Long-Time Anderson Localization for the Nonlinear Schrödinger Equation Revisited

Hongzi Cong · Yunfeng Shi · Zhifei Zhang

Received: 7 June 2020 / Accepted: 5 December 2020 / Published online: 7 January 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract
In this paper, we confirm the conjecture of Wang and Zhang (J Stat Phys 134 (5-6):953–968, 2009) in a long time scale, i.e., the displacement of the wavefront for 1D nonlinear random Schrödinger equation is of logarithmic order in time $|t|$.

Keywords Anderson localization · Birkhoff normal form · Nonlinear random Schrödinger equation

1 Introduction

Anderson localization was originally discussed by Anderson [3] in the context of wave propagation of non-interacting quantum particles through random disordered media. Since this seminal work, a great deal of attention has been paid to this topic both in physics and mathematics community. The Anderson model is a discrete linear Schrödinger operator defined on $l^2(\mathbb{Z}^d)$

$$H_0 = -\epsilon_1 \Delta + \lambda v_n(\omega) \delta_{nn'},$$

(1.1)

where $\Delta$ is the discrete Laplacian: $(\Delta q)_n = \sum_{|e|=1} q_{n+e} \cdot q_{n'}$ (with $|e|=1$) and $(v_n(\omega))_{n \in \mathbb{Z}^d}$ is a family of identical independent distributed (i.i.d.) random variables with uniform distribution on $[0, 1]$ (i.e., $dv_n(\omega) = \chi_{[0,1]}(\omega_n)d\omega_n$). The constant $\epsilon_1 \geq 0$ is the coupling for...
describing the strength of random disorder. We say that $H_0$ has Anderson localization (AL) if its spectrum is pure point with exponentially decaying eigenfunctions. In many cases, we are interested in the dynamics of the time dependent (linear) Schrödinger equation associated with $H_0$

$$i\dot{q} = H_0 q,$$  \hspace{1cm} (1.2)

where $q \in \ell^2(\mathbb{Z}^d)$. The standard spectral theorem of self-adjoint operators deduces that (1.2) has a unique global solution $q(t) = e^{-i t H_0} q(0)$ for each initial data $q(0) \in \ell^2(\mathbb{Z}^d)$. The evolution operator $e^{-i t H_0}$ is unitary for each $t \in \mathbb{R}$ and thus preserves the $\ell^2$-norm. If we want to know more precise information about the wave packet propagation, we can introduce the concept of dynamical localization (DL) for

$$\sup_{t \in \mathbb{R}} \sum_{n \in \mathbb{Z}^d} (1 + |n|)^{2\alpha} |q_n(t)|^2 < \infty,$$  \hspace{1cm} (1.3)

where $q(t) = e^{-i t H_0} q(0)$ and $|n| = \max_{1 \leq j \leq d} |n_j|$. The DL implies that the particle is concentrated near the origin uniformly for all time.

The first mathematical rigorous proof of localization for random operators was due to Goldsheid–Molchanov–Pastur [18] for 1D continuous random Schrödinger operators. In high dimensions, Fröhlich–Spencer [16] proved, either at high disorder (i.e., $\epsilon_1 \ll 1$) or low energy, the absence of diffusion for Anderson model by developing the celebrated multiscale analysis (MSA) method. Based on MSA of [10,15,16,24] finally obtained the Anderson localization at either high disorder or low energy. An alternative method for the proof of localization for random operators, known as the fractional moment method (FMM), was developed by Aizenman–Molchanov [2]. Remarkably, by employing FMM, Aizenman proved the first DL for Anderson model [1].

When a nonlinear perturbation is added in (1.2), we are led to the study of the so called nonlinear Schrödinger equations with a random potential. In this paper, we focus on the following 1D nonlinear Schrödinger equation (NLSE)

$$i\dot{q}_j = \epsilon_1 (q_{j-1} + q_{j+1}) + v_j(\omega) q_j + \epsilon_2 |q_j|^2 q_j,$$  \hspace{1cm} (1.4)

and in particular the solution $q(t)$ of (1.4) with an initial state $q(0) \in \ell^2(\mathbb{Z})$ as $t \to \infty$. The NLSE also has important applications in a variety of physical systems, especially the Bose–Einstein Condensation [9] (we refer to [14] for an excellent review on NLSE). Since in nonlinear case the spectral theorem becomes invalid, the study of AL for a NLSE seems vacuous. However, in linear case the famous RAGE Theorem (see [4,20]) claims that $H_0$ has pure point spectrum if and only if, for any $q(0) \in \ell^2(\mathbb{Z})$,

$$\lim_{N \to \infty} \sup_{t \in \mathbb{R}} \sum_{|j| > N} |q_j(t)|^2 = 0,$$  \hspace{1cm} (1.5)

where $q(t) = e^{-i t H_0} q(0)$. Thus, it is natural to define AL for a NLSE via (1.5) by noting that (1.4) is globally well-posed for any initial data belongs to $\ell^2(\mathbb{Z})$.

The numerical results found by Pikovsky–Shepelyansky [21] and by Flach and coworkers [6,13,22,23] suggested that an initially localized wavepacket spreads eventually in the presence of nonlinearity. Particularly, in the weak nonlinearity case (i.e., $0 < \epsilon_2 \ll 1$), it was numerically established in [13] that AL occurs up to some time scale $T_{\epsilon_2} > 0$ which increases with decreasing $\epsilon_2$. Moreover, for $t > T_{\epsilon_2}$, the wavepacket starts to spread sub-diffusively. However, all rigorous theories predict that the spreading cannot be faster than logarithmic
in time. This seems due to the fact that numerical calculations for chaotic systems are quite sensitive to numerical errors (see [14] for details).

