\int \frac{\partial P}{\partial I} d\theta|_{q=\bar{q}=0,I=0}$. With $\mathcal{O}_+$ defined as in (4.14), we have

Proposition 2 *There exist two real-analytic Hamiltonians*

$$F = \sum_{\substack{k\neq 0 \\ 1\le 2|l|+|\alpha|+|\beta|\le 2}} F_{kl\alpha\beta} q^\alpha \bar{q}^\beta I^l e^{\mathrm{i}\langle k,\theta\rangle}, \quad \grave{P} = \sum_{\substack{k \\ 1\le |\alpha|+|\beta|\le 2}} \grave{P}_{k0\alpha\beta} q^\alpha \bar{q}^\beta e^{\mathrm{i}\langle k,\theta\rangle},$$

and a Hermitian matrix W', all of which are C^1_W parametrized by $\xi \in \mathcal{O}_+$, such that

$$\{\mathcal{N}, F\} + R = e' + \langle \omega', I\rangle + \langle W' q, \bar{q}\rangle + \grave{P}. \tag{4.17}$$

Moreover, both of F and $\grave{P}$ have gauge invariance, and for ε sufficiently small,

$$|F_{kl\alpha\beta}|_{\mathcal{O}_+} \le \varepsilon^{\frac{4}{5}} |k|^{2\tau+1} e^{-|k|r} e^{-\rho n^*_{\alpha\beta}}, \tag{4.18}$$

$$|\grave{P}_{k0\alpha\beta}|_{\mathcal{O}_+} \le \varepsilon^{\frac{7}{5}} |k|^{2\tau+1} e^{-|k|r} e^{-\rho^{(1)} n^*_{\alpha\beta}}, \tag{4.19}$$

$$|W'_{mn}|_{\mathcal{O}_+} \le \begin{cases} \varepsilon e^{-\rho \max\{|m|,|n|\}}, & |m|,|n| \le N_+, \quad m,n \notin \mathcal{J} \\ 0, & otherwise \end{cases}, \tag{4.20}$$

$$\partial_{q_n} F = \partial_{\bar{q}_n} F = \partial_{q_n} \grave{P} = \partial_{\bar{q}_n} \grave{P} \equiv 0, \quad n \in \mathcal{J}. \tag{4.21}$$

Proof of Proposition 2: We decompose the proof into the following parts.

- **Truncation and approximate linearized equations**

At first, we rewrite R as

$$R = \sum_{\substack{k \\ |l|\le 1}} P_{kl00} e^{\mathrm{i}\langle k,\theta\rangle} I^l + \sum_k (\langle P^{k10}, q\rangle + \langle P^{k01}, \bar{q}\rangle + \langle P^{k20} q, q\rangle + \langle P^{k11} q, \bar{q}\rangle + \langle P^{k02} \bar{q}, \bar{q}\rangle) e^{\mathrm{i}\langle k,\theta\rangle},$$

where P^{k10}, P^{k01}, P^{k20}, P^{k11}, P^{k02} respectively denote

$$\begin{aligned}
&\left(P^{k10}_n\right) := \left(P_{k0e_n0}\right), \qquad \left(P^{k01}_n\right) := \left(P_{k00e_n}\right), \\
&\left(P^{k20}_{mn}\right) := \left(P_{k0(e_m+e_n)0}\right), \quad \left(P^{k11}_{mn}\right) := \left(P_{k0e_me_n}\right), \quad \left(P^{k02}_{mn}\right) := \left(P_{k00(e_m+e_n)}\right).
\end{aligned}$$

The gauge invariance **(A5)** implies that P^{010}, P^{001}, P^{020}, $P^{002} \equiv 0$.

We try to construct a Hamiltonian F, of the same form as R, such that

$$\{\mathcal{N}, F\} + R = e' + \langle \omega', I\rangle + \langle P^{011} q, \bar{q}\rangle. \tag{4.22}$$

By a straightforward calculation and simple comparison of coefficients, Equation (4.22) is equivalent to the following equations for $k \neq 0$ and $|l| \leq 1$,

$$\langle k, \omega\rangle F_{kl00} = \mathrm{i} P_{kl00}, \tag{4.23}$$

$$(\langle k, \omega\rangle I_{\mathbb{Z}} - \Omega) F^{k10} = \mathrm{i} P^{k10}, \tag{4.24}$$

$$(\langle k, \omega\rangle I_{\mathbb{Z}} + \Omega) F^{k01} = \mathrm{i} P^{k01}, \tag{4.25}$$

$$(\langle k, \omega\rangle I_{\mathbb{Z}} - \Omega) F^{k20} - F^{k20}\Omega = \mathrm{i} P^{k20}, \tag{4.26}$$

$$(\langle k, \omega\rangle I_{\mathbb{Z}} - \Omega) F^{k11} + F^{k11}\Omega = \mathrm{i} P^{k11}, \tag{4.27}$$

$$(\langle k, \omega\rangle I_{\mathbb{Z}} + \Omega) F^{k02} + F^{k02}\Omega = \mathrm{i} P^{k02}. \tag{4.28}$$

In view of the definition of $\mathcal{O}_+$, we know that (4.23) is solved on $\mathcal{O}_+$, with

$$|F_{kl00}|_{\mathcal{O}_+} \leq \gamma^{-2} |k|^{2\tau+1} \varepsilon e^{-|k|r}.$$

As for (4.24) – (4.28), consider the equations

$$\left(\langle k, \omega\rangle I_{\mathbb{Z}} - (D + \hat{A})\right) \hat{F}^{k10} = \mathrm{i} \hat{R}^{k10}, \tag{4.29}$$

$$\left(\langle k, \omega\rangle I_{\mathbb{Z}} + (D + \hat{A})\right) \hat{F}^{k01} = \mathrm{i} \hat{R}^{k01}, \tag{4.30}$$

$$\left(\langle k, \omega\rangle I_{\mathbb{Z}} - (D + \hat{A})\right) \hat{F}^{k20} - \hat{F}^{k20}(D + \hat{A}) = \mathrm{i} \hat{R}^{k20}, \tag{4.31}$$

$$\left(\langle k, \omega\rangle I_{\mathbb{Z}} - (D + \hat{A})\right) \hat{F}^{k11} + \hat{F}^{k11}(D + \hat{A}) = \mathrm{i} \hat{R}^{k11}, \tag{4.32}$$

$$\left(\langle k, \omega\rangle I_{\mathbb{Z}} + (D + \hat{A})\right) \hat{F}^{k02} + \hat{F}^{k02}(D + \hat{A}) = \mathrm{i} \hat{R}^{k02} \tag{4.33}$$

instead, where D and $\hat{A}$ are defined in the previous subsection, and for $k \neq 0$,

$$\hat{R}^{kx}_n = \begin{cases} (U^* P^{kx})_n, & |n| \leq K_+ \\ 0, & \text{otherwise} \end{cases}, \quad x = \text{“10”}, \text{“01”}, \tag{4.34}$$

$$\hat{R}^{kx}_{mn} = \begin{cases} (U^* P^{kx} U)_{mn}, & |m|, |n| \leq K_+ \\ 0, & \text{otherwise} \end{cases}, \quad x = \text{“20”}, \text{“11”}, \text{“02”}. \tag{4.35}$$

By (4.6) and (4.15), combining with Lemma 2.1, there exists $c_3 > 0$ such that

$$|(U^* P^{kx})_n|_{\mathcal{O}} \leq c_3 (\sigma - \rho)^{-1} \varepsilon e^{-\rho|n|} e^{-|k|r}, \tag{4.36}$$

$$|(U^* P^{kx} U)_{mn}|_{\mathcal{O}} \leq c_3 (\sigma - \rho)^{-2} \varepsilon e^{-\rho \max\{|m|, |n|\}} e^{-|k|r}. \tag{4.37}$$

This means

$$|(U^*P^{kx} - \hat{R}^{kx})_n|_{\mathcal{O}} \le \frac{1}{4}\varepsilon^{\frac{7}{5}} e^{-\rho^{(1)}|n|} e^{-|k|r}, \tag{4.38}$$

$$|(U^*P^{kx}U - \hat{R}^{kx})_{mn}|_{\mathcal{O}} \le \frac{1}{4}\varepsilon^{\frac{7}{5}} e^{-\rho^{(1)}\max\{|m|,|n|\}} e^{-|k|r} \tag{4.39}$$

under the assumption that

(C3) $c_3(\sigma-\rho)^{-4}e^{-(\rho-\rho^{(1)})K_+} \le \frac{1}{4}\varepsilon^{\frac{2}{5}}$.

Equation (4.29) – (4.33) provide us with approximate solutions to (4.24) – (4.28), with the error estimated later.

- **Block-diagonalization and construction of** F

Consider the equations

$$\left(\langle k,\omega\rangle I_\Lambda - (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{k10} = \mathrm{i}\tilde{R}^{k10}, \tag{4.40}$$

$$\left(\langle k,\omega\rangle I_\Lambda + (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{k01} = \mathrm{i}\tilde{R}^{k01}, \tag{4.41}$$

$$\left(\langle k,\omega\rangle I_\Lambda - (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{k20} - \tilde{F}^{k20}(\tilde{D}_\Lambda + \tilde{A}) = \mathrm{i}\tilde{R}^{k20}, \tag{4.42}$$

$$\left(\langle k,\omega\rangle I_\Lambda - (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{k11} + \tilde{F}^{k11}(\tilde{D}_\Lambda + \tilde{A}) = \mathrm{i}\tilde{R}^{k11}, \tag{4.43}$$

$$\left(\langle k,\omega\rangle I_\Lambda + (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{k02} + \tilde{F}^{k02}(\tilde{D}_\Lambda + \tilde{A}) = \mathrm{i}\tilde{R}^{k02}, \tag{4.44}$$

where $\tilde{D}_\Lambda$, $\tilde{A}$ are defined as in (4.11) via the orthogonal matrix Q, and

$$\tilde{R}^{kx} := \begin{cases} Q^*\hat{R}^{kx}, & x = \text{“10”, “01”} \\ Q^*\hat{R}^{kx}Q, & x = \text{“20”, “11”, “02”} \end{cases}.$$

Note that $Q_{mn} = 0$ if $|m-n| > N$, then by (4.34) and (4.35), we have

$$\tilde{R}^{kx}_n \equiv 0, \quad \text{if} \quad |n| > K_+ + N, \quad x = \text{“10”, “01”},$$
$$\tilde{R}^{kx}_{mn} \equiv 0, \quad \text{if} \quad |m| \ \text{or} \ |n| > K_+ + N, \quad x = \text{“20”, “11”, “02”}.$$

Thus, recalling that $\Lambda := \bigcup\{\Lambda_j : \Lambda_j \cap [-(K_+ + N), \ K_+ + N] \neq \emptyset\}$, solutions of these finite-dimensional equations satisfy

$$\tilde{F}^{kx}_n \equiv 0, \quad \text{if} \quad n \notin \Lambda, \quad x = \text{“10”, “01”},$$
$$\tilde{F}^{kx}_{mn} \equiv 0, \quad \text{if} \quad m \ \text{or} \ n \notin \Lambda, \quad x = \text{“20”, “11”, “02”}.$$

Then, in view of the facts

$$\left(\langle k,\omega\rangle I_{\mathbb{Z}} \pm (\tilde{D} + \tilde{A})\right)\tilde{F}^{kx} = \left(\langle k,\omega\rangle I_\Lambda \pm (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{kx}, \quad x = \text{“10”, “01”},$$
$$\left(\langle k,\omega\rangle I_{\mathbb{Z}} \pm (\tilde{D} + \tilde{A})\right)\tilde{F}^{kx} = \left(\langle k,\omega\rangle I_\Lambda \pm (\tilde{D}_\Lambda + \tilde{A})\right)\tilde{F}^{kx}, \quad x = \text{“20”, “11”, “02”},$$
$$\tilde{F}^{kx}(\tilde{D} + \tilde{A}) = \tilde{F}^{kx}(\tilde{D}_\Lambda + \tilde{A}), \quad x = \text{“20”, “11”, “02”},$$

they are also solutions of

$$\begin{aligned}
\Big(\langle k,\omega\rangle I_{\mathbb{Z}} - (\tilde{D}+\tilde{A})\Big)\,\tilde{F}^{k10} &= \mathrm{i}\tilde{R}^{k10},\\
\Big(\langle k,\omega\rangle I_{\mathbb{Z}} + (\tilde{D}+\tilde{A})\Big)\,\tilde{F}^{k01} &= \mathrm{i}\tilde{R}^{k01},\\
\Big(\langle k,\omega\rangle I_{\mathbb{Z}} - (\tilde{D}+\tilde{A})\Big)\,\tilde{F}^{k20} - \tilde{F}^{k20}(\tilde{D}+\tilde{A}) &= \mathrm{i}\tilde{R}^{k20},\\
\Big(\langle k,\omega\rangle I_{\mathbb{Z}} - (\tilde{D}+\tilde{A})\Big)\,\tilde{F}^{k11} + \tilde{F}^{k11}(\tilde{D}+\tilde{A}) &= \mathrm{i}\tilde{R}^{k11},\\
\Big(\langle k,\omega\rangle I_{\mathbb{Z}} + (\tilde{D}+\tilde{A})\Big)\,\tilde{F}^{k02} + \tilde{F}^{k02}(\tilde{D}+\tilde{A}) &= \mathrm{i}\tilde{R}^{k02},
\end{aligned}$$

which are respectively equivalent to Equation (4.29) – (4.33) since D can be block-diagonalized by the orthogonal matrix Q.