The first rigorous result towards nonlinear AL for NLSE with i.i.d. random potential was obtained by Fröhlich–Spencer–Wayne [17]: they showed that, with high probability and weak nonlinearity, any sup-exponentially localized initial state always stayed in a full dimensional KAM tori. Their proof is based on an extension of the KAM techniques. Later, if the initial state is polynomially localized, by using Birkhoff normal form method, Benettin–Fröhlich–Giorgilli [5] got that the propagation remains localized in very long-time for some $dD$ lattice nonlinear oscillation equations with i.i.d. Gaussian random potential. Recently, Bourgain–Wang [8] constructed many quasi-periodic solutions for some random NLSE by combining Nash–Moser iteration and the improved MSA. We would also like to mention the works of Yuan [27] and Geng–You–Zhao [19], in which the persistence of quasi-periodic solutions for some 1D discrete nonlinear equations was proved via the KAM type iterations scheme.

The most important result for nonlinear AL with non-localized initial state was due to Wang–Zhang [26]: they proved the first “truly” long-time AL for the 1D NLSE. More precisely, they established that Given $A \geq 2$, $\delta > 0$, let $q(0) \in \ell^2(\mathbb{Z})$ be any initial state satisfying $\sum_{|j| > j_0}|q_j(0)|^2 \leq \delta$. Then there exist $\epsilon = \epsilon(A) > 0$, $C = C(A) > 0$ such that for $0 < \epsilon = \epsilon_1 + \epsilon_2 \leq \epsilon$ and $t \leq \delta C^{-1} \epsilon^{-A}$,

$$\sum_{|j| > j_0 + N} |q_j(t)|^2 \leq 2\delta$$

with probability at least $1 - \exp\left(-\frac{\delta}{N} e^{-2N c(A)^{-1}}\right)$ and $N = N(A) \geq A^2$. In this theorem, they required actually both high disorder and weak nonlinearity. The proof depends on some type of Birkhoff normal form borrowed from Bourgain–Wang [7]. Remarkably, Fishman–Krivolapov–Soffer [11] obtained the long-time exponentially DL (i.e., with $(1 + |n|)^{2\alpha}$ being replaced by the exponential bound in (1.3)) of time $t \leq \epsilon^{-2}$ under just weak nonlinearity assumption. Their proof differs from that of Wang–Zhang and is based on perturbation theory combined with FMM of Aizenman–Molchanov [2]. Subsequently, some results of [11] have been improved to time of order $\epsilon^{-A}$ for any $A \geq 2$ [12] by the same authors, but the proof is partly rigorous: in some parts it relies on conjectures that they tested numerically.

Wang–Zhang’s result mentioned as above indicates that if $\epsilon \leq \epsilon(A)$, AL holds of time scale $T_{\epsilon} \sim \epsilon^{-A}$, and as a result the wavefront $N$ depends on time in the following way

$$N \sim (\ln T_{\epsilon})^{2+}.$$ (1.6)

In addition, it was proven in [25] that the growth of Sobolev norms is at most logarithmic in $t$. These enable them to raise the conjecture:

**Conjecture 1.1** [26] As $t \to \infty$, the displacement of the wavefront $N$ is of order $t^{0^+}$ (possibly logarithmic).

The main motivation of the present paper comes from this conjecture. In fact, we prove the following main result.

**Theorem 1.2** Given $\delta > 0$, for all initial datum $q(0) \in \ell^2(\mathbb{Z})$, let $j_0 \in \mathbb{N}$ be such that

$$\sum_{|j| > j_0} |q_j(0)|^2 < \delta.$$
Fix $0 < \alpha < 1/100$. Then there exists constant $\epsilon = \epsilon(\alpha) > 0$ such that the following holds: for $0 < \epsilon := \epsilon_1 + \epsilon_2 < \epsilon$ and for all

$$|t| \leq \delta \exp \left( \frac{|\ln \epsilon|^2}{200 \ln |\ln \epsilon|} \right)$$

one has

$$\sum_{|j| > j_0 + N} |q_j(t)|^2 < 2\delta$$

with probability at least

$$1 - e^{\alpha/2},$$

where

$$N = \left| \frac{\ln \epsilon}{200 \ln |\ln \epsilon|} \right|^2.$$

**Remark 1.1** • As an easy corollary, one has for $|t| \leq T_\epsilon = \delta \exp \left( \frac{|\ln \epsilon|^2}{200 \ln |\ln \epsilon|} \right)$,

$$N(\epsilon) \sim \ln T_\epsilon.$$

Moreover, $T_\epsilon \to \infty$ in the exponential rate as $\epsilon \to 0$. This confirms Wang-Zhang’s conjecture in a long time scale.

• Our result can’t be derived directly from Wang-Zhang’s by choosing $A \sim \frac{|\ln \epsilon|}{\ln |\ln \epsilon|}$. It is because the perturbation $\epsilon(A)$ in their argument depends sensitively on $A$. In order to improve Wang-Zhang’s polynomial bound to the exponential one, it requires new ideas.

We then outline the proof. The main scheme of our proof is definitely adapted from Bourgain–Wang [7] and Wang–Zhang [26], which uses Birkhoff normal form type transformations to construct barriers centered at some $\pm j_0$, $j_0 > 1$ of width $N$, where the terms responsible for propagation are small enough. However, while our localized time is significantly much longer, our argument can also be viewed as both a clarification and at the same time streamlining of [26]. This is due to several important technical improvements that we add to Wang–Zhang’s scheme:

1. One important highlight is that, we make use of $\ell^1$-norm (with an exponential weight) rather than $\ell^\infty$-norm for the Hamiltonian. This will lead to more clear and effective estimate on some key ingredients, such as the Poisson bracket, symplectic transformations and particularly the small divisors when performing the Birkhoff normal form. In addition, we deal with those elements in a separated fashion, which makes the proof more tractable.

2. Another issue we want to highlight is that we introduce new ideas originated from Benettin–Fröhlich–Giorgilli [5] in our proof. In the iteration scheme, we always assume that both the width $N$ of the barriers and the total iteration steps $M$ are non-negligible as compared with the perturbation $\epsilon$. Then our main result follows from optimal choices of $N, M$ depending on $\epsilon$. To achieve this goal, one needs to take care of all terms in the barriers and thus needs to use the $\ell^1$-norm.

The structure of the paper is as follows. Some important facts on Hamiltonian dynamics, such as the Poisson bracket, symplectic transformation and non-resonant conditions are presented in Sect. 2. The Birkhoff normal form type theorem is proved in Sect. 3. The estimate on the probability when handling the small divisors can be found in Sect. 4. The proof of our main theorem is finished in Sect. 5.
2 Structure of the Transformed Hamiltonian

We recast (1.4) as a Hamiltonian equation

$$i \dot{q}_j = 2 \frac{\partial H}{\partial \bar{q}_j},$$

where

$$H(q, \bar{q}) = \frac{1}{2} \left( \sum_{j \in \mathbb{Z}} v_j |q_j|^2 + \epsilon_1 \sum_{j \in \mathbb{Z}} (\bar{q}_j q_{j+1} + q_j \bar{q}_{j+1}) + \frac{1}{2} \epsilon_2 \sum_{j \in \mathbb{Z}} |q_j|^4 \right). \quad (2.1)$$

As is well-known, the $\ell^2$-norm of the solution $q(t)$ is conserved, i.e.,

$$\sum_{j \in \mathbb{Z}} |q_j(t)|^2 = \sum_{j \in \mathbb{Z}} |q_j(0)|^2 \quad \text{for } \forall \, t \in \mathbb{R}.$$ 

In order to prove the main result, we need to control the time derivative of the truncated sum of higher modes

$$\frac{d}{dt} \sum_{|j| > j_0} |q_j(t)|^2. \quad (2.2)$$

In what follows, we will deal extensively with monomials in $q_j$. Rewrite any monomials in the form

$$\prod_{j \in \mathbb{Z}} q_j^{n_j} \bar{q}_j^{n'_j}. \quad (2.3)$$

Let

$$n = (n_j, n'_j)_{j \in \mathbb{Z}} \in \mathbb{N}^\mathbb{Z} \times \mathbb{N}^\mathbb{Z}.$$ 

We define

$$\text{supp } n = \{ j \in \mathbb{Z} : n_j \neq 0 \text{ or } n'_j \neq 0 \},$$

$$\Delta(n) = \sup_{j, j' \in \text{supp } n} |j - j'|,$$

$$|n| = \sum_{j \in \mathbb{Z}} (n_j + n'_j).$$

If $n_j = n'_j$ for all $j \in \text{supp } n$, then the monomial (2.3) is called resonant. Otherwise it is called non-resonant. Note that non-resonant monomials contribute to the truncated sum in (2.2), where resonant ones do not. We define the (resonant) set as

$$\mathcal{N} = \left\{ n \in \mathbb{N}^\mathbb{Z} \times \mathbb{N}^\mathbb{Z} : n_j = n'_j \text{ for } \forall \, j \right\}. \quad (2.4)$$

Given $j_0$ and $N \in \mathbb{N}$, let

$$A(j_0, N) := [j_0 - N, j_0 + N] \cup [-j_0 - N, -j_0 + N].$$

**Definition 2.1** Given a Hamiltonian

$$H(q, \bar{q}) = \sum_{n \in \mathbb{N}^\mathbb{Z} \times \mathbb{N}^\mathbb{Z}} H(n) \prod_{n \in \text{supp } n} q_j^{n_j} \bar{q}_j^{n'_j},$$
for \( j_0, N \in \mathbb{N} \) and \( r > 2 \), we define

\[
\|H\|_{j_0,N,r} = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2 \atop \text{supp } n \cap A(j_0,N) \neq \emptyset} |H(n)| \cdot |n| \cdot r^{\Delta(n)+|n|-1}
\]  

(2.5)

and

\[
\|H\|_{j_0,N,r}^L = \sup_{j \in \mathbb{Z}} \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2 \atop \text{supp } n \cap A(j_0,N) \neq \emptyset} \left| \partial_{v_j} H(n) \right| \cdot |n| \cdot r^{\Delta(n)+|n|-1},
\]  

(2.6)

where \( v = (v_j)_{j \in \mathbb{Z}} \) is the potential. Define

\[
|||H|||_{j_0,N,r} = \|H\|_{j_0,N,r} + \|H\|_{j_0,N,r}^L.
\]  

Definition 2.2 Given

\[
H(q, \bar{q}) = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} H(n) \prod_{\text{supp } n} q_j^n q_j^{\bar{n}}
\]

and

\[
G(q, \bar{q}) = \sum_{m \in \mathbb{N}^2 \times \mathbb{N}^2} G(m) \prod_{\text{supp } m} q_j^m q_j^{\bar{m}}
\]

the Poisson bracket of \( H \) and \( G \) is defined as

\[
\{H, G\} := i \sum_{n,m \in \mathbb{N}^2 \times \mathbb{N}^2} \sum_{k \in \mathbb{Z}} H(n)G(m)(n_k m'_k - n'_k m_k)q_k^{n_k+m_k-1} q_k^{n'_k+m'_k-1} \left( \prod_{j \neq k} q_j^{n_j+m_j} q_j^{\bar{n}_j+m'_j} \right).
\]

We have the following key estimate.