Now we focus on the following equations

$$\begin{aligned}
(\langle k,\omega\rangle - \mu_n)\check{F}^{k10}_n &= \mathrm{i}(O^*\tilde{R}^{k10})_n,\\
(\langle k,\omega\rangle + \mu_n)\check{F}^{k01}_n &= \mathrm{i}(O^*\tilde{R}^{k01})_n,\\
(\langle k,\omega\rangle - \mu_m - \mu_n)\check{F}^{k20}_{mn} &= \mathrm{i}(O^*\tilde{R}^{k20}O)_{mn},\\
(\langle k,\omega\rangle - \mu_m + \mu_n)\check{F}^{k11}_{mn} &= \mathrm{i}(O^*\tilde{R}^{k11}O)_{mn},\\
(\langle k,\omega\rangle + \mu_m + \mu_n)\check{F}^{k02}_{mn} &= \mathrm{i}(O^*\tilde{R}^{k02}O)_{mn},
\end{aligned}$$

for $k \neq 0$ and $m,n \in \Lambda$, which is transformed from (4.40) – (4.44) by diagonalizing $\tilde{D}_\Lambda + \tilde{A}$ via the orthogonal matrix O. Obviously, these equations can be solved in $\mathcal{O}_+$. Hence, (4.29) – (4.33) are solved with

$$\hat{F}^{kx} = \begin{cases} QO\check{F}^{kx}, & x = \text{“10”, “01”}\\ QO\check{F}^{kx}O^*Q^*, & x = \text{“20”, “11”, “02”} \end{cases}.$$

Let

$$F^{kx} := \begin{cases} U\hat{F}^{kx}, & x = \text{“10”, “01”}\\ U\hat{F}^{kx}U^*, & x = \text{“20”, “11”, “02”} \end{cases},$$

then we obtain a Hamiltonian

$$F = \sum_{\substack{k\neq 0\\ |l|\leq 1}} F_{kl00} e^{\mathrm{i}\langle k,\theta\rangle} I^l + \sum_{k\neq 0} (\langle F^{k10}, q\rangle + \langle F^{k01}, \bar{q}\rangle + \langle F^{k20}q, q\rangle + \langle F^{k11}q, \bar{q}\rangle + \langle F^{k02}\bar{q}, \bar{q}\rangle) e^{\mathrm{i}\langle k,\theta\rangle}.$$

It is easy to see that $\bar{F} = F$, by noting

$$\begin{aligned}
&\overline{F_{(-k)l00}} = F_{kl00}, \quad \overline{F^{(-k)10}} = F^{k01}, \quad \overline{F^{(-k)01}} = F^{k10},\\
&\overline{F^{(-k)20}} = F^{k02}, \quad (F^{(-k)11})^* = F^{k11}, \quad \overline{F^{(-k)02}} = F^{k20}.
\end{aligned}$$

- **Estimates for coefficients of** F

Let us consider F^{k20}_{mn} for instance, and the other terms can be treated in an analogous way. By the construction above, one sees that

$$F^{k20}_{mn} = \mathrm{i}\sum_{\mathcal{F}_0} \frac{U_{mn_1}Q_{n_1n_2}O_{n_2n_3}O^*_{n_3n_4}Q^*_{n_4n_5}\hat{R}^{k20}_{n_5n_6}Q_{n_6n_7}O_{n_7n_8}O^*_{n_8n_9}Q^*_{n_9n_{10}}U^*_{n_{10}n}}{\langle k,\omega\rangle - \mu_{n_3} - \mu_{n_8}}, \tag{4.45}$$

where the summation notation $\mathcal{F}_0$ denotes

$$\left\{\begin{array}{l} n_1\in\mathbb{Z},\quad |n_2-n_1|\le N,\quad |n_3-n_2|,\,|n_4-n_3|\le 2(M+2)N,\quad |n_5-n_4|\le N,\\ n_{10}\in\mathbb{Z},\quad |n_9-n_{10}|\le N,\quad |n_8-n_9|,\,|n_7-n_8|\le 2(M+2)N,\quad |n_6-n_7|\le N \end{array}\right\}$$

by virtue of the structure of Q and O, i.e, (4.8) and (4.12). Then, by (4.37) and Lemma 2.1,

$$\sup_{\xi\in\mathcal{O}_+}|F^{k20}_{mn}(\xi)|\le c(\gamma^{-1}|k|^{\tau}N_+^4)(\sigma-\rho)^{-4}M^4N^8e^{(4M+10)N\rho}\varepsilon e^{-|k|r}e^{-\rho\max\{|m|,|n|\}}.$$

Here we have applied the property of the orthogonal matrices Q and O, and used the factor $e^{(4M+10)N\rho}$ to recover the exponential decay.

To estimate $|\partial_{\xi_j}F^{k20}_{mn}|$, we need to differentiate both sides of (4.42) with respect to ξ_j, $j=1,2,\cdots,b$. Then we obtain the equation about $\partial_{\xi_j}\tilde{F}^{k20}$

$$(\langle k,\omega\rangle I_\Lambda-(\tilde{D}_\Lambda+\tilde{A}))(\partial_{\xi_j}\tilde{F}^{k20})-(\partial_{\xi_j}\tilde{F}^{k20})(\tilde{D}_\Lambda+\tilde{A})=\breve{R}^{k20}_{\xi_j},$$

which can also be solved by diagonalizing $\tilde{D}_\Lambda+\tilde{A}$ via O as above, where

$$\breve{R}^{k20}_{\xi_j}:=\mathrm{i}\partial_{\xi_j}\tilde{R}^{k20}+\tilde{F}^{k20}(\partial_{\xi_j}\tilde{A})-(\partial_{\xi_j}(\langle k,\omega\rangle I-\tilde{A}))\tilde{F}^{k20}.$$

We get the formulation

$$\partial_{\xi_j}F^{k20}_{mn}=\sum_{\mathcal{F}_1}\frac{U_{mn_1}Q_{n_1n_2}O_{n_2n_3}O^*_{n_3n_4}(\breve{R}^{k20}_{\xi_j})_{n_4n_5}O_{n_5n_6}O^*_{n_6n_7}Q^*_{n_7n_8}U^*_{n_8n}}{\langle k,\omega\rangle-\mu_{n_3}-\mu_{n_6}},$$

with $\mathcal{F}_1$ denotes

$$\left\{\begin{array}{l} n_1\in\mathbb{Z},\quad |n_2-n_1|\le N,\quad |n_3-n_2|,\,|n_4-n_3|\le 2(M+2)N,\\ n_8\in\mathbb{Z},\quad |n_7-n_8|\le N,\quad |n_6-n_7|,\,|n_5-n_6|\le 2(M+2)N \end{array}\right\}.$$

By the decay property of $\hat{R}^{k20}$ and $\partial_{\xi_j}\hat{A}$, we have that

$$\sup_{\xi\in\mathcal{O}_+}|(\breve{R}^{k20}_{\xi_j})_{mn}|\le c(\gamma^{-1}|k|^{\tau+1}N_+^4)(\sigma-\rho)^{-4}M^4N^8e^{(4M+11)N\rho}\varepsilon e^{-|k|r}e^{-\rho\max\{|m|,|n|\}}.$$

Thus there exists $c_4>0$ such that

$$\begin{aligned}&\sup_{\xi\in\mathcal{O}_+}(|F^{k20}_{mn}|+|\partial_\xi F^{k20}_{mn}|)\\ \le\;& c_4(\gamma^{-2}|k|^{2\tau+1}N_+^8)(\sigma-\rho)^{-6}M^8N^{14}e^{(8M+20)N\rho}\varepsilon e^{-\rho\max\{|m|,|n|\}}e^{-|k|r}\\ \le\;& \varepsilon^{\frac{4}{5}}|k|^{2\tau+1}e^{-|k|r}e^{-\rho\max\{|m|,|n|\}},\end{aligned}$$

under the assumption

(C4) $c_4\gamma^{-2}(\sigma-\rho)^{-6}N_+^8M^8N^{14}e^{(8M+20)N\rho}\varepsilon^{\frac{1}{5}}\le 1.$

Suppose that $\sum_{i=1}^b k_i+2\ne 0$, which means $P^{k20}\equiv 0$. Then $\hat{R}^{k20}\equiv 0$, since it is a truncation of $U^*P^{k20}U$. By the formulation of F^{k20}_{mn} in (4.45), $F^{k20}\equiv 0$.

Doing the same thing for F^{k11}, F^{k02}, F^{k10}, F^{k01} as above, we obtain the gauge invariance of F and the inequality (4.18).

- **Estimates for coefficients of $\grave{P}$**

Let W' be the truncation of P^{011}, satisfying

$$W'_{mn} = \begin{cases} P^{011}_{mn}, & |m|, |n| \leq N_+ \\ 0, & \text{otherwise} \end{cases},$$

and

$$\grave{P} = \langle \grave{P}^{011} q, \bar{q} \rangle + \sum_{k \neq 0} (\langle \grave{P}^{k10}, q \rangle + \langle \grave{P}^{k01}, \bar{q} \rangle + \langle \grave{P}^{k20} q, q \rangle + \langle \grave{P}^{k11} q, \bar{q} \rangle + \langle \grave{P}^{k02} \bar{q}, \bar{q} \rangle) e^{\mathrm{i}\langle k, \theta \rangle}$$

with

$$\begin{aligned}
\grave{P}^{011} &:= P^{011} - W', \\
\grave{P}^{k10} &:= (P^{k10} - U\hat{R}^{k10}) - \mathrm{i}(\grave{A} + \grave{Z}) F^{k10}, \\
\grave{P}^{k01} &:= (P^{k01} - U\hat{R}^{k01}) + \mathrm{i}(\grave{A} + \grave{Z}) F^{k01}, \\
\grave{P}^{k20} &:= (P^{k20} - U\hat{R}^{k20}U^*) - \mathrm{i}(\grave{A} + \grave{Z}) F^{k20} - \mathrm{i}F^{k20}(\grave{A} + \grave{Z}), \\
\grave{P}^{k11} &:= (P^{k11} - U\hat{R}^{k11}U^*) - \mathrm{i}(\grave{A} + \grave{Z}) F^{k11} + \mathrm{i}F^{k11}(\grave{A} + \grave{Z}), \\
\grave{P}^{k02} &:= (P^{k02} - U\hat{R}^{k02}U^*) + \mathrm{i}(\grave{A} + \grave{Z}) F^{k02} + \mathrm{i}F^{k02}(\grave{A} + \grave{Z}),
\end{aligned}$$

where $\grave{A} := (A + W) - U\hat{A}U^*$, and $\grave{Z} := UZU^*$. Then we obtain

$$\{\mathcal{N}, F\} + R = e' + \langle \omega', I \rangle + \langle W' q, \bar{q} \rangle + \grave{P}. \tag{4.46}$$

By (4.4) and (4.5), we have (4.20) holds and

$$|\grave{P}^{011}_{mn}|_{\mathcal{O}_+} \leq \varepsilon e^{-\rho \max\{|m|, |n|\}} \leq \varepsilon^{\frac{7}{5}} e^{-\rho^{(1)} \max\{|m|, |n|\}},$$

under the assumption

(C5) $e^{-(\rho - \rho^{(1)}) N_+} \leq \varepsilon^{\frac{2}{5}}$.

As for the case $k \neq 0$ in (4.19), we only estimate $\grave{P}^{k20}$, with the others entirely analogous. By (4.39) and **(C3)**, combining with Lemma 2.1,

$$\left| \left(P^{k20} - U\hat{R}^{k20}U^* \right)_{mn} \right|_{\mathcal{O}} = \left| \left(U(U^* P^{k20} U - \hat{R}^{k20}) U^* \right)_{mn} \right|_{\mathcal{O}} \leq \frac{1}{4} \varepsilon^{\frac{7}{5}} e^{-\rho^{(1)} \max\{|m|, |n|\}} e^{-|k| r}. \tag{4.47}$$

In view of (4.7) and (4.10),

$$|\grave{A}_{mn}|_{\mathcal{O}} \leq c(\sigma - \rho)^{-2} \varepsilon e^{-\rho \max\{|m|, |n|\}}, \quad |\grave{Z}_{mn}| \leq c(\sigma - \rho)^{-2} \varepsilon e^{-\rho |m - n|}.$$

Then, by applying Lemma 2.1 again, there exists $c_6 > 0$ such that

$$\begin{aligned}
& \left| \left(F^{k20} (\grave{A} + \grave{Z}) \right)_{mn} \right|_{\mathcal{O}_+}, \ \left| \left((\grave{A} + \grave{Z}) F^{k20} \right)_{mn} \right|_{\mathcal{O}_+} \\
\leq\ & c_6 (\sigma - \rho)^{-2} (\rho - \rho^{(1)})^{-1} \varepsilon^{\frac{9}{5}} |k|^{2\tau + 1} e^{-|k| r} e^{-\rho^{(1)} \max\{|m|, |n|\}} \\
\leq\ & \frac{1}{4} \varepsilon^{\frac{7}{5}} |k|^{2\tau + 1} e^{-|k| r} e^{-\rho^{(1)} \max\{|m|, |n|\}},
\end{aligned} \tag{4.48}$$

provided that

(C6) $c_6(\sigma-\rho)^{-2}(\rho-\rho^{(1)})^{-1}\varepsilon^{\frac{2}{5}} \le \frac{1}{4}$.

Thus, we can obtain the estimate for $\grave{P}^{k20}$ by putting (4.47) and (4.48) together.

By the construction of $\grave{P}$, the gauge invariance is easily verified.

- **Verification of (4.21)**

In view of the construction of R and W' above, the objects in (4.46) that may depend on the variables $(q_n,\bar{q}_n)_{n\in\mathcal{J}}$ are F and $\grave{P}$. Let

$$\begin{aligned}\acute{F} &= \sum_{k\neq 0}\left(\sum_{n\in\mathcal{J}}(F_n^{k10}q_n+F_n^{k01}\bar{q}_n)+\sum_{m\,\mathrm{or}\,n\in\mathcal{J}}(F_{mn}^{k20}q_mq_n+F_{mn}^{k11}q_m\bar{q}_n+F_{mn}^{k02}\bar{q}_m\bar{q}_n)\right)e^{\mathrm{i}\langle k,\theta\rangle}\\ &=: \sum_{k\neq 0}\left(\langle \acute{F}^{k10},q\rangle+\langle \acute{F}^{k01},\bar{q}\rangle+\langle \acute{F}^{k20}q,q\rangle+\langle \acute{F}^{k11}q,\bar{q}\rangle+\langle \acute{F}^{k02}\bar{q},\bar{q}\rangle\right)e^{\mathrm{i}\langle k,\theta\rangle}.\end{aligned}$$

For m or $n\in\mathcal{J}$, by (4.2), we have

$$\begin{aligned}\left((\langle k,\omega\rangle I_{\mathbb{Z}}-\Omega)\acute{F}^{k20}-\acute{F}^{k20}\Omega\right)_{mn} &= \langle k,\omega\rangle\acute{F}_{mn}^{k20}-\sum_{l\notin\mathcal{J}}\Omega_{ml}\acute{F}_{ln}^{k20}-\sum_{l\notin\mathcal{J}}\acute{F}_{ml}^{k20}\Omega_{ln}\\ &= \begin{cases}\langle k,\omega\rangle F_{mn}^{k20}, & m,n\in\mathcal{J}\\ \langle k,\omega\rangle F_{mn}^{k20}-\sum_{l\notin\mathcal{J}}\Omega_{ml}F_{ln}^{k20}, & m\notin\mathcal{J},\ \ n\in\mathcal{J}\\ \langle k,\omega\rangle F_{mn}^{k20}-\sum_{l\notin\mathcal{J}}F_{ml}^{k20}\Omega_{ln}, & m\in\mathcal{J},\ \ n\notin\mathcal{J}\end{cases}\\ &= \left((\langle k,\omega\rangle I_{\mathbb{Z}}-\Omega)F^{k20}-F^{k20}\Omega\right)_{mn}.\end{aligned}$$