Proposition 2.3 (Poisson Bracket) For \( j_0, N \in \mathbb{N} \), let \( a \) and \( b \) satisfy

\[ [a, b] \subset [j_0 - N, j_0 + N]. \]

Let

\[
H(q, \bar{q}) = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} H(n) \prod_{\text{supp } n} q_j^n q_j^{\bar{n}}
\]

and

\[
G(q, \bar{q}) = \sum_{m \in \mathbb{N}^2 \times \mathbb{N}^2} G(m) \prod_{\text{supp } m} q_j^m q_j^{\bar{m}}
\]

with

\[ \text{supp } n \subset [-b, -a] \cup [a, b] \text{ for any } n. \]

(2.7)

Then for any \( 0 < \sigma < r/2 \), we have

\[
|||\{H, G\}|||_{j_0,N,r-\sigma} \leq \frac{1}{\sigma} |||H|||_{j_0,N,r} \cdot |||G|||_{j_0,N,r}.
\]

(2.8)
Proof First of all, we write

\[ \{H, G\} = \sum_{l \in \mathbb{N}^2 \times \mathbb{N}^2} \{H, G\}(l) \prod_{\text{supp } l} q^{l_j} \bar{q}^{l_j}, \]

where

\[ \{H, G\}(l) = i \sum_{k \in \mathbb{Z}} \left( \sum_{n, m \in \mathbb{N}^2 \times \mathbb{N}^2} H(n)G(m) \left( n_k m'_k - n'_k m_k \right) \right) \] (2.9)

and the sum is taken as

\[ l_j = n_j + m_j - 1, \quad l'_j = n'_j + m'_j - 1 \text{ for } j = k, \]

\[ l_j = n_j + m_j, \quad l'_j = n'_j + m'_j \text{ for } j \neq k. \]

Secondly, let

\[ \tilde{G} = \sum_{m \in \mathbb{N}^2 \times \mathbb{N}^2, \text{supp } m \cap A(j_0, N) = \emptyset} G(m) \prod_{\text{supp } m} q^{m_j} \bar{q}^{m'_j} \]

and then following (2.7), one has \( \{H, \tilde{G}\} = 0. \) Hence, we always assume that

\[ G = \sum_{m \in \mathbb{N}^2 \times \mathbb{N}^2, \text{supp } m \cap A(j_0, N) \neq \emptyset} G(m) \prod_{\text{supp } m} q^{m_j} \bar{q}^{m'_j}. \] (2.10)

Without loss of generality, we assume that \( H \) and \( G \) are homogeneous polynomials with degrees \( n^* \) and \( m^* \) respectively, i.e.,

\[ H(q, \bar{q}) = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2, |n| = n^*} H(n) \prod_{\text{supp } n} q^{n_j} \bar{q}^{n'_j} \]

and

\[ G = \sum_{m \in \mathbb{N}^2 \times \mathbb{N}^2, |m| = m^*} G(m) \prod_{\text{supp } m} q^{m_j} \bar{q}^{m'_j}. \]

Since \( r > 2 \) and \( 0 < \sigma < r/2, \) one has

\[ 1 < r - \sigma < r. \] (2.11)

In view of (2.11) and

\[ \Delta(l) \leq \Delta(n) + \Delta(m), \]

one has

\[ \sum_{l \in \mathbb{N}^2 \times \mathbb{N}^2} \left| \sum_{k \in \mathbb{Z}} \sum_{n, m \in \mathbb{N}^2 \times \mathbb{N}^2} H(n)G(m) \left( n_k m'_k - n'_k m_k \right) \right| (r - \sigma)^\Delta(l). \]
\[
\begin{align*}
\sum_{n,m \in \mathbb{Z} \times \mathbb{Z}} |H(n)| |G(m)| \sum_{k \in \mathbb{Z}} (n_k m'_k + n'_k m_k) (r - \sigma)^{\Delta(n) + \Delta(m)} \\
\leq \left( \sum_{n \in \mathbb{Z} \times \mathbb{Z}} |H(n)| \cdot |n| \cdot r^{\Delta(n)} \right) \left( \sum_{m \in \mathbb{Z} \times \mathbb{Z}} |G(m)| \cdot |m| \cdot r^{\Delta(m)} \right).
\end{align*}
\]

In view of (2.9), (2.12) and using \(|l| = |n| + |m| - 2\), we have

\[
\begin{align*}
\|\{H, G\}\|_{j_0, N, r - \sigma} & \leq (|n| + |m| - 2) (r - \sigma)^{|n|+|m|-3} \\
& \times \left( \sum_{n \in \mathbb{Z} \times \mathbb{Z}} |H(n)| \cdot |n| \cdot r^{\Delta(n)} \right) \left( \sum_{m \in \mathbb{Z} \times \mathbb{Z}} |G(m)| \cdot |m| \cdot r^{\Delta(m)} \right) \\
& \leq \frac{1}{\sigma} \left( \sum_{n \in \mathbb{Z} \times \mathbb{Z}} |H(n)| \cdot |n| \cdot r^{\Delta(n) + |n| - 1} \right) \left( \sum_{m \in \mathbb{Z} \times \mathbb{Z}} |G(m)| \cdot |m| \cdot r^{\Delta(m) + |m| - 1} \right),
\end{align*}
\]

where the last inequality is based on

\[((|n| + |m| - 2)(r - \sigma)^{|n|+|m|-3} \leq \frac{1}{\sigma} r^{|n|+|m|-2}.\]