This means, by comparing the coefficients in both side of Equation (4.46),

$$\left((\langle k,\omega\rangle I_{\mathbb{Z}}-\Omega)\acute{F}^{k20}-\acute{F}^{k20}\Omega\right)_{mn}=-\mathrm{i}\grave{P}_{mn}^{k20}\quad m\ \text{ or }\ n\in\mathcal{J}.$$

Similarly,

$$\begin{aligned}\left((\langle k,\omega\rangle I_{\mathbb{Z}}-\Omega)\acute{F}^{k10}\right)_n &= -\mathrm{i}\grave{P}_n^{k10}, \quad n\in\mathcal{J},\\ \left((\langle k,\omega\rangle I_{\mathbb{Z}}+\Omega)\acute{F}^{k01}\right)_n &= -\mathrm{i}\grave{P}_n^{k01}, \quad n\in\mathcal{J},\\ \left((\langle k,\omega\rangle I_{\mathbb{Z}}-\Omega)\acute{F}^{k11}+\acute{F}^{k11}\Omega\right)_{mn} &= -\mathrm{i}\grave{P}_{mn}^{k11}, \quad m\ \text{ or }\ n\in\mathcal{J},\\ \left((\langle k,\omega\rangle I_{\mathbb{Z}}+\Omega)\acute{F}^{k02}+\acute{F}^{k02}\Omega\right)_{mn} &= -\mathrm{i}\grave{P}_{mn}^{k02}, \quad m\ \text{ or }\ n\in\mathcal{J}.\end{aligned}$$

Thus, $\{\mathcal{N},\acute{F}\}$ equals to

$$\sum_{k\neq 0}\left(\sum_{n\in\mathcal{J}}(\grave{P}_n^{k10}q_n+\grave{P}_n^{k01}\bar{q}_n)+\sum_{m\,\mathrm{or}\,n\in\mathcal{J}}(\grave{P}_{mn}^{k20}q_mq_n+\grave{P}_{mn}^{k11}q_m\bar{q}_n+\grave{P}_{mn}^{k02}\bar{q}_m\bar{q}_n)\right)e^{\mathrm{i}\langle k,\theta\rangle}.$$

Hence, if we substitute F with $F-\acute{F}$, which is independent of the variables $(q_n,\bar{q}_n)_{n\in\mathcal{J}}$, then $\grave{P}$ in the homological equation (4.46) is replaced correspondingly, independent of $(q_n,\bar{q}_n)_{n\in\mathcal{J}}$. (4.21) is satisfied. ∎

4.3 Verification of assumptions after one sub-step

We proceed to estimate the norm of X_F, and to study properties of Φ_F^1 on smaller domains $\mathcal{D}_i := \mathcal{D}_{d,\rho_+}(r^{(1)} + \frac{i}{4}(r - r^{(1)}),\ \frac{i}{4}s)$, $i = 1, 2, 3, 4$.

Lemma 4.1 *For ε sufficiently small, we have $\|X_F\|_{\mathcal{D}_3,\mathcal{O}_+} \le \varepsilon^{\frac{3}{4}}$ and $\|X_{\grave{P}}\|_{\mathcal{D}_3,\mathcal{O}_+} \le \varepsilon^{\frac{5}{4}}$.*

Proof: In view of the decay property of F in Proposition 2, it follows that

$$\frac{1}{s^2}\|\partial_\theta F\|_{\mathcal{D}_3,\mathcal{O}_+},\ \|\partial_I F\|_{\mathcal{D}_3,\mathcal{O}_+} \le c(r - r^{(1)})^{-(2\tau+b+1)}\varepsilon^{\frac{4}{5}},$$

and

$$\begin{aligned} & \sup_{\mathcal{D}_3}\frac{1}{s}\sum_{n\in\mathbb{Z}}\left(\|\partial_{q_n}F\|_{\mathcal{O}_+} + \|\partial_{\bar q_n}F\|_{\mathcal{O}_+}\right)\langle n\rangle^d e^{\rho_+|n|} \\ \le\ & \sup_{\mathcal{D}_3}\frac{c}{s}\sum_{n\in\mathbb{Z}}\sum_{k\ne 0}\left(|F_n^{k10}|_{\mathcal{O}_+} + |F_n^{k01}|_{\mathcal{O}_+}\right)e^{|k|(r-\frac{1}{4}(r-r^{(1)}))}\langle n\rangle^d e^{\rho_+|n|} \\ & + \sup_{\mathcal{D}_3}\frac{c}{s}\sum_{n\in\mathbb{Z}}\sum_{\substack{k\ne 0\\ m\in\mathbb{Z}}}(|F_{mn}^{k20}|_{\mathcal{O}_+} + |F_{mn}^{k11}|_{\mathcal{O}_+} + |F_{mn}^{k02}|_{\mathcal{O}_+})|q_m|e^{|k|(r-\frac{1}{4}(r-r^{(1)}))}\langle n\rangle^d e^{\rho_+|n|} \\ \le\ & c(r - r^{(1)})^{-(2\tau+b+1)}(\rho-\rho_+)^{-2}\varepsilon^{\frac{4}{5}}. \end{aligned}$$

Putting together the estimates above, there is a constant $c_7 > 0$ such that

$$\|X_F\|_{\mathcal{D}_3,\mathcal{O}_+} \le c_7(r - r^{(1)})^{-(2\tau+b+1)}(\rho-\rho_+)^{-2}\varepsilon^{\frac{4}{5}}.$$

In an entirely analogous way, we have

$$\|X_{\grave{P}}\|_{\mathcal{D}_3,\mathcal{O}_+} \le c_7(r - r^{(1)})^{-(2\tau+b+1)}(\rho^{(1)}-\rho_+)^{-2}\varepsilon^{\frac{7}{5}}.$$

Moreover, if

(C7) $c_7(r - r^{(1)})^{-(2\tau+b+1)}(\rho^{(1)}-\rho_+)^{-2}\varepsilon^{\frac{1}{20}} \le \frac{1}{3}$,

then Lemma 4.1 follows. ■

Let $\mathcal{D}_{i\eta} = \mathcal{D}_{d,\rho_+}(r^{(1)} + \frac{i}{4}(r - r^{(1)}), \frac{i}{4}\eta s)$, $i = 1, 2, 3, 4$.

Lemma 4.2 *For ε sufficiently small, we have $\Phi_F^t : \mathcal{D}_{2\eta} \to \mathcal{D}_{3\eta}$, $-1 \le t \le 1$ and moreover,*

$$\|D\Phi_F^t - I\|_{\mathcal{D}_{1\eta}} < 2\varepsilon^{\frac{3}{4}}.$$

Let $F^{(1)}$, $e^{(1)}$, $\omega^{(1)}$, $W^{(1)}$, $\grave{P}^{(1)}$ be the corresponding quantities in (4.17) respectively, which means that we are in the 1st sub-step. Define $H^{(1)}$ as

$$\begin{aligned} H^{(1)} &:= H \circ \Phi_{F^{(1)}}^1 \\ &= (\mathcal{N} + \breve{P} + R)\circ\Phi_{F^{(1)}}^1 + (P - R)\circ\Phi_{F^{(1)}}^1 \\ &= \mathcal{N} + \breve{P} + \{\mathcal{N}, F^{(1)}\} + R + \int_0^1 (1-t)\{\{\mathcal{N}, F^{(1)}\}, F^{(1)}\}\circ\Phi_{F^{(1)}}^t\,dt \\ &\quad + \int_0^1 \{\breve{P} + R, F^{(1)}\}\circ\Phi_{F^{(1)}}^t\,dt + (P - R)\circ\Phi_{F^{(1)}}^1 \\ &= \mathcal{N} + \breve{P} + e^{(1)} + \langle\omega^{(1)}, I\rangle + \langle W^{(1)}q, \bar q\rangle + P^{(1)}, \end{aligned}$$

where

$$P^{(1)} := \grave{P}^{(1)} + \int_0^1 \{(1-t)\{\mathcal{N}, F^{(1)}\} + \breve{P} + R, F^{(1)}\} \circ \Phi^t_{F^{(1)}}\, dt + (P-R) \circ \Phi^1_{F^{(1)}}.$$

Let $R(t) := (1-t)(e^{(1)} + \langle \omega^{(1)}, I\rangle + \langle W^{(1)}q, \bar{q}\rangle + \grave{P}^{(1)}) + tR$, which satisfies $\|X_{R(t)}\|_{\mathcal{D}_3} \le c\varepsilon$. Then $P^{(1)}$ can be written as

$$P^{(1)} = \grave{P}^{(1)} + \int_0^1 \{R(t) + \breve{P},\, F^{(1)}\} \circ \Phi^t_{F^{(1)}}\, dt + (P-R) \circ \Phi^1_{F^{(1)}}.$$

Hence,

$$X_{P^{(1)} - \grave{P}^{(1)}} \;=\; \int_0^1 (\Phi^t_{F^{(1)}})^* X_{\{R(t)+\breve{P},\, F^{(1)}\}}\, dt \;+\; (\Phi^1_{F^{(1)}})^* X_{(P-R)}.$$

By Lemma A.4,

$$\|X_{\{R(t)+\breve{P},\, F^{(1)}\}}\|_{\mathcal{D}_{2\eta}} \le c\eta^{-2}\varepsilon^{\frac{7}{4}} = \varepsilon^{\frac{27}{20}}.$$

Then, combining with (4.16), recalling the conclusion of Lemma 4.1 and 4.2,

$$\|X_{P^{(1)}}\|_{\mathcal{D}^{(1)}, \mathcal{O}_+} \le \frac{1}{2}\varepsilon^{\frac{6}{5}} + 2\varepsilon^{\frac{5}{4}} + 2c\varepsilon^{\frac{27}{20}} \le \varepsilon^{\frac{6}{5}} = \varepsilon^{(1)}.$$

Now we need to show $P^{(1)}$ satisfies assumptions **(A4)** and **(A5)**. Note that

$$\begin{aligned} P^{(1)} \;=\;& \grave{P}^{(1)} + P - R + \{\breve{P}, F^{(1)}\} + \{P, F^{(1)}\} \\ &+ \frac{1}{2!}\{\{\mathcal{N}, F^{(1)}\}, F^{(1)}\} + \frac{1}{2!}\{\{\breve{P}, F^{(1)}\}, F^{(1)}\} + \frac{1}{2!}\{\{P, F^{(1)}\}, F^{(1)}\} + \cdots \\ &+ \frac{1}{n!}\{\cdots\{\mathcal{N}, \underbrace{F^{(1)}\}\cdots, F^{(1)}}_{n}\} + \frac{1}{n!}\{\cdots\{\breve{P}, \underbrace{F^{(1)}\}\cdots, F^{(1)}}_{n}\} \\ &+ \frac{1}{n!}\{\cdots\{P, \underbrace{F^{(1)}\}\cdots, F^{(1)}}_{n}\} + \cdots. \end{aligned}$$

Since all of $\mathcal{N}$, $\breve{P}$, P, $F^{(1)}$, $\grave{P}^{(1)}$ have gauge invariance, independent of variables $(q_n, \bar{q}_n)_{n\in\mathcal{J}}$, so does $P^{(1)}$ due to Lemma A.5 and A.6 in Appendix.

For $P - R = \sum_{2|l|+|\alpha|+|\beta|\ge 3} P_{kl\alpha\beta} e^{\mathrm{i}\langle k,\theta\rangle} I^l q^\alpha \bar{q}^\beta$, we have

$$\|P_{\alpha\beta}\|_{\mathcal{D}^{(1)}} \le \begin{cases} \frac{1}{4}\varepsilon^{(1)} e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta| \le 2 \\ e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta| \ge 3 \end{cases}.$$

Here we applied the estimate $|I| \le s^{(1)} \le \frac{1}{4}\varepsilon^{(1)}$ to handle the case that $|\alpha|+|\beta| \le 2$ and $2|l|+|\alpha|+|\beta| \ge 3$.

The decay property of remaining terms, which are made up of several Poisson brackets, is covered by the following lemmas.

Lemma 4.3 *For ε sufficiently small, $\{P, F^{(1)}\}$ satisfies*

$$\|\{P, F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_{3\eta}, \mathcal{O}_+} \le \begin{cases} \varepsilon^{\frac{5}{4}} e^{-\rho^{(1)} n^*_{\alpha\beta}}, & |\alpha|+|\beta| \le 2 \\ \varepsilon^{\frac{1}{4}} e^{-\rho^{(1)} n^*_{\alpha\beta}}, & |\alpha|+|\beta| \ge 3 \end{cases}.$$

Proof: A straightforward calculation yields that

$$\{P,F^{(1)}\}_{\alpha\beta} \;=\; \mathrm{i}\sum_{\substack{n\in\mathbb{Z}\\ (\check\alpha,\check\beta)+(\hat\alpha,\hat\beta)=(\alpha,\beta)}} \left(P_{\check\alpha+e_n,\check\beta}F^{(1)}_{\hat\alpha,\hat\beta+e_n} - P_{\check\alpha,\check\beta+e_n}F^{(1)}_{\hat\alpha+e_n,\hat\beta}\right) \tag{4.49}$$

$$+\sum_{(\check\alpha,\check\beta)+(\hat\alpha,\hat\beta)=(\alpha,\beta)} \left\{P_{\check\alpha\check\beta},F^{(1)}_{\hat\alpha\hat\beta}\right\}. \tag{4.50}$$

- Terms in (4.49)
  Let us consider terms $P_{\check\alpha+e_n,\check\beta}F^{(1)}_{\hat\alpha,\hat\beta+e_n}$ first.

  i) $|\alpha|+|\beta|\le 2$

  Since $|\hat\alpha|+|\hat\beta+e_n|=1$ or 2 in view of the construction of $F^{(1)}$, we have that

$$|\check\alpha+e_n|+|\check\beta| = |\alpha|+|\beta|+1-(|\hat\alpha|+|\hat\beta|)\le 3. \tag{4.51}$$

  If $|\check\alpha+e_n|+|\check\beta|\le 2$, then, noting that $n^*_{\alpha\beta}\le\max\{n^*_{\check\alpha+e_n,\check\beta},\, n^*_{\hat\alpha,\hat\beta+e_n}\}$, we have

$$\|P_{\check\alpha+e_n,\check\beta}F^{(1)}_{\hat\alpha,\hat\beta+e_n}\|_{\mathcal{D}_3,\mathcal{O}_+}\le \varepsilon e^{-\rho n^*_{\check\alpha+e_n,\check\beta}}\cdot\varepsilon^{\frac34}e^{-\rho n^*_{\hat\alpha,\hat\beta+e_n}}\le\varepsilon^{\frac74}e^{-\rho n^*_{\alpha\beta}}. \tag{4.52}$$

  If $|\check\alpha+e_n|+|\check\beta|=3$, then, by (4.51), $(\hat\alpha,\hat\beta)=(0,0)$, $(\check\alpha,\check\beta)=(\alpha,\beta)$. By the definition of norm $\|X_P\|_{\mathcal{D},\mathcal{O}}$ and the construction of $F^{(1)}$,

$$\|P_{\alpha+e_n,\beta}\|_{\mathcal{D}_3,\mathcal{O}}\le e^{-\rho n^*_{\alpha+e_n,\beta}},\quad \|F^{(1)}_{0,e_n}\|_{\mathcal{D}_3,\mathcal{O}_+}\le s\varepsilon^{\frac34}e^{-\rho|n|}.$$

  Thus, noting that $n^*_{\alpha\beta}\le\max\{n^*_{\alpha+e_n,\beta},\,|n|\}$,

$$\|P_{\alpha+e_n,\beta}F^{(1)}_{0,e_n}\|_{\mathcal{D}_3,\mathcal{O}_+}\le s\varepsilon^{\frac34}e^{-\rho n^*_{\alpha\beta}}\le\frac14\varepsilon^{\frac74}e^{-\rho n^*_{\alpha\beta}}. \tag{4.53}$$

  ii) $|\alpha|+|\beta|\ge 3$

  By the same argument as above,

$$\|P_{\check\alpha+e_n,\check\beta}F^{(1)}_{\hat\alpha,\hat\beta+e_n}\|_{\mathcal{D}_3,\mathcal{O}_+}\le e^{-\rho n^*_{\check\alpha+e_n,\check\beta}}\cdot\varepsilon^{\frac34}e^{-\rho n^*_{\hat\alpha,\hat\beta+e_n}}\le\varepsilon^{\frac34}e^{-\rho n^*_{\alpha\beta}}. \tag{4.54}$$

  Doing the same for $P_{\check\alpha,\check\beta+e_n}F^{(1)}_{\hat\alpha+e_n,\hat\beta}$, we finish estimates for terms in (4.49).

- Terms in (4.50)
  By Lemma A.3 and the inequality $n^*_{\alpha\beta}\le\max\{n^*_{\check\alpha\check\beta},\,n^*_{\hat\alpha\hat\beta}\}$, we have

$$\|\{P_{\check\alpha\check\beta},F^{(1)}_{\hat\alpha\hat\beta}\}\|_{\mathcal{D}_{3\eta}}\le c(r-r^{(1)})^{-1}\eta^{-2}\begin{cases}\varepsilon^{\frac74}e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta|\le 2\\ \varepsilon^{\frac34}e^{-\rho n^*_{\alpha\beta}}, & |\alpha|+|\beta|\ge 3\end{cases}. \tag{4.55}$$

Combining (4.52) – (4.55), there exists $c_8>0$ such that

$$\|\{P,F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_{3\eta}}\le c_8(r-r^{(1)})^{-1}\eta^{-2}(\rho-\rho^{(1)})^{-2}\begin{cases}\varepsilon^{\frac74}e^{-\rho^{(1)}n^*_{\alpha\beta}}, & |\alpha|+|\beta|\le 2\\ \varepsilon^{\frac34}e^{-\rho^{(1)}n^*_{\alpha\beta}}, & |\alpha|+|\beta|\ge 3\end{cases},$$

applying the fact that $|\hat\alpha|+|\hat\beta|\le 2$. Moreover, if

(C8) $c_8(r-r^{(1)})^{-1}\eta^{-2}(\rho-\rho^{(1)})^{-2}\varepsilon^{\frac{1}{2}} \le \frac{1}{4}$,

Lemma 4.3 is proved. ■

By (4.15), (4.19) and (4.20), it is evident that the coefficients of

$$\{\mathcal{N}, F^{(1)}\} = e^{(1)} + \langle \omega^{(1)}, I\rangle + \langle W^{(1)}q, \bar{q}\rangle + \grave{P}^{(1)} - R$$

satisfies $\|\{\mathcal{N}, F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_3,\mathcal{O}_+} \le c\varepsilon e^{-\rho^{(1)}n^*_{\alpha\beta}}$. Then we have the following lemma, whose proof is analogous to that of Lemma 4.3.

Lemma 4.4 *For ε sufficiently small, $\{\{\mathcal{N}, F^{(1)}\}, F^{(1)}\}$ satisfies*

$$\|\{\{\mathcal{N}, F^{(1)}\}, F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_{3\eta},\mathcal{O}_+} \le \frac{1}{4}\varepsilon^{\frac{6}{5}} e^{-\rho^{(1)}n^*_{\alpha\beta}}.$$

Lemma 4.5 *For ε sufficiently small, $\{\breve{P}, F^{(1)}\}$ satisfies*

$$\|\{\breve{P}, F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_3,\mathcal{O}_+} \le \varepsilon^{\frac{1}{4}} e^{-\rho^{(1)}n^*_{\alpha\beta}}, \quad |\alpha|+|\beta| \ge 3.$$

Proof: It can be calculated that

$$\{\breve{P}, F^{(1)}\}_{\alpha\beta} = \mathrm{i} \sum_{\substack{n\in\mathbb{Z} \\ (\breve{\alpha},\breve{\beta})+(\hat{\alpha},\hat{\beta})=(\alpha,\beta)}} \left(\breve{P}_{\breve{\alpha}+e_n,\breve{\beta}} F^{(1)}_{\hat{\alpha},\hat{\beta}+e_n} - \breve{P}_{\breve{\alpha},\breve{\beta}+e_n} F^{(1)}_{\hat{\alpha}+e_n,\hat{\beta}}\right). \tag{4.56}$$

For $\breve{P}_{\breve{\alpha}+e_n,\breve{\beta}} F^{(1)}_{\hat{\alpha},\hat{\beta}+e_n}$ in (4.56), since $|\hat{\alpha}|+|\hat{\beta}+e_n| = 1$ or 2 and $|\breve{\alpha}+e_n|+|\breve{\beta}| \ge 4$ here, it is obvious that $|\alpha|+|\beta| = |\breve{\alpha}|+|\breve{\beta}|+|\hat{\alpha}|+|\hat{\beta}| \le 3$.

Note that $n^*_{\alpha\beta} \le \max\{n^*_{\breve{\alpha}+e_n,\breve{\beta}},\, n^*_{\hat{\alpha},\hat{\beta}+e_n}\}$, and

$$n^*_{\breve{\alpha}+e_n,\breve{\beta}} = \max\{n^+_{\breve{\alpha}+e_n,\breve{\beta}},\, -n^-_{\breve{\alpha}+e_n,\breve{\beta}}\}, \quad n^*_{\hat{\alpha},\hat{\beta}+e_n} = \max\{n^+_{\hat{\alpha},\hat{\beta}+e_n},\, -n^-_{\hat{\alpha},\hat{\beta}+e_n}\}.$$

Then $n^+_{\breve{\alpha}+e_n,\breve{\beta}} - n^-_{\breve{\alpha}+e_n,\breve{\beta}} + n^*_{\hat{\alpha},\hat{\beta}+e_n} \ge n^*_{\alpha\beta}$, and hence

$$\left\|\breve{P}_{\breve{\alpha}+e_n,\breve{\beta}} F^{(1)}_{\hat{\alpha},\hat{\beta}+e_n}\right\|_{\mathcal{D}_3,\mathcal{O}_+} \le e^{-\rho(n^+_{\breve{\alpha}+e_n,\breve{\beta}} - n^-_{\breve{\alpha}+e_n,\breve{\beta}})} \cdot \varepsilon^{\frac{3}{4}} e^{-\rho n^*_{\hat{\alpha},\hat{\beta}+e_n}} \le \varepsilon^{\frac{3}{4}} e^{-\rho n^*_{\alpha\beta}}.$$

Doing the estimate for $\breve{P}_{\breve{\alpha},\breve{\beta}+e_n} F^{(1)}_{\hat{\alpha}+e_n,\hat{\beta}}$ in (4.56) similarly, we have that

$$\|\{\breve{P}, F^{(1)}\}_{\alpha\beta}\|_{\mathcal{D}_3,\mathcal{O}_+} \le c_8(\rho-\rho^{(1)})^{-2}\varepsilon^{\frac{3}{4}} e^{-\rho^{(1)}n^*_{\alpha\beta}} \le \varepsilon^{\frac{1}{4}} e^{-\rho^{(1)}n^*_{\alpha\beta}}, \quad |\alpha|+|\beta| \ge 3,$$

if **(C8)** holds. ■

Summarize the analysis above, then the decay property for $P^{(1)}$ can be expressed as

Proposition 3 *For ε sufficiently small, $P^{(1)} = \sum_{\alpha,\beta} P^{(1)}_{\alpha\beta}(\theta, I; \xi) q^\alpha \bar{q}^\beta$ satisfies*

$$\|P^{(1)}_{\alpha\beta}\|_{\mathcal{D}^{(1)},\mathcal{O}_+} \le \begin{cases} \varepsilon^{(1)} e^{-\rho^{(1)}n^*_{\alpha\beta}}, & |\alpha|+|\beta| \le 2 \\ e^{-\rho^{(1)}n^*_{\alpha\beta}}, & |\alpha|+|\beta| \ge 3 \end{cases}.$$

4.4 A succession of symplectic transformations

With the verification of assumptions **(A4)** and **(A5)** completed, we finish one sub-step of KAM iteration. Suppose that we have arrived at the j^{th} sub-step, $j = 1, \cdots, J$, with $J = \left[\frac{5}{2}\varepsilon^{\frac{a}{2}}\right]$, then we encounter the Hamiltonian

$$\begin{aligned} H^{(j-1)} &= H \circ \Phi^1_{F^{(1)}} \circ \cdots \circ \Phi^1_{F^{(j-1)}} \\ &= \mathcal{N} + \breve{P} + \sum_{i=1}^{j-1} \left(e^{(i)} + \langle \omega^{(i)}, I \rangle + \langle W^{(i)} q, \bar{q} \rangle \right) + P^{(j-1)}, \end{aligned}$$

with the superscript "(0)" labeling quantities before the 1^{st} sub-step in particular. Let

$$R^{(j-1)} := \sum_{\substack{k \\ 2|l|+|\alpha|+|\beta|\le 2}} P^{(j-1)}_{kl\alpha\beta} e^{\mathrm{i}\langle k,\theta\rangle} I^l q^\alpha \bar{q}^\beta. \tag{4.57}$$

As demonstrated in Proposition 2, on $\mathcal{O}_+$, the following homological equation

$$\{\mathcal{N}, F^{(j)}\} + R^{(j-1)} = e^{(j)} + \langle \omega^{(j)}, I \rangle + \langle W^{(j)} q, \bar{q} \rangle + \grave{P}^{(j)}, \tag{4.58}$$

can be solved, with $F^{(j)}$, $e^{(j)}$, $\omega^{(j)}$, $W^{(j)}$, $\grave{P}^{(j)}$ having properties similar to $F^{(1)}$, $e^{(1)}$, $\omega^{(1)}$, $W^{(1)}$, $\grave{P}^{(1)}$ respectively. Then we obtain

$$H^{(j)} = H^{(j-1)} \circ \Phi^1_{F^{(j)}} = \mathcal{N} + \breve{P} + \sum_{i=1}^{j} \left(e^{(i)} + \langle \omega^{(i)}, I \rangle + \langle W^{(i)} q, \bar{q} \rangle \right) + P^{(j)}.$$

The estimates for $F^{(j)}$ and the verification of assumptions for $P^{(j)}$ can be done similarly as in subsection 4.3.

The process above can be summarized as

Proposition 4 *Consider the Hamiltonian H in (4.1). There exist J symplectic transformations $\Phi^{(j)} = \Phi^1_{F^{(j)}}$, $j = 1, \cdots, J$, generated by the corresponding real-analytic Hamiltonians $F^{(j)}$ respectively, such that*

$$H^{(j)} = H \circ \Phi^{(1)} \circ \cdots \circ \Phi^{(j)} = \mathcal{N} + \breve{P} + G_j + P^{(j)},$$

is real-analytic on $\mathcal{D}^{(j)} = \mathcal{D}_{d,\rho_+}(r^{(j)}, s^{(j)})$, with $G_j = \sum_{i=1}^{j} \left(e^{(i)} + \langle \omega^{(i)}, I \rangle + \langle W^{(i)} q, \bar{q} \rangle \right)$. For $i = 1, 2, 3, 4$, $\eta = \varepsilon^{\frac{1}{5}}$, let

$$\begin{aligned} \mathcal{D}^{(j)}_i &= \mathcal{D}_{d,\rho_+}(r^{(j+1)} + \frac{i}{4}(r^{(j)} - r^{(j+1)}), \frac{i}{4} s^{(j)}), \\ \mathcal{D}^{(j)}_{i\eta} &= \mathcal{D}_{d,\rho_+}(r^{(j+1)} + \frac{i}{4}(r^{(j)} - r^{(j+1)}), \frac{i}{4} \eta s^{(j)}). \end{aligned}$$

(a) *With $R^{(j-1)}$ defined in (4.57), $F^{(j)}$ satisfies the homological equation (4.58) on $\mathcal{O}_+$, $\|X_{F^{(j)}}\|_{\mathcal{D}^{(j-1)}_3, \mathcal{O}_+} \le \varepsilon^{-\frac{1}{4}} \varepsilon^{(j-1)}$, $\Phi^t_{F^{(j)}} : \mathcal{D}^{(j-1)}_{2\eta} \to \mathcal{D}^{(j-1)}_{3\eta}$, $-1 \le t \le 1$, and*

$$\|D\Phi^t_{F^{(j)}} - I\|_{\mathcal{D}^{(j-1)}_{1\eta}} < 2\varepsilon^{-\frac{1}{4}} \varepsilon^{(j-1)},$$

$$\|F^{(j)}_{\alpha\beta}\|_{\mathcal{D}^{(j-1)}_3, \mathcal{O}_+} \le \begin{cases} \varepsilon^{-\frac{1}{4}} \varepsilon^{(j-1)} e^{-\rho^{(j-1)} n^*_{\alpha\beta}}, & |\alpha| + |\beta| \le 2 \\ 0, & |\alpha| + |\beta| \ge 3 \end{cases},$$

$$\partial_{q_n} F^{(j)} = \partial_{\bar{q}_n} F^{(j)} \equiv 0, \quad \forall n \in \mathcal{J}.$$