Using (2.7), (2.10) and Definition 2.1, we have

\[
\|\{H, G\}\|_{j_0, N, r - \sigma} \leq \frac{1}{\sigma} \|H\|_{j_0, N, r} \cdot \|G\|_{j_0, N, r}.
\]

Finally, recalling

\[
\partial v_j (H(n)G(m)) = \partial v_j H(n) \cdot G(m) + H(n) \cdot \partial v_j G(m)
\]

and following the proof of (2.13), one has

\[
\|\{H, G\}\|_{L^j_{j_0, N, r - \sigma}} \leq \frac{1}{\sigma} \left( \|H\|_{L^j_{j_0, N, r}} \cdot \|G\|_{j_0, N, r} + \|H\|_{j_0, N, r} \cdot \|G\|_{L^j_{j_0, N, r}} \right).
\]

Combining (2.13) and (2.14), we finish the proof of (2.8).

**Proposition 2.4** Let \(H\) and \(G\) be as in Proposition 2.3. Assume further that

\[
\left(\frac{e}{\sigma}\right) \|H\|_{j_0, N, r} \leq \frac{1}{2}.
\]

Then

\[
\|G \circ X^1_H\|_{j_0, N, r - \sigma} \leq 2 \|G\|_{j_0, N, r},
\]

where \(X^1_H\) is the time-1 map generated by the flow of \(H\).

**Proof** First of all, we expand \(G \circ X^1_H\) into the Taylor series

\[
G \circ X^1_H = \sum_{n \geq 0} \frac{1}{n!} G^{(n)},
\]

 Springer
where $G^{(n)} = \{ G^{(n-1)}, H \}$ and $G^{(0)} = G$. We will estimate $\|\| G^{(n)} \|\|_{j_0, N, r - \sigma}$ by repeatedly using of Proposition 2.3:

\[
\|\| G^{(n)} \|\|_{j_0, N, r - \sigma} = \|\| \{ G^{(n-1)}, H \} \|\|_{j_0, N, r - \sigma} \\
\leq \left( \frac{n}{\sigma} \right)^2 \left( \|\| H \|\|_{j_0, N, r} \right)^2 \|\| G^{(n-2)} \|\|_{j_0, N, r - (n-2)\sigma} \\
\ldots \\
\leq \left( \frac{n}{\sigma} \right)^n \left( \|\| H \|\|_{j_0, N, r} \right)^n \|\| G \|\|_{j_0, N, r}.
\]

Then

\[
\frac{1}{n!} \|\| G^{(n)} \|\|_{j_0, N, r - \sigma} \leq \left( \frac{e \|\| H \|\|_{j_0, N, r}}{\sigma} \right)^n \|\| G \|\|_{j_0, N, r}, \tag{2.17}
\]

where we use the inequality $n^n < n!e^n$. Hence combining (2.16) and (2.17), we obtain

\[
\|\| G \circ X_H^1 \|\|_{j_0, N, r - \sigma} \leq \sum_{n \geq 0} \left( \frac{e \|\| H \|\|_{j_0, N, r}}{\sigma} \right)^n \|\| G \|\|_{j_0, N, r} \\
\leq 2 \|\| G \|\|_{j_0, N, r},
\]

where the last inequality is based on (2.15).

\[\square\]

**Remark 2.1** In general, we have

\[
\|\| G \circ X_H^1 - G \|\|_{j_0, N, r - \sigma} \leq \frac{e}{\sigma} \cdot \|\| H \|\|_{j_0, N, r} \cdot \|\| G \|\|_{j_0, N, r}, \tag{2.18}
\]

and

\[
\|\| G \circ X_H^1 - G - \{ G, H \} \|\|_{j_0, N, r - \sigma} \leq \left( \frac{e}{\sigma} \right)^2 \|\| H \|\|_{j_0, N, r} \cdot \|\| G \|\|_{j_0, N, r}. \tag{2.19}
\]

Let

\[
\epsilon = \epsilon_1 + \epsilon_2, \tag{2.20}
\]

and introduce the non-resonant conditions.

**Definition 2.5** *(Non-resonant condition)* Given $\epsilon > 0$, $\alpha \in (0, 1/100)$ and $N \in \mathbb{N}$, we say that the frequency $v = (v_j)_{j \in \mathbb{Z}}$ is $(\epsilon, \alpha, N)$-nonresonant if for any $0 \neq k \in \mathbb{Z}$,

\[
\left| \sum_{j \in \mathbb{Z}} k_j v_j \right| \geq \frac{\epsilon \alpha}{N \Delta^2(k)|k|^{\Delta(k)+1}}. \tag{2.21}
\]

### 3 Analysis and Estimates of the Symplectic Transformations

We now construct the symplectic transformation $\Gamma$ by a finite step induction.
At the first step, i.e., \( s = 1 \) (in view of (2.1))
\[
H_1 = H = \frac{1}{2} \left( \sum_{j \in \mathbb{Z}} v_j |q_j|^2 + \epsilon_1 \sum_{j \in \mathbb{Z}} (\tilde{a}_j q_{j+1} + q_j \tilde{a}_{j+1}) + \frac{1}{2} \epsilon_2 \sum_{j \in \mathbb{Z}} |q_j|^4 \right),
\] (3.1)
which can be rewritten as
\[
H_1 = D_1 + Z_1 + R_1
= \frac{1}{2} \sum_{j \in \mathbb{Z}} v_{1j} |q_j|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} Z_1(n) \prod_{\text{supp } n} |q_{nj}|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_1(n) \prod_{\text{supp } n} q_{nj} q_{nj}',
\]
where
\[
v_{1j} = v_j, \quad Z_1(n) = \frac{\epsilon_2}{4}, \quad R_1(n) = \frac{\epsilon_1}{2}.
\]
From (2.20), we see that
\[
\| H_1 - D_1 \|_{j_0, N, r} \leq 10 N r^3 \epsilon
\] (3.2)
and
\[
\| H_1 - D_1 \|_{L, j_0, N, r} = 0,
\]
which implies
\[
\| |H_1 - D_1|||_{j_0, N, r} \leq 10 N r^3 \epsilon.
\]

3.1 One Step of Birkhoff Normal Form

Let
\[
N_s = N - 20(s - 1), \quad s \geq 1.
\] (3.3)

Lemma 3.1 Let \( v_1 = (v_{1j})_{j \in \mathbb{Z}} \) satisfy the \((\epsilon, \alpha, N)\)-nonresonant conditions (2.21). Assume \( 0 < \sigma < r/2 \) and
\[
\frac{2^6 \epsilon}{\sigma} \cdot 10 N^3 r^3 \epsilon^{1-2\alpha} \leq \frac{1}{2}.
\] (3.4)
Then there exists a change of variables \( \Gamma_1 := X_{F_1}^1 \) such that
\[
H_2 = H_1 \circ X_{F_1}^1 = D_2 + Z_2 + R_2
= \frac{1}{2} \sum_{j \in \mathbb{Z}} v_{2j} |q_j|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} Z_2(n) \prod_{\text{supp } n} |q_{nj}|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_2(n) \prod_{\text{supp } n} q_{nj} q_{nj}'.
\]
Moreover, one has
\[
| |F_1|||_{j_0, N, r} \leq 2^6 \cdot 10 N^3 r^3 \epsilon^{1-2\alpha},
\] (3.5)
\[
| |Z_2|||_{j_0, N, r-\sigma} \leq 10 N r^3 \epsilon \left( \sum_{i=0}^{1} 2^{-i} \right),
\] (3.6)
\[ |||R_2|||_{j_0,N,r - \sigma} \leq 10N r^3 \epsilon \left( \sum_{i=0}^{1} 2^{-i} \right), \quad (3.7) \]

and

\[ |||R_2|||_{j_0,N,r - \sigma} \leq 10N r^3 \epsilon \left( \frac{26^6 \epsilon}{\sigma} \cdot 10N^2 r^3 \epsilon^{1-2\alpha} \right), \quad (3.8) \]

where

\[ \mathcal{R}_2 = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_2(n) \prod_{\text{supp } n \cap A(\bar{j}_0, N_3) \neq \emptyset} q_j^{n_j} q_j^{n_j'}, \quad (3.9) \]

Furthermore, for any \( A \geq 3 \) the following estimate holds

\[ \left| \sum_{\Delta(n) + |n| = A} (|Z_2(n)| + |R_2(n)|) \prod_{\text{supp } n} q_j^{n_j} q_j^{n_j'} \right|_{j_0,N,r - \sigma} \leq 10N r^3 \epsilon \left( \frac{26^6 \epsilon}{\sigma} \cdot 10N^2 r^3 \epsilon^{1-2\alpha} \right)^{A-3}. \quad (3.10) \]

**Proof** By the Birkhoff normal form theory, one knows that \( F_1 \) satisfies the homological equation

\[ L_{v_1} F_1 = \mathcal{R}_1, \quad (3.11) \]

where the *Lie derivative* operator is defined by

\[ L_{v_1} : H \mapsto L_{v_1} H := i \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} \left( \sum_{j \in \mathbb{Z}} (n_j - n_j')v_{1j} \right) H(n) \prod_{\text{supp } n} q_j^{n_j} q_j^{n_j'} \]

and

\[ \mathcal{R}_1 = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_1(n) \prod_{\text{supp } n \cap A(\bar{j}_0, N_2) \neq \emptyset} q_j^{n_j} q_j^{n_j'}. \]

Unless \( n \in \mathcal{N} \) (see (2.4)), one has

\[ F_1(n) = \frac{R_1(n)}{\sum_{j \in \mathbb{Z}} (n_j - n_j')v_{1j}}. \]

Note that frequency \( v_1 \) satisfies the nonresonant conditions (2.21). Then we have

\[ |F_1(n)| \leq |R_1(n)| \cdot \left( e^{-\alpha} \cdot N \cdot \Delta^2(n) \cdot |n|^{\Delta(n)+1} \right). \quad (3.12) \]

Noting that \( |n| \leq 2 \) and \( \Delta(n) \leq 1 \), then

\[ \| F_1 \|_{j_0,N,r} \leq \| R_1 \|_{j_0,N,r} \cdot e^{-\alpha} \cdot N \cdot 2^2 \leq 2^2 \cdot 10N^2 r^3 \epsilon^{1-\sigma}, \quad (3.13) \]

where the last inequality is based on (3.2). On the other hand, for any \( \bar{j} \in \mathbb{Z} \) we have

\[ \partial_{v_1} F_1(n) = \frac{\partial_{v_1} R_1(n)}{\sum_{j \in \mathbb{Z}} (n_j - n_j')v_{1j}} - \frac{R_1(n)}{\left( \sum_{j \in \mathbb{Z}} (n_j - n_j')v_{1j} \right)^2} \partial_{v_1} \left( \sum_{j \in \mathbb{Z}} (n_j - n_j')v_{1j} \right). \]
Then following the proof of (3.13), one has
\[ \|F_1\|_{j_0,N,r} \leq 2^5 \cdot 10N^3 r^3 \epsilon^{1-2\alpha}. \] (3.14)

Then we finish the proof of (3.5) by using (3.13) and (3.14).