(b) G_j *satisfies that* $\|X_{G_j}\|_{\mathcal{D}_3^{(j)},\mathcal{O}_+} \le c\varepsilon$ *and for* $i = 1, 2, \cdots, j$,

$$|\omega^{(i)}|_{\mathcal{O}_+} \le \varepsilon^{(i-1)},$$
$$|W^{(i)}_{mn}|_{\mathcal{O}_+} \le \begin{cases} \varepsilon^{(i-1)} e^{-\rho^{(i-1)} \max\{|m|,|n|\}}, & |m|, |n| \le N_+, \quad m, n \notin \mathcal{J} \\ 0, & \textit{otherwise} \end{cases}.$$

(c) $P^{(j)}$ *satisfies* $\|X_{\breve{P}+P^{(j)}}\|_{\mathcal{D}^{(j)},\mathcal{O}_+} \le \varepsilon^{(j)}$ *and assumptions* **(A4)**, **(A5)**, *which include*

$$\|P^{(j)}_{\alpha\beta}\|_{\mathcal{D}^{(j)},\mathcal{O}_+} \le \begin{cases} \varepsilon^{(j)} e^{-\rho^{(j)} n^*_{\alpha\beta}}, & |\alpha| + |\beta| \le 2 \\ e^{-\rho^{(j)} n^*_{\alpha\beta}}, & |\alpha| + |\beta| \ge 3 \end{cases},$$
$$\partial_{q_n} P^{(j)} = \partial_{\bar{q}_n} P^{(j)} \equiv 0, \quad \forall n \in \mathcal{J}.$$

Let $s_+ = s^{(J)} = 2^{-3J} \varepsilon^{\frac{J}{5}} s$, $\Phi = \Phi^{(1)} \circ \cdots \circ \Phi^{(J)}$, and

$$\mathcal{N}_+ = e_+ + \langle \omega_+, I \rangle + \langle \Omega_+ q, \bar{q} \rangle,$$

with $\Omega_+ = T + A + W_+$, and

$$e_+ = e + \sum_{j=1}^{J} e^{(j)}, \quad \omega_+ = \omega + \sum_{j=1}^{J} \omega^{(j)}, \quad W_+ = W + \sum_{j=1}^{J} W^{(j)}.$$

Then $\Phi : \mathcal{D}_+ \times \mathcal{O}_+ \to \mathcal{D} \times \mathcal{O}$. From the estimates of $\omega^{(j)}$ and $W^{(j)}$, we have

$$|\omega_+ - \omega|_{\mathcal{O}_+} \le c\varepsilon, \tag{4.59}$$
$$|(W_+ - W)_{mn}|_{\mathcal{O}_+} \le \begin{cases} \varepsilon^{\frac{1}{2}} e^{-\frac{\rho}{2} \max\{|m|,|n|\}}, & |m|, |n| \le N_+, \quad m, n \notin \mathcal{J} \\ 0, & \text{otherwise} \end{cases}. \tag{4.60}$$

Since $W^* = W$ and $(W^{(i)})^* = W^{(i)}$, W_+ is still a Hermitian matrix. Then, by (4.59) and (4.60), (**A1**) and (**A2**) hold with $p_+ = p + \varepsilon^{\frac{1}{2}}$ and $\sigma_+ := \frac{1}{3}\rho$.

Let $P_+ = P^{(J)}$. It has been verified that the assumptions **(A4)** and **(A5)** for $P^{(J)}$ hold, which is an analogue to the process in subsection 4.3.

This completes one step of KAM iterations.

5 Proof of the KAM theorem

With $\varepsilon_0 = \epsilon^{\frac{1}{4}}$, $\sigma_0 = 1$, $\hat{N} = |\ln \varepsilon_0|$, and

$$M_0 = \max\left\{ 2^{\tilde{s}+4} C \frac{L^{\tilde{s}+1}((\tilde{s}+1)!)^2}{\tilde{\xi}}, 2\tilde{\tau}, 8, \frac{12(2\tau + b + 3)}{\tilde{\tau}} \right\}, \quad N_0 = 6|\ln \varepsilon_0|, \quad \rho_0 = N_0^{-1},$$

one can define the following sequences as in [11],

$$M_{\nu+1} = M_\nu^{\tilde{s} M_\nu^3}, \quad a_\nu = \frac{1}{\tilde{\tau}} M_\nu^{-3\tilde{s} M_\nu^3}, \quad \varepsilon_{\nu+1} = \varepsilon_\nu^{\frac{1}{2}\varepsilon_\nu^{-a_\nu/2}},$$
$$N_{\nu+1} = \varepsilon_\nu^{-a_\nu}, \quad \rho_{\nu+1} = \varepsilon_\nu^{a_\nu}, \quad \sigma_{\nu+1} = \frac{1}{3}\rho_\nu.$$

Given $p_0=\varepsilon_0^{\frac{1}{2}}$, $r_0=r$, $s_0=s$, the other sequences are defined as

$$p_{\nu+1}=p_\nu+\varepsilon_\nu^{\frac{1}{2}}, \qquad K_{\nu+1}=N_{\nu+1}-(M_\nu+1)N_\nu, \quad J_\nu=\left[\frac{5}{2}\varepsilon_\nu^{-\frac{a_\nu}{2}}\right],$$
$$r_\nu=r_0\left(1-\sum_{i=2}^{\nu+1}2^{-i}\right), \quad s_{\nu+1}=2^{-3J_\nu}\varepsilon_\nu^{\frac{J_\nu}{5}}s_\nu, \qquad \gamma_\nu=\varepsilon_\nu^{\frac{1}{80}}.$$

Let $\mathcal{D}_\nu$ and $\mathcal{O}_\nu$ be as defined in Section 4.

5.1 Iteration lemma

The preceding analysis can be summarized as follows.

Lemma 5.1 *There exists ε_0 sufficiently small such that the following holds for all $\nu=0,1,\cdots$.*

(a) *$H_\nu=\mathcal{N}_\nu+\breve{P}+P_\nu$ is real-analytic on $\mathcal{D}_\nu$, and C^1_W parametrized by $\xi\in\mathcal{O}_\nu$, where*

$$\mathcal{N}_\nu=e_\nu+\langle\omega_\nu,I\rangle+\langle\Omega_\nu q,\bar{q}\rangle, \quad P_\nu=\sum_{\alpha,\beta}(P_\nu)_{\alpha\beta}(\theta,I)q^\alpha\bar{q}^\beta,$$

with $\Omega_\nu=T+A+W_\nu$ satisfying

$$(\Omega_\nu)_{mn}\equiv 0 \quad if \quad m \quad or \quad n\in\mathcal{J},$$
$$|(W_\nu)_{mn}|_{\mathcal{O}_\nu}\le\begin{cases} p_\nu e^{-\sigma_\nu\max\{|m|,|n|\}}, & |m|,|n|\le N_\nu \\ 0, & otherwise\end{cases},$$
$$|\omega_{\nu+1}-\omega_\nu|_{\mathcal{O}_{\nu+1}}\le\varepsilon_\nu,$$
$$|(W_{\nu+1}-W_\nu)_{mn}|_{\mathcal{O}_{\nu+1}}\le\begin{cases}\varepsilon_\nu^{\frac{1}{2}}e^{-\frac{\rho_\nu}{2}\max\{|m|,|n|\}}, & |m|,|n|\le N_{\nu+1}, \quad m,n\notin\mathcal{J} \\ 0, & otherwise\end{cases}.$$

Moreover, P_ν has gauge invariance and $\|X_{\breve{P}+P_\nu}\|_{\mathcal{D}_\nu,\mathcal{O}_\nu}\le\varepsilon_\nu$,

$$\|(P_\nu)_{\alpha\beta}\|_{\mathcal{D}_\nu,\mathcal{O}_\nu}\le\begin{cases}\varepsilon_\nu e^{-\rho_\nu n^*_{\alpha\beta}}, & |\alpha|+|\beta|\le 2\\ e^{-\rho_\nu n^*_{\alpha\beta}}, & |\alpha|+|\beta|\ge 3\end{cases},$$
$$\partial_{q_n}P_\nu=\partial_{\bar{q}_n}P_\nu\equiv 0, \quad \forall n\in\mathcal{J}.$$

(b) *For each ν, there is a symplectic transformation $\Phi_\nu:\mathcal{D}_{\nu+1}\to\mathcal{D}_\nu$ with*

$$\|D\Phi_\nu-Id\|_{\mathcal{D}_{\nu+1},\mathcal{O}_{\nu+1}}\le\varepsilon_\nu^{\frac{1}{2}},$$

such that $H_{\nu+1}=H_\nu\circ\Phi_\nu$.

Proof: Let $c_0:=8e^{20}\max\{c_1,\cdots,c_8\}$. We need to verify the assumptions $(\mathbf{C1})-(\mathbf{C8})$ for $\nu=0,1,\cdots$. By noting that

$$N_{\nu+1}=\varepsilon_\nu^{a_\nu}=\rho_{\nu+1}^{-1}, \quad \sigma_{\nu+1}=\frac{1}{3}\rho_\nu, \quad r_\nu^{(j)}-r_\nu^{(j+1)}=\frac{r_\nu-r_{\nu+1}}{2J_\nu}, \quad \rho_\nu^{(j)}-\rho_\nu^{(j+1)}=\frac{\rho_\nu-\rho_{\nu+1}}{2J_\nu},$$

it is sufficient for us to check:

(D1) $c_0 s_\nu \le \varepsilon_\nu$,

(D2) $c_0 \left(\frac{r_\nu - r_{\nu+1}}{2J_\nu}\right)^{-(2\tau+b+1)} \left(\frac{\rho_\nu - \rho_{\nu+1}}{2J_\nu}\right)^{-2} \le \varepsilon_\nu^{-\frac{1}{20}}$,

(D3) $c_0 N_{\nu+1}^8 M_\nu^8 N_\nu^{20} e^{8M_\nu N_\nu \rho_\nu} \le \varepsilon_\nu^{-\frac{7}{40}}$,

(D4) $e^{-\frac{\rho_\nu K_{\nu+1}}{2J_\nu}} \le \varepsilon_\nu^{\frac{2}{5}}$,

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

By the choice of s_0, the condition (**D1**) clearly holds for $\nu = 0$. Suppose that it holds for some ν, then it is easy to see that

$$c_0 s_{\nu+1} = 2^{-3J_\nu} \varepsilon_\nu^{\frac{J_\nu}{5}} \cdot c_0 s_\nu < 2^{-3J_\nu} \varepsilon_\nu^{\frac{J_\nu}{5}} \cdot \varepsilon_\nu < \varepsilon_{\nu+1}.$$

Hence (**D1**) holds for all ν.

Let us first take ε_0 sufficiently small such that

$$\varepsilon_0^{\frac{1}{20} - \frac{1}{2}a_0(2\tau+b+3)} \le \frac{1}{c_0}\left(\frac{r_0}{20}\right)^{2\tau+b+1} \left(\frac{1-\varepsilon_0^{a_0}}{5}\right)^2.$$

Here we have applied $M_0 \ge \frac{12}{\tilde{\tau}}(2\tau+b+3)$ and $a_0 = M_0^{-3\tilde{s}M_0^3}$ such that $\frac{1}{20} - \frac{1}{2}a_0(2\tau+b+3) > 0$. Then, recalling that $r_\nu - r_{\nu+1} = \frac{r_0}{2^{2+\nu}}$ and $J_\nu = \left[\frac{5}{2}\varepsilon_\nu^{-\frac{a_\nu}{2}}\right]$,

$$c_0 \left(\frac{r_0 - r_1}{2J_0}\right)^{-(2\tau+b+1)} \left(\frac{\rho_0 - \rho_1}{2J_0}\right)^{-2} \le \varepsilon_0^{-\frac{1}{20}},$$

i.e., (**D2**) holds for $\nu = 0$. Since for $\nu \ge 1$ and for ε_0 sufficiently small,

$$\varepsilon_\nu^{\frac{1}{40} - \frac{1}{2}a_\nu(2\tau+b+3)} \ll \varepsilon_0^{\left(\frac{6}{5}\right)^\nu} \ll \frac{1}{2^{\nu(2\tau+b+1)}c_0}\left(\frac{r_0}{20}\right)^{2\tau+b+1}, \quad \varepsilon_\nu^{\frac{1}{40}} \ll \left(\frac{\varepsilon_{\nu-1}^{a_{\nu-1}} - \varepsilon_\nu^{a_\nu}}{5}\right)^2,$$

we have

$$c_0 \left(\frac{r_\nu - r_{\nu+1}}{2J_\nu}\right)^{-(2\tau+b+1)} \left(\frac{\rho_\nu - \rho_{\nu+1}}{2J_\nu}\right)^{-2} \le \varepsilon_\nu^{-\frac{1}{20}}.$$

Thus, (**D2**) holds true.

In Section 6 of [11], the basic smallness assumption of ε_ν, i.e., the inequality (A.1) in Lemma A.1, has been verified, then all other assumptions are immediate, including the inequality

$$\Gamma_\nu N_\nu^2 e^{6M_\nu N_\nu \rho_\nu} \le \varepsilon_\nu^{-\frac{1}{8}},$$

where Γ_ν increases superexponentially in M_ν. Since all of M_ν, N_ν, ρ_ν and ε_ν here are defined in the same way as [11], we can apply this inequality. So (**D3**) has been verified.

By the definition of ρ_ν, a_ν and ε_ν, we have

$$\rho_\nu \varepsilon_\nu^{-\frac{1}{2}a_\nu} > \ln\frac{1}{\varepsilon_\nu}.$$

Then we see that (**D4**) holds for $\nu = 0, 1, \cdots$. ■

5.2 Convergence

Now we fix $x \in \tilde{\mathcal{X}}$, with $\tilde{\mathcal{X}}$ defined as in Proposition 1. This means that the blocks mentioned in Proposition 1 are eventually stationary after some step, i.e., for each $n \in \mathbb{Z}$, there is a $\nu_0(n)$ such that

$$\Lambda^{\nu+1}(n) = \Lambda^{\nu}(n), \quad \forall \nu \geq \nu_0(n).$$

In this case, the local decay rate for n may not shrink with ν necessarily(ρ_ν is the global upper bound of the rates for all $n \in \mathbb{Z}$).

Define $\Psi^\nu = \Phi_0 \circ \Phi_1 \circ \cdots \circ \Phi_{\nu-1}$, $\nu = 1, 2, \cdots$. An induction argument shows that $\Psi^\nu : \mathcal{D}_{\nu+1} \to \mathcal{D}_0$, and

$$H_0 \circ \Psi^\nu = H_\nu = \mathcal{N}_\nu + \breve{P} + P_\nu.$$

Let $\mathcal{O}_{\varepsilon_0} = \cap_{\nu=0}^{\infty} \mathcal{O}_\nu$. As in standard arguments (e.g. [24, 29]), thanks to Lemma 4.2, it concludes that H_ν, $\mathcal{N}_\nu$, P_ν, Ψ^ν, e_ν, ω_ν and W_ν converge uniformly on $\mathcal{D}_{d,0}(\frac{1}{2}r_0, 0) \times \mathcal{O}_{\varepsilon_0}$ to, say, H_∞, $\mathcal{N}_\infty$, P_∞, Ψ^∞, e_∞, ω_∞ and W_∞ respectively, in which case it is clear that

$$\mathcal{N}_\infty = e_\infty + \langle \omega_\infty, I \rangle + \langle (T + A + W_\infty) q, \bar{q} \rangle,$$

with $\Omega_\infty = T + A + W_\infty$ satisfying $(\Omega_\infty)_{mn} \equiv 0$ if m or $n \in \mathcal{J}$. Since $\|X_{P_\nu}\|_{\mathcal{D}_\nu, \mathcal{O}_\nu} \leq \varepsilon_\nu$ with $\varepsilon_\nu \to 0$, it follows that $\|X_{P_\infty}\|_{\mathcal{D}_{d,0}(\frac{1}{2}r_0,0), \mathcal{O}_{\varepsilon_0}} = 0$.

Since $H_0 \circ \Psi^\nu = H_\nu$, we have $\Phi^t_{H_0} \circ \Psi^\nu = \Psi^\nu \circ \Phi^t_{H_\nu}$, with $\Phi^t_{H_0}$ denoting the flow of the Hamiltonian vector field X_{H_0}. The uniform convergence of Ψ^ν and X_{H_ν} implies that one can pass the limit in the above and conclude that

$$\Phi^t_{H_0} \circ \Psi^\infty = \Psi^\infty \circ \Phi^t_{H_\infty}, \quad \Psi^\infty : \mathcal{D}_{d,0}(\frac{1}{2}r_0, 0) \to \mathcal{D}_0.$$

Hence,

$$\Phi^t_{H_0}(\Psi^\infty(\mathbb{T}^b \times \{\xi\})) \; = \; \Psi^\infty \Phi^t_{\mathcal{N}_\infty}(\mathbb{T}^b \times \{\xi\}) \; = \; \Psi^\infty(\mathbb{T}^b \times \{\xi\}), \quad \forall \xi \in \mathcal{O}_{\varepsilon_0}.$$

This means that $\Psi^\infty(\mathbb{T}^b \times \{\xi\})$ is an embedded invariant torus of the original perturbed Hamiltonian system at $\xi \in \mathcal{O}_{\varepsilon_0}$. Moreover, the frequencies $\omega_\infty(\xi)$ associated with $\Psi^\infty(\mathbb{T}^b \times \{\xi\})$ are slightly deformed from the unperturbed ones, $\omega(\xi)$.

5.3 Measure estimate

At the ν^{th} step of KAM iteration, we need to exclude the following resonant parameter set

$$\mathcal{R}^\nu_k := \mathcal{R}^{\nu 1}_k \bigcup \left(\bigcup_{n \in \Lambda^\nu} \mathcal{R}^{\nu 2}_{kn} \right) \bigcup \left(\bigcup_{m,n \in \Lambda^\nu} \mathcal{R}^{\nu 3}_{kmn} \right) \bigcup \left(\bigcup_{m,n \in \Lambda^\nu} \mathcal{R}^{\nu 4}_{kmn} \right), \quad k \neq 0$$

for any fixed $x \in \tilde{\mathcal{X}}$, where

$$\begin{aligned} \mathcal{R}^{\nu 1}_k &:= \left\{ \xi \in \mathcal{O}_\nu : |\langle k, \omega_\nu \rangle| < \frac{\gamma_\nu}{|k|^\tau} \right\}, \\ \mathcal{R}^{\nu 2}_{kn} &:= \left\{ \xi \in \mathcal{O}_\nu : |\langle k, \omega_\nu \rangle + \mu^\nu_n| < \frac{\gamma_\nu}{|k|^\tau N^2_{\nu+1}} \right\}, \end{aligned}$$

$$\mathcal{R}^{\nu 3}_{kmn} := \left\{\xi \in \mathcal{O}_\nu : |\langle k, \omega_\nu\rangle + \mu^\nu_m + \mu^\nu_n| < \frac{\gamma_\nu}{|k|^\tau N^4_{\nu+1}}\right\},$$
$$\mathcal{R}^{\nu 4}_{kmn} := \left\{\xi \in \mathcal{O}_\nu : |\langle k, \omega_\nu\rangle + \mu^\nu_m - \mu^\nu_n| < \frac{\gamma_\nu}{|k|^\tau N^4_{\nu+1}}\right\},$$

with $\{\mu^\nu_j\}_{j\in\Lambda^\nu}$ eigenvalues of $\tilde{D}^\nu_{\Lambda^\nu} + \tilde{A}_\nu$. It is clear $\mathcal{O}_0 \setminus \mathcal{O}_{\varepsilon_0} \subseteq \bigcup_{\nu\geq 0}\bigcup_{k\neq 0}\mathcal{R}^\nu_k$.

Recalling that ω_0 is a diffeomorphism of ξ, together with the estimates in (4.13), (4.59) and (4.60), we have

$$|\partial_\xi(\langle k, \omega_\nu\rangle + \mu^\nu_m - \mu^\nu_n)| \geq |\partial_\xi\langle k, \omega_0\rangle| - \varepsilon_0^{\frac{1}{4}}|k| - p = O(|k|)$$

for the set $\mathcal{R}^{\nu 4}_{kmn}$. The cases for $\mathcal{R}^{\nu 1}_k$, $\mathcal{R}^{\nu 2}_{kn}$, $\mathcal{R}^{\nu 3}_{kmn}$ can be handled in an entirely analogous way. Thus

$$\left|\mathcal{R}^{\nu 1}_k \bigcup \left(\bigcup_{n\in\Lambda^\nu}\mathcal{R}^{\nu 2}_{kn}\right)\bigcup\left(\bigcup_{m,n\in\Lambda^\nu}\mathcal{R}^{\nu 3}_{kmn}\right)\bigcup\left(\bigcup_{m,n\in\Lambda^\nu}\mathcal{R}^{\nu 4}_{kmn}\right)\right| \leq \frac{c\gamma_\nu}{|k|^{\tau+1}}.$$

Since $\tau \geq b$, we have that

$$|\mathcal{O}_0 \setminus \mathcal{O}_{\varepsilon_0}| \leq \left|\bigcup_{\nu\geq 0}\bigcup_{k\neq 0}\mathcal{R}^\nu_k\right| \leq c\sum_{\nu\geq 0}\sum_{k\neq 0}\frac{\gamma_\nu}{|k|^{\tau+1}} = c\sum_{\nu\geq 0}\gamma_\nu \sim \gamma_0 = \varepsilon_0^{\frac{1}{80}}.$$

A Appendix

A.1 Outline of the proof of Proposition 1

For any smooth function f defined on $\mathcal{I} \subset \mathbb{R}/\mathbb{Z}$, let $|f|_{C^j} := \max_{0\leq k\leq j}\sup_{x\in\mathcal{I}}\frac{1}{k!}|\partial^k_x f(x)|$.

The operator T in (2.1) can be viewed as a sum of two infinite-dimensional matrices, i.e., $\mathrm{diag}\{V(x+n\tilde{\alpha})\}_{n\in\mathbb{Z}} + \epsilon\Delta$ with Δ denoting the discrete Laplacian. It is natural to define an abstract normal form containing $\mathrm{diag}\{V(x+n\tilde{\alpha})\}_{n\in\mathbb{Z}}$.

Definition A.1 *Given a symmetric matrix D, smoothly parametrized by $x \in \mathbb{R}/\mathbb{Z}$ and satisfying the shift condition*

$$D_{m+k,n+k}(x) = D_{mn}(x + k\tilde{\alpha}), \quad \forall k \in \mathbb{Z},$$

where $\tilde{\alpha}$ is a Diophantine number, i.e., for some $\tilde{\gamma} > 0$ and $\tilde{\tau} > 1$,

$$|n\tilde{\alpha}|_1 \geq \frac{\tilde{\gamma}}{|n|^{\tilde{\tau}}}, \quad n \neq 0.$$

We say that D is in **normal form** *if the following conditions hold.*

(a) Short-range.

$$|D_{mn}|_{C^k} \leq \begin{cases} Ce^{-\rho|m-n|}L^k, & |m-n| \leq N \\ 0, & \textit{otherwise} \end{cases}, \quad \forall k \geq 0.$$

(b) Block diagonalization. *Fix any $x_* \in \mathbb{R}/\mathbb{Z}$. There exist an interval $\mathcal{I}$ centered in x_*, a disjoint decomposition $\bigcup_j \Lambda_j = \mathbb{Z}$ and a smooth orthogonal matrix Q on $\mathcal{I}$ such that* (b1) $\sharp\Lambda_j \leq M$ *and* $\mathrm{diam}\Lambda_j \leq MN$ *for each* j.

(b2) *the conjugated matrix* $\tilde{D} = Q^* D Q$ *is a product of commuting blocks*

$$\prod_j \tilde{D}_{\Lambda_j}(x), \quad \forall x \in \mathcal{I}.$$

(b3) $Q_{mn} \equiv 0$ *if* $|m-n| > N$. *Moreover, for all* m, $Q_{mn} \not\equiv 0$ *for at most* M *different* n.

(b4) $|Q|_{C^k} \leq L^k$ *for each* $k \geq 0$.

(c) Eigenvalues. *There is a piecewise smooth function* $E(x)$ *such that for each* j,

$$\{E(x_* + n\tilde{\alpha})\}_{n\in\Lambda_j} \;\; \textit{are the eigenvalues of} \;\; \tilde{D}_{\Lambda_j}(x_*),$$

and there are sets $\Omega_j \supset \Lambda_j$ *such that*

(c1) *for each* n, *if* $\inf_{l\in\Lambda_j} |E(x_*+l\tilde{\alpha}) - E(x_*+n\tilde{\alpha})| < \kappa$, *then*

$$x_* + n\tilde{\alpha} \in x_* + m\tilde{\alpha} + \frac{1}{2}(\mathcal{I} - x_*) \;\; for \;\; some \;\; m \in \Omega_j,$$

$$Q(x)(\mathbb{R}^{\Lambda(n)}) \subset \mathbb{R}^{\Omega_j + n - m}, \quad \forall x \in \mathcal{I}.$$

(c2) *the resultant*

$$u_{\Omega_j}(\varphi, x) = \mathrm{Res}\left(\det(D(x+\varphi)_{\Omega_j} - tI_{\Omega_j}), \det(D(x)_{\Omega_j} - tI_{\Omega_j})\right)^{5}$$

satisfies

$$|u_{\Omega_j}|_{C^k} < (4MC)^{2M^2} B^k, \quad \forall k \leq \tilde{s}M^2 + 1,^{6}$$

$$\max_{0\leq k\leq \tilde{s}M^2} \left| \frac{1}{\nu! B^k} \partial_\varphi^k u_{\Omega_j}(\varphi, x) \right| \geq \vartheta, \quad \forall \varphi, \, \forall x \in \mathbb{R}/\mathbb{Z}.$$

(c3) $\sharp\Omega_j \leq M$ *and* $\mathrm{diam}\Omega_j \leq \left(\frac{1}{\lambda}\right)^{\tilde{\tau}+2}$.

(c4) *the intervals* $\{n\tilde{\alpha} + \mathcal{I}\}_{\mathrm{dist}(n,\Omega_j)<N}$ *are pairwise disjoint.*

(c5) *for each* $\varphi \in \mathcal{I}$, $u_{\Omega_j}(\varphi, x)$ *satisfies*

$$|u_{\Omega_j}|_{C^k} < (2MC)^{2M^2} B^k, \quad \forall k \leq \tilde{s}M^2 + 1,^{7}$$

$$\max_{0\leq k\leq \tilde{s}M^2} \left| \frac{1}{\nu! B^k} \partial_x^k u_{\Omega_j}(\varphi, x) \right| \geq \vartheta \left(\prod_{m,n\in\Omega_j} |\varphi + (m-n)\tilde{\alpha}|_1 \right), \quad \forall x \in \mathbb{R}/\mathbb{Z}.$$

[5] The resultant of two monic polynomials P and Q is defined as the product $\mathrm{Res}(P,Q) = \prod_{\substack{P(x)=0 \\ Q(y)=0}} (x-y)$.

[6] The norm is with respect to the variable φ.

[7] The norm is with respect to the variable x.

Remark A.1 *Condition* (a) *implies an estimate of D in the operator norm on $\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$:*

$$\|D\|_{C^k} \le C\frac{e^\rho+1}{e^\rho-1}L^k \le C\frac{4}{\rho}L^k, \quad \forall k \ge 0,$$

if $\rho \le 1$.