Using Taylor’s formula yields
\[ \begin{aligned}
H_2 &:= H_1 \circ X_{F_1}^1 = D_1 + Z_1 \\
&\quad + \{D_1, F_1\} + R_1 + (X_{F_1}^1 - \text{id} - \{\cdot, F_1\}) D_1 + (X_{F_1}^1 - \text{id}) (Z_1 + R_1) \\
&= D_2 + Z_2 + R_2
\end{aligned} \]
where by (3.11),
\[ \begin{aligned}
R_2 &= (R_1 - \mathcal{R}_1) + (X_{F_1}^1 - \text{id} - \{\cdot, F_1\}) D_1 + (X_{F_1}^1 - \text{id}) (Z_1 + R_1) \\
&= \sum_{n \in \mathbb{Z}^2 \times \mathbb{N}^2} R_2(n) \prod_n q_{j}^n - q_{j}^{n'},
\end{aligned} \]
and
\[ \begin{aligned}
(X_{F_1}^1 - \text{id} - \{\cdot, F_1\}) D_1 &:= D_1 \circ X_{F_1}^1 - D_1 - \{D_1, F_1\}, \\
(X_{F_1}^1 - \text{id}) (Z_1 + R_1) &:= (Z_1 + R_1) \circ X_{F_1}^1 - (Z_1 + R_1).
\end{aligned} \]

In the first step, we have \( v_2 = v_1 \) and \( Z_2 = Z_1 \), which implies \( D_2 = D_1 \) and \( Z_2 = Z_1 \). Hence, the estimate (3.6) holds true.

Write
\[ R_2 = \mathcal{R}_2 + (R_2 - \mathcal{R}_2), \]
where \( \mathcal{R}_2 \) is defined by (3.9). By (2.18) and (2.19) in Remark 2.1 and (3.11), for any \( 0 < \sigma < r/2 \) one has
\[ \begin{aligned}
\|\|\mathcal{R}_2\|\|_{j_0,N,r-\sigma} &\leq \left( \frac{\epsilon}{\sigma} \right) \cdot \|\|F_1\|\|_{j_0,N,r} \cdot \|\|H_1\|\|_{j_0,N,r} \\
&\leq 10N r^3 \epsilon \left( \frac{26 \epsilon}{\sigma} \cdot 10N^3 r^3 \epsilon^{1-2\alpha} \right),
\end{aligned} \]
where the last inequality follows from (3.2) and (3.5). This finishes the proof of (3.8). Similarly, we have
\[ \begin{aligned}
\|\|R_2 - \mathcal{R}_2\|\|_{j_0,N,r-\sigma} &\leq 10N r^3 \epsilon + 10N r^3 \epsilon \left( \frac{26 \epsilon}{\sigma} \cdot 10N^3 r^3 \epsilon^{1-2\alpha} \right) \\
&\leq 10N r^3 \epsilon \left( \sum_{i=0}^{1} 2^{-i} \right),
\end{aligned} \]
where the last inequality is based on (3.4).

Finally, the estimate (3.10) follows from (3.2) and (3.5) by using by induction about \( A \).

Precisely, the term in \( R_2 \) comes from \( \frac{1}{j} Z_1^{(j)} \) and \( \frac{1}{j} R_1^{(j)} \) for some \( j \in \mathbb{N} \), where \( Z_1^{(j)} = \{ Z_1^{(j-1)} , H \} \), \( Z_1^{(0)} = Z_1 \), \( R_1^{(j)} = \{ R_1^{(j-1)} , H \} \) and \( R_1^{(0)} = R_1 \). Following the proof of

\( \Downarrow \) Springer
Furthermore, assume for any $A \geq 3$ the following holds

$$
\left| \left\| \sum_{\Delta(n) + |n| = A} (|Z_s(n)| + |R_s(n)|) \prod_{q_j, q_j'} q_j^{n_j} q_j'^{-n_j'} \right\| \right|_{j_0, N, r - (s-1)\sigma} \leq 10N r^3 \epsilon \left( \frac{(10s)^{10s} \cdot 2^{6s} \cdot N^3 r^3 \epsilon^{1-2\alpha}}{\sigma} \right)^{A-3}.
$$

(3.20)

Then there exists a change of variables $\Phi_s := X_{F_s}^1$

$$
H_{s+1} = H_s \circ X_{F_s}^1
$$

$$
= \frac{1}{2} \sum_{j \in \mathbb{Z}} v_{(s+1)j} |q_j|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} Z_{s+1}(n) \prod_{q_j} q_j^{n_j} + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_{s+1}(n) \prod_{q_j} q_j^{n_j} q_j'^{-n_j'}.
$$

3.2 Iterative Lemma

Lemma 3.2 For $s \in \mathbb{N}$ and $1 \leq s \leq \sqrt{N} - 1$, consider the Hamiltonian $H_s(q, \tilde{q})$ of the form

$$
H_s = D_s + Z_s + R_s
$$

$$
= \frac{1}{2} \sum_{j \in \mathbb{Z}} v_{s,j} |q_j|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} Z_s(n) \prod_{q_j} q_j^{n_j} + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_s(n) \prod_{q_j} q_j^{n_j} q_j'^{-n_j'}.
$$