Lemma A.1 (The inductive lemma of [11]) *Let D be in normal form on an interval $\mathcal{I} \subset \mathbb{R}/\mathbb{Z}$ with parameters C, L, ρ, M, N, κ, B, ϑ, λ, and let $a < g < h$ be numbers restricted by*

$$\frac{1}{\tilde{\tau}M^{3\tilde{s}M^3}} \le a < \frac{g}{20\tilde{s}\tilde{\tau}M^4} < \frac{h}{100\tilde{s}^2\tilde{\tau}M^8}, \quad h \le \frac{1}{5\tilde{s}M^{2\tilde{s}M^3}}.$$

Assume, as simplification, that

$$1 \le B \le L, \quad M \ge 8, \quad 1 < C < 2, \quad \rho, \kappa, \vartheta \le 1.$$

Let Z be a symmetric matrix, smoothly parametrized on $\mathbb{R}/\mathbb{Z}$, satisfying the shift condition. Assume that

$$\lambda \le |\mathcal{I}| \le \vartheta/B,$$

$$|Z_{mn}|_{C^k} < \varepsilon e^{-\rho|m-n|}L^k, \quad k \ge 0.$$

If there is a constant $\Gamma = \Gamma(\tilde{\gamma}, \tilde{\tau}, \tilde{s}, M)$, super-exponentially decaying in M, such that

$$|\varepsilon| < \Gamma\left[\frac{\rho^{\tilde{\tau}}\kappa\vartheta\lambda^{\tilde{\tau}^2}}{LN^{\tilde{\tau}}}e^{-N\rho}\right]^{e^{e^{\tilde{s}M^4}}}, \tag{A.1}$$

then there is a smooth orthogonal matrix $\tilde{U}$, satisfying the shift condition, such that

$$|(\tilde{U}-I)_{mn}|_{C^k} < \varepsilon^{\frac{1}{2}}e^{-\rho'|m-n|}L'^k$$

and

$$\tilde{U}^*(D+Z)\tilde{U} = D' + Z',$$

with Z' a symmetric matrix, smoothly parametrized on $\mathbb{R}/\mathbb{Z}$, satisfying the shift condition, and D' in normal form on an interval $\mathcal{I}' \subset \mathcal{I}$, with parameters

$$\begin{array}{lll} C' = (1+\varepsilon^{\frac{1}{2}})C, & L' = \varepsilon^{-h}L, & \rho' = \frac{1}{2}\rho, \\ \lambda' = 9^{-M'}\lambda, & M' = M^{\tilde{s}M^3}, & N' = \varepsilon^{-a}, \\ \kappa' = \varepsilon^h, & B' = L, & \vartheta' = \varepsilon^g L, \end{array}$$

and

$$2\lambda' \le |\mathcal{I}'| \le \varepsilon^g,$$

$$|Z'_{mn}|_{C^k} < \varepsilon^{\frac{1}{2}\varepsilon^{-a/2}}e^{-\rho'|m-n|}L'^k.$$

In addition,

$$|E(x_* + m\tilde{\alpha}) - E(x_* + n\tilde{\alpha})| < M'\frac{L}{\rho}\varepsilon^g, \quad \forall m \in \Lambda'(n),$$

$$Q'(x)(\mathbb{R}^{\Lambda'(n)}) \subset \sum_{m\in\Lambda'(n)} Q(x)(\mathbb{R}^{\Lambda(m)}), \quad \forall x \in \mathcal{I}',$$

$$D' \text{ is in normal form with the same parameters also on } x_* + \frac{1}{2}(\mathcal{I}' - x_*).$$

Finally, if $M \geq 2\tilde{\tau}$ *then the closure of the sets*

$$\{x_* + m\tilde{\alpha} : |E(x_* + m\tilde{\alpha}) - E(x_* + (m+n)\tilde{\alpha})| < 2M'\frac{L}{\rho}\varepsilon^g\}, \quad \forall\, 4(1/\lambda)^{\tilde{\tau}+2} < |n| < M'N',$$

$$\{x_* + m\tilde{\alpha} : |E(x_* + m\tilde{\alpha}) - E(x_* + (m+n)\tilde{\alpha})| < 2\varepsilon^{\frac{1}{8}}\}, \quad \forall\, M'N' < |n| < 4(1/\lambda')^{\tilde{\tau}+2}$$

are unions of, respectively, at most $\varepsilon^{-\frac{g}{5\tilde{s}M^2}}$ *and* $\varepsilon^{-M^4 g}$ *many components, each component being of length, respectively, at most* $\varepsilon^{\frac{g}{4\tilde{s}M^2}}$ *and* $\varepsilon^{2M^4 g}$.

For the detail of proof, which contains the construction of new blocks Λ'_i, i.e., the new equivalence relation on $\mathbb{Z}$, and the new orthogonal transformation Q', see Section 5 of Reference [11].

Recall that (1.5) – (1.8) have defined quantities $\tilde{\gamma}$, $\tilde{\tau}$, C, L, $\tilde{s}$, $\tilde{\xi}$ associated with the Diophantine number $\tilde{\alpha}$ and the nonconstant real-analytic function V. It has been proved by Eliasson in Section 6 of [11] that $D_0 = \mathrm{diag}\{V(x + n\tilde{\alpha})\}_{n\in\mathbb{Z}}$ is in normal form with $C_0 = C$, $L_0 = L$, any

$$M_0 \geq \max\left\{2^{\tilde{s}+4}C\frac{L^{\tilde{s}+1}((\tilde{s}+1)!)^2}{\tilde{\xi}},\, 2\tilde{\tau},\, 8\right\}, \quad N_0 \geq 1, \quad \rho = N_0^{-1},$$

and other suitable parameters κ_0, B_0, λ_0, ϑ_0. With $\varepsilon_0 = \epsilon^{\frac{1}{4}}$, $Z_0 = \epsilon\Delta$ satisfies

$$|(Z_0)_{mn}|_{C^k} < \varepsilon_0 e^{-\rho_0|m-n|} L_0^k.$$

For $\nu = 0, 1, 2, \cdots$, let $M_{\nu+1} = M_\nu^{\tilde{s}M_\nu^3}$, and

$$a_\nu = \frac{1}{\tilde{\tau}}\left(\frac{1}{M_\nu}\right)^{3\tilde{s}M_\nu^3}, \quad g_\nu = 20\tilde{s}\tilde{\tau}M_\nu^4 a_\nu, \quad h_\nu = \frac{1}{5\tilde{s}}\left(\frac{1}{M_\nu}\right)^{2\tilde{s}M_\nu^3}.$$

The other sequences can be defined as

$$\begin{array}{lll} \varepsilon_{\nu+1} = \varepsilon_\nu^{\frac{1}{2}\varepsilon_\nu^{-a_\nu/2}}, & C_{\nu+1} = (1 + \varepsilon_\nu^{1/2})C_\nu, & L_{\nu+1} = \varepsilon_\nu^{-h_\nu} L_\nu, \\ N_{\nu+1} = \varepsilon_\nu^{-a_\nu}, & \rho_{\nu+1} = \varepsilon_\nu^{a_\nu}, & \kappa_{\nu+1} = \varepsilon_\nu^{h_\nu}, \\ B_{\nu+1} = L_\nu, & \lambda_{\nu+1} = 9^{-M_\nu}\varepsilon_\nu^{g_\nu}, & \vartheta_{\nu+1} = \varepsilon_\nu^{g_\nu} L_\nu. \end{array}$$

The inequality (A.1), about parameters at the ν^{th} step, has been verified in Section 6 of [11], so we can apply Lemma A.1 iteratively. For each $\nu \geq 0$, there is an orthogonal matrix $\tilde{U}_\nu$ satisfying the shift condition, such that

$$|(\tilde{U}_\nu - I_{\mathbb{Z}})_{mn}|_{C^k} < \varepsilon_\nu^{\frac{1}{2}} e^{-\frac{\rho_\nu}{2}|m-n|} L_{\nu+1}^k$$

and

$$(\tilde{U}_0 \cdots \tilde{U}_\nu)^*(D_0 + Z_0)(\tilde{U}_0 \cdots \tilde{U}_\nu) = D_{\nu+1} + Z_{\nu+1},$$

where $D_{\nu+1}$ is in normal form with parameters $C_{\nu+1}$, $L_{\nu+1}$, $\rho_{\nu+1}$, $M_{\nu+1}$, $N_{\nu+1}$, $\kappa_{\nu+1}$, $B_{\nu+1}$, $\vartheta_{\nu+1}$, $\lambda_{\nu+1}$, and $Z_{\nu+1}$ is a symmetric matrix, smoothly parametrized on $\mathbb{R}/\mathbb{Z}$, satisfying the shift condition and

$$|(Z_{\nu+1})_{mn}|_{C^k} \leq \varepsilon_{\nu+1} e^{-\rho_{\nu+1}|m-n|} L_{\nu+1}^k.$$

Hence, in the operator norm $\|\cdot\|_{C^k}$,

$$\tilde{U}_0 \cdots \tilde{U}_\nu \to U, \quad Z_\nu \to 0, \quad D_\nu \to D_\infty.$$

Let $U_{\nu+1} = \tilde{U}_0 \cdots \tilde{U}_\nu$, by a simple calculation, we have

$$|(U_{\nu+1} - I_{\mathbb{Z}})_{mn}|_{C^k} < \varepsilon_0^{\frac{1}{2}} e^{-\frac{\rho_\nu}{2}|m-n|} L_{\nu+1}^k.$$

Clearly there is a uniform limit $E^\nu(x) \to E^\infty(x)$ which describes the spectrum of $D_\infty(x)$–it is the closure of the image of E^∞. Consider now the closure S_ν of the set of all x such that

$$|E_\infty(x) - E_\infty(x+n\tilde{\alpha})| < \frac{3}{2} M_{\nu+1} \frac{L_\nu}{\rho_\nu} \varepsilon_\nu^{g_\nu} \quad \text{for some } \ 4(1/\lambda_\nu)^{\tilde{\tau}+2} < |n| < M_{\nu+1} N_{\nu+1}$$

or

$$|E_\infty(x) - E_\infty(x+n\tilde{\alpha})| < \frac{3}{2} \varepsilon_\nu^{\frac{1}{8}} \quad \text{for some } \ M_{\nu+1} N_{\nu+1} < |n| < 4(1/\lambda_{\nu+1})^{\tilde{\tau}+2}.$$

According to the final statement of Lemma A.1, this set is of measure less than $c\varepsilon_\nu^{g_\nu/20\tilde{s}M_\nu^2}$. By Borel-Cantelli Lemma, we conclude that there is a full-measure subset $\tilde{\mathcal{X}}$ of $\mathbb{R}/\mathbb{Z}$ such that for any $x \in \tilde{\mathcal{X}}$, each $x + n\tilde{\alpha}$ will belong to only finitely many S_ν's. Choose $x = x_*$ of this sort, i.e., for all $n \in \mathbb{Z}$ there is a $\nu_0(n)$ such that $x_* + n\tilde{\alpha} \notin S_\nu$ for $\nu \geq \nu_0(n)$. Hence for such ν's,

$$|E^\nu(x_* + n\tilde{\alpha}) - E^\nu(x_* + n\tilde{\alpha} + m\tilde{\alpha})| \geq 2M_{\nu+1} \frac{L_\nu}{\rho_\nu} \varepsilon_\nu^{g_\nu}, \quad \forall\, 4(1/\lambda_\nu)^{\tilde{\tau}+2} < |m| < M_{\nu+1} N_{\nu+1},$$

$$|E^\nu(x_* + n\tilde{\alpha}) - E^\nu(x_* + n\tilde{\alpha} + m\tilde{\alpha})| \geq 2\varepsilon_\nu^{\frac{1}{8}}, \quad \forall\, M_{\nu+1} N_{\nu+1} < |m| < 4(1/\lambda_{\nu+1})^{\tilde{\tau}+2}.$$

This implies that $\Lambda^\nu(n) \subset [n - 4(1/\lambda_{\nu_0(n)})^{\tilde{\tau}+2},\ n + 4(1/\lambda_{\nu_0(n)})^{\tilde{\tau}+2}]$ for $\nu \geq \nu_0(n)$. The blocks $\Lambda^\nu(n)$ therefore become eventually stationary:

$$\Lambda^{\nu+1}(n) = \Lambda^\nu(n), \quad \forall \nu \geq \nu_0(n).$$

A.2 Hamiltonian vector field and Poisson bracket

For $d, \rho, r, s > 0$, let F, G be two real-analytic functions on $\mathcal{D} = \mathcal{D}_{d,\rho}(r,s)$, both of which C_W^1 depend on the parameter $\xi \in \mathcal{O}$.

Lemma A.2 *The norm $\|\cdot\|_{\mathcal{D},\mathcal{O}}$ has the Banach algebraic property, i.e.,*

$$\|FG\|_{\mathcal{D},\mathcal{O}} \leq \|F\|_{\mathcal{D},\mathcal{O}} \|G\|_{\mathcal{D},\mathcal{O}}.$$

Proof: Since $(FG)_{kl\alpha\beta} = \sum_{\substack{\check{k}+\hat{k}=k,\ \check{l}+\hat{l}=l \\ \check{\alpha}+\hat{\alpha}=\alpha,\ \check{\beta}+\hat{\beta}=\beta}} F_{\check{k}\check{l}\check{\alpha}\check{\beta}} G_{\hat{k}\hat{l}\hat{\alpha}\hat{\beta}}$, we have that

$$\begin{aligned}
\|FG\|_{\mathcal{D},\mathcal{O}} &= \sup_{\mathcal{D}} \sum_{k,l,\alpha,\beta} |(FG)_{kl\alpha\beta}|_{\mathcal{O}} |q^\alpha||\bar{q}^\beta||I^l|e^{|k||\mathrm{Im}\theta|} \\
&\le \sup_{\mathcal{D}} \sum_{k,l,\alpha,\beta} \sum_{\substack{\check{k}+\hat{k}=k,\ \check{l}+\hat{l}=l \\ \check{\alpha}+\hat{\alpha}=\alpha,\ \check{\beta}+\hat{\beta}=\beta}} |F_{\check{k}\check{l}\check{\alpha}\check{\beta}} G_{\hat{k}\hat{l}\hat{\alpha}\hat{\beta}}|_{\mathcal{O}} |q^\alpha||\bar{q}^\beta||I^l|e^{(|\check{k}|+|\hat{k}|)|\mathrm{Im}\theta|} \\
&\le \|F\|_{\mathcal{D},\mathcal{O}} \|G\|_{\mathcal{D},\mathcal{O}}.
\end{aligned}$$

■

Lemma A.3 (Generalized Cauchy Inequalities) *The various components of the Hamiltonian vector field X_F satisfy: for any $0 < r' < r$, $0 < \rho' < \rho$,*

$$\|\partial_\theta F\|_{\mathcal{D}_{d,\rho}(r',\,s)} \le \frac{c}{r-r'}\|F\|_{\mathcal{D}},$$

$$\|\partial_I F\|_{\mathcal{D}_{d,\rho}(r,\,\frac{s}{2})} \le \frac{c}{s^2}\|F\|_{\mathcal{D}},$$

$$\sup_{\mathcal{D}_{d,\rho}(r,\frac{s}{2})} \sum_{n\in\mathbb{Z}} \left(\|\partial_{q_n}F\|_{\mathcal{O}} + \|\partial_{\bar{q}_n}F\|_{\mathcal{O}}\right) \langle n\rangle^d e^{|n|\rho'} \le \frac{c}{s(\rho-\rho')}\|F\|_{\mathcal{D}}.$$

Proof: We only prove the third inequality, with others shown analogously. Given $\omega \in \ell^1_{d,\rho}(\mathbb{Z}) \setminus \{0\}$, $f(t) = F(\cdot,\cdot,q+t\omega,\cdot)$ is an analytic function on the the complex disc $\{z \in \mathbb{C} : |z| < \frac{s}{\|\omega\|_{d,\rho}}\}$. Hence

$$|f'(0)| = \left|\sum_{n\in\mathbb{Z}} \omega_n \cdot \partial_{q_n} F\right| \le \frac{c}{s}\|F\|_{\mathcal{D}} \cdot \|\omega\|_{d,\rho},$$

by the usual Cauchy inequality. As a linear operator on $\ell^1_{d,\rho}(\mathbb{Z})$, $\partial_q F$ satisfies

$$\|\partial_q F\|_{\mathrm{op}} := \sup_{\omega\neq 0} \frac{|\sum_{n\in\mathbb{Z}_1} \omega_n \cdot \partial_{q_n} F|}{\|\omega\|_{d,\rho}} \le \frac{c}{s}\|F\|_{\mathcal{D}}.$$

Let $\|\omega\|_{d,\rho} = \frac{s}{2}$, then

$$|\partial_{q_n} F| \le \sup_{\|\omega\|_{d,\rho}=\frac{s}{2}} \frac{|\partial_{q_n}F|\cdot|\omega_n|}{\|\omega\|_{d,\rho}} \le \frac{\|\partial_q F\|_{\mathrm{op}}|\omega_n|}{\frac{s}{2}} \le \frac{c}{s}\|F\|_{\mathcal{D}}\langle n\rangle^{-d}e^{-|n|\rho}.$$

Hence, for any $0 < \rho' < \rho$,

$$\sum_{n\in\mathbb{Z}} |\partial_{q_n}F|\langle n\rangle^d e^{|n|\rho'} \le \sum_{n\in\mathbb{Z}_1} \frac{c}{s}\|F\|_{\mathcal{D}} e^{-|n|(\rho-\rho')} \le \frac{c}{s(\rho-\rho')}\|F\|_{\mathcal{D}}.$$

With $\tilde{F} = \sum_{k,l,\alpha,\beta}(\partial_\xi F_{kl\alpha\beta}) I^l e^{\mathrm{i}\langle k,\theta\rangle} q^\alpha \bar{q}^\beta$, it can be proved similarly that

$$\sum_{n\in\mathbb{Z}} |\partial_{q_n}\tilde{F}|e^{|n|\rho'} \le \frac{c}{s(\rho-\rho')}\|F\|_{\mathcal{D}}.$$

Since in the process above, $\xi \in \mathcal{O}$ and $(\theta, I, q, \bar{q}) \in \mathcal{D}_{d,\rho}(r, \frac{s}{2})$ are arbitrarily chosen, this inequality is proved. ■

Remark A.2 *These inequalities can be seen as a generalization of the standard Cauchy estimates, which is similar to Lemma A.3 in [30].*

Let $\{\cdot,\cdot\}$ denote the Poisson bracket of smooth functions, i.e.,

$$\{F,G\} = \langle \partial_I F, \partial_\theta G\rangle - \langle \partial_\theta F, \partial_I G\rangle + \mathrm{i}\sum_{n\in\mathbb{Z}_1}\left(\partial_{q_n}F\cdot\partial_{\bar{q}_n}G - \partial_{\bar{q}_n}F\cdot\partial_{q_n}G\right).$$

Lemma A.4 *If $\|X_F\|_{\mathcal{D}} < \varepsilon'$, $\|X_G\|_{\mathcal{D}} < \varepsilon''$, then*

$$\|X_{\{F,G\}}\|_{\mathcal{D}_{d,\rho}(r-\sigma,\eta s)} < c\sigma^{-1}\eta^{-2}\varepsilon'\varepsilon'',$$

for any $0<\sigma<r$ and $0<\eta\ll 1$.

The proof is similar to that of Lemma 7.3 in [14].

Remark A.3 *For more information about the norm $\|\cdot\|_{\mathcal{D},\mathcal{O}}$, see references [3, 5, 26].*

Lemma A.5 *If both of F and G have gauge invariance, then $\{F,G\}$ has gauge invariance.*

Proof: F and G can be written as

$$F = \sum_{k,\alpha,\beta} F_{k\alpha\beta}(I;\xi)e^{\mathrm{i}\langle k,\theta\rangle}q^\alpha\bar{q}^\beta, \quad G = \sum_{k,\alpha,\beta} G_{k\alpha\beta}(I;\xi)e^{\mathrm{i}\langle k,\theta\rangle}q^\alpha\bar{q}^\beta,$$

with $F_{k\alpha\beta} = G_{k\alpha\beta} \equiv 0$ if $\sum_{j=1}^b k_i + |\alpha| - |\beta| \neq 0$. By a simple calculation, we have

$$\begin{aligned}\{F,G\}_{k\alpha\beta} &= \mathrm{i}\sum_{\substack{\check{k}+\hat{k}=k\\ \check{\alpha}+\hat{\alpha}=\alpha\\ \check{\beta}+\hat{\beta}=\beta}}\Big(\langle\partial_I F_{\check{k}\check{\alpha}\check{\beta}},\hat{k}\rangle G_{\hat{k}\hat{\alpha}\hat{\beta}} - \langle\check{k},\partial_I G_{\hat{k}\hat{\alpha}\hat{\beta}}\rangle F_{\check{k}\check{\alpha}\check{\beta}}\Big) &\text{(A.2)}\\ &\quad + \mathrm{i}\sum_{\substack{\check{k}+\hat{k}=k\\ \check{\alpha}+\hat{\alpha}=\alpha\\ \check{\beta}+\hat{\beta}=\beta}}\sum_{m\in\mathbb{Z}}\Big(F_{\check{k}(\check{\alpha}+e_m)\check{\beta}}G_{\hat{k}\hat{\alpha}(\hat{\beta}+e_m)} - F_{\check{k}\check{\alpha}(\check{\beta}+e_m)}G_{\hat{k}(\hat{\alpha}+e_m)\hat{\beta}}\Big). &\text{(A.3)}\end{aligned}$$

Assume $\sum_{j=1}^b k_i + |\alpha| - |\beta| \neq 0$. Then, in the summation above, it is impossible that

$$\sum_{j=1}^b \check{k}_j + |\check{\alpha}| - |\check{\beta}| = \sum_{j=1}^b \hat{k}_j + |\hat{\alpha}| - |\hat{\beta}| = 0,$$

or

$$\sum_{j=1}^b \check{k}_j + |\check{\alpha}+e_m| - |\check{\beta}| = \sum_{j=1}^b \hat{k}_j + |\hat{\alpha}| - |\hat{\beta}+e_m| = 0,$$

$$\sum_{j=1}^b \check{k}_j + |\check{\alpha}| - |\check{\beta}+e_m| = \sum_{j=1}^b \hat{k}_j + |\hat{\alpha}+e_m| - |\hat{\beta}| = 0.$$

This means, in (A.2) and (A.3), each term $\equiv 0$. Thus Lemma A.5 is obtained. ∎

Lemma A.6 *If there exists $n_* \in \mathbb{Z}$ such that*

$$\partial_{q_{n_*}} F = \partial_{\bar{q}_{n_*}} F = \partial_{q_{n_*}} G = \partial_{\bar{q}_{n_*}} G \equiv 0,$$

then $\partial_{q_{n_*}}\{F, G\} = \partial_{\bar{q}_{n_*}}\{F, G\} \equiv 0$.

Proof: Since

$$\begin{aligned}
\partial_{q_{n_*}}\{F,G\} &= \partial_{q_{n_*}}\left(\langle \partial_I F,\, \partial_\theta G\rangle - \langle \partial_\theta F,\, \partial_I G\rangle + \mathrm{i}\sum_{m\in\mathbb{Z}}\left(\partial_{q_m} F\cdot \partial_{\bar{q}_m} G - \partial_{\bar{q}_m} F\cdot \partial_{q_m} G\right)\right) \\
&= \langle \partial_I(\partial_{q_{n_*}} F),\, \partial_\theta(\partial_{q_{n_*}} G)\rangle - \langle \partial_\theta(\partial_{q_{n_*}} F),\, \partial_I(\partial_{q_{n_*}} G)\rangle \\
&\quad + \mathrm{i}\sum_{m\in\mathbb{Z}}\left(\partial_{q_m}(\partial_{q_{n_*}} F)\cdot \partial_{\bar{q}_m}(\partial_{q_{n_*}} G) - \partial_{\bar{q}_m}(\partial_{q_{n_*}} F)\cdot \partial_{q_m}(\partial_{q_{n_*}} G)\right) \\
&\equiv 0
\end{aligned}$$

and similarly, $\partial_{\bar{q}_{n_*}}\{F,G\} \equiv 0$, this lemma is proved. ∎

Acknowledgements. We would like to thank L.H.Eliasson for many fruitful discussions about this work. We also thank the anonymous referees for valuable comments and suggestions on the manuscript. Z. Zhao wishes to thank COLLEGE de FRANCE for supporting his related work in Institut de Mathematique de Jussieu.

This work is partially supported by NNSF of China (Grant 11031003 and Grant 11271180), and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Anderson, P.: Absence of diffusion in certain random lattices. Phys. Rev. **109**, 1492(1958).

[2] Aubry, S., André, G.: Analyticity breaking and Anderson localization in incommensurate lattices. Ann. Israel. Phys. Soc. **3**, 33(1980).

[3] Bambusi, D., Grébert, B.: Birkhoff normal form for partial differential equations with tame modulus. Duke Math. J. **135**, no. 3, 507–567(2006).

[4] Bellissard, J., Lima, R., Scoppola, E.: Localization in ν-dimensional incommensurate structures. Commun. Math. Phys. **88**, 465–477(1983).

[5] Berti, M., Biasco, L., Procesi, M.: KAM theory for the Hamiltonian derivative wave equation. Annales Scientifiques de l'ENS **46**, fasc. 2(2013).

[6] Bourgain, J., Goldstein, M.: On nonperturbative localization with quasi-periodic potential. Ann. Math. **152**, 835–879(2000).

[7] Bourgain, J., Wang, W.-M.: Quasi-periodic solutions of nonlinear random Schrödinger equations. J. Eur. Math. Soc. **10**, 1–45(2008).

[8] Bourgain, J., Wang, W.-M.: Diffusion bound for a nonlinear Schrödinger equation. *Mathematical Aspect of Nonlinear Dispersive Equations. Ann. of Math. Stud.*, 21–42. Princeton University Press, Princeton(2007).

[9] Chulaevsky, V. A., Dinaburg, E. I.: Methods of KAM-theory for long-range quasiperiodic operators on $\mathbb{Z}^\nu$. Pure point spectrum. Commun. Math. Phys. **153**, 559–577(1993).

[10] Craig, W., Wayne, C. E.: Newton's method and periodic solutions of nonlinear wave equations. Commun. Pure. Appl. Math. **46**, 1409–1498(1993).

[11] Eliasson, L. H.: Discrete one-dimensional quasi-periodic Schrödinger operators with pure point spectrum. Acta. Math. **179**, 153–196(1997).

[12] Fröhlich, J., Spencer, T., Wayne, C. E.: Localization in disordered, nonlinear dynamical systems. J. Statist. Phys. **42**, 247–274(1986).

[13] Fröhlich, J., Spencer, T., Wittwer, P.: Localization for a class of one dimensional quasi-periodic Schrödinger operators. Commun. Math. Phys. **132**, 5–25(1990).

[14] Geng, J., You, J.: A KAM theorem for one dimensional Schrödinger equation with periodic boundary conditions. J. Diff. Eqs. **209**, 1–56(2005).

[15] Geng, J., Zhao, Z.: Quasi-periodic solutions for one-dimensional discrete nonlinear Schrödinger equations with tangent potential. Preprint, 2011.

[16] Germinet, F.: Dynamical localization II with an application to the almost Mathieu operator. J. Stat. Phys. **95**, no. 1-2, 273–286(1999).

[17] Germinet, F., Jitomirskaya, S. Ya.: Strong dynamical localization for the almost Mathieu model. Rev. Math. Phys. **13**, 755–765(2001).

[18] Gross, E. P.: Structure of a quantized vortex in boson systems. Nuovo Cimento **20**, 454–477(1961).

[19] Harper, P. G.: Single band motion of conduction electrons in a uniform magnetic field. Proc. Phys. Soc. London Sect. A **68**, 874–878(1955).

[20] Hiramoto, H., Abe, S.: Dynamics of an Electron in Quasiperiodic Systems. II. Harper's Model. J. Phys. Soc. Jpn. **57** 1365–1371(1988).

[21] Jitomirskaya, S. Ya.: Anderson localization for the almost Mathieu equation: a nonperturbative proof. Commun. Math. Phys. **165**, 49–57(1994).

[22] Jitomirskaya, S. Ya.: Metal-insulator transition for the almost Mathieu operator. Annal. Math., **150**, 1159–1175(1999).

[23] Klein, S.: Anderson localization for the discrete one-dimensional quasi-periodic Schrödinger operator with potential defined by a Gevrey-class function. J. Funct. Anal. **218**, 255–292(2005).

[24] Kuksin, S. B.: *Nearly integrable infinite dimensional Hamiltonian systems.* Lecture Notes in Mathematics, 1556. Berlin: Springer, 1993.

[25] Larcher, M., Dalfovo, F., Modugno, M.: Effects of interaction on the diffusion of atomic matter waves in one-dimensional quasiperiodic potentials. Phys. Rev. A **80**, 053606(2009).

[26] Nikolenko, N. V.: The method of Poincaré normal forms in problems of integrability of equations of evolution type. Russian Math. Surveys **41**, no. 5, 63–114(1986).

[27] Pitaevskii, L. P.: Vortex lines in an imperfect Bose gas. Sov. Phys. JETP **13**, 451(1961).

[28] Pöschel, J.: Small divisors with spatial structrue in infinite dimensional Hamiltonian systems. Comment. Math. Phys. **127**, 351–393(1990).

[29] Pöschel, J.: Quasi-periodic solutions for a nonlinear wave equation. Comment. Math. Helvetici **71**, 269–296(1996).

[30] Pöschel, J.: A KAM Theorem for some nonlinear partial differential equations. Ann. Sc. Norm. Sup. Pisa Cl. Sci. **23**, 119–148(1996).

[31] Roati, G., D'Errico, C., Fallani, L., Fattori, M., Fort, C., Zaccanti, M., Modugno, G., Modugno, M., Inguscio, M,: Anderson localization of a non-interacting Bose-Einstein condensate. Nature **453**, 895–898(2008).

[32] Sinai, Ya. G.: Anderson localization for the one-dimensional difference Schrödinger operator with a quasi-periodic potential. J. Statist. Phys. **46**, 861–909(1987).

[33] Trombettoni, A., Smerzi, A.:Discrete solitons and breathers with dilute Bose-Einstein condensates. Phys. Rev. Lett. **86**, 2353–2356(2001).

[34] Vittot, M., Bellissard, J.: Invariant tori for an infinite lattice of coupled classical rotators. CPT-Marseille, 1985.

[35] Wang, W.-M., Zhang, Z.: Long time Anderson localization for the nonlinear random Schrödinger equation. J. Stat. Phys. **134**, no. 5-6, 953–968(2009).