Let $v_s = (v_{s,j})_{j \in \mathbb{Z}}$ satisfy the $(\epsilon, \alpha, N)$-nonresonant condition (2.21). Assume that $0 < \sigma < r/2$ and

$$
\frac{(10(s+1))^{10(s+1)} \cdot 2^{6s} \cdot N^3(s+1) r^3 \epsilon^{1-2\alpha}}{\sigma} \leq \frac{1}{2},
$$

(3.15)

Then there exists a change of variables $\Phi_1$ such that

$$
H_{s+1} = H_s \circ X_{F_s}^1
$$

$$
= \frac{1}{2} \sum_{j \in \mathbb{Z}} v_{(s+1)j} |q_j|^2 + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} Z_{s+1}(n) \prod_{q_j} q_j^{n_j} + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_{s+1}(n) \prod_{q_j} q_j^{n_j} q_j'^{-n_j'}.
$$

$\Phi_1$.
Moreover, one has
\[ ||R_{s+1}||_{j_0,N,r-s\sigma} \leq 10Nr^3\epsilon \left( \sum_{i=0}^{s} 2^{-i} \right), \]
where
\[ R_{s+1}(n) = \prod_{j} q_j^{n_j-n_j'} q_j'. \]

Moreover, we have
\[
\left\| \sum_{\Delta(n)+|n|=A} (|Z_{s+1}(n)| + |R_{s+1}(n)|) \prod_{\supp n} q_j^{n_j-n_j'} \right\|_{j_0,N,r-s\sigma} \leq 10Nr^3\epsilon \left( \frac{(10(s+1))^{10(s+1)}}{\sigma} \cdot 2^6 \cdot N^{3(s+1)}r^3\epsilon^{1-2\alpha} \right)^s.
\]

**Proof** As done before, we know that \( F_s \) will satisfy the homological equation
\[ L_{vs} F_s = \tilde{R}_s, \]
where
\[ \tilde{R}_s(q, \tilde{q}) := \sum_{n\in\mathbb{N}^2 \times \mathbb{N}^2} R_s(n) \prod_{\supp n} q_j^{n_j-n_j'} \cdot q_j'. \]

By the direct computations, one has
\[ F_s(n) = \frac{R_s(n)}{\sum_{j\in\mathbb{Z}} (n_j-n_j') v_{sj}}, \]
unless \( n \in \mathcal{N} \). Since the frequency \( v_s \) satisfies the \((\epsilon, \alpha, N)\)-nonresonant condition (2.21), we get
\[ |F_s(n)| \leq |R_s(n)| \cdot \epsilon^{-\alpha} \cdot N \cdot \Delta^2(n) \cdot |n|^{|\Delta(n)|+1}. \]

In view of (3.18) we have
\[
||F_s||_{j_0,N,r-(s-1)\sigma} \leq 10Nr^3\epsilon \left( \frac{(10s)^{10s} \cdot 2^6 e}{\sigma} \cdot N^{3s}r^3\epsilon^{1-2\alpha} \right)^{s-1} \cdot \epsilon^{-\alpha} \cdot N \cdot \Delta^2(n) \cdot |n|^{|\Delta(n)|+1} \leq \left( 10N^2r^3\epsilon^{1-\alpha} (4s)^{4s} \right) \left( \frac{(10s)^{10s} \cdot 2^6 e}{\sigma} \cdot N^{3s}r^3\epsilon^{1-2\alpha} \right)^{s-1}. \]
where the last inequality is based on $\Delta(n) + |n| \leq s + 2$. Similarly, one has

$$\|F_s\|_{j_0, N, r-(s-1)} \leq (10N^3 r^3 e^{1-2\alpha} (4s)^8) \left( \frac{(10s)^{10s} \cdot 2^6 e}{\sigma} \cdot N^{3s} r^3 e^{1-2\alpha} \right)^{s-1}. \quad (3.29)$$

In view of (3.28) and (3.29), one has

$$\|F_s\|_{j_0, N, r-(s-1)} \leq (20N^3 r^3 e^{1-2\alpha} (4s)^8) \left( \frac{(10s)^{10s} \cdot 2^6 e}{\sigma} \cdot N^{3s} r^3 e^{1-2\alpha} \right)^{s-1},$$

which finishes the proof of (3.21).

Using Taylor’s formula again shows

$$H_{s+1} := H_s \circ X_{F_s}^1$$

$$= D_s + \{D_s, F_s\} + Z_s + R_s + (X_{F_s}^1 - \text{id} - \{\cdot, F_s\}) D_s + \{X_{F_s}^1 - \text{id} \} (Z_s + R_s)$$

$$= D_{s+1} + Z_{s+1} + R_{s+1}$$

$$= \frac{1}{2} \sum_{j \in \mathbb{Z}} u_{(s+1,j)|q_j|^2 + \sum_{n \in \mathbb{N}^2, |n| \geq 4} Z_{s+1}(n) \prod_{\text{supp } n} |q_j^n|^2$$

$$+ \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_{s+1}(n) \prod_{\text{supp } n} q_j^n \bar{q}_j^{n'}. \quad (3.30)$$

Precisely, let

$$G_{s+1} = \{D_s, F_s\} + R_s + (X_{F_s}^1 - \text{id} - \{\cdot, F_s\}) D_s + (X_{F_s}^1 - \text{id}) (Z_s + R_s)$$

$$= \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} G_{s+1}(n) \prod_{\text{supp } n} q_j^n \bar{q}_j^{n'},$$

and then one has

$$D_{s+1} = D_s + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2, n \in \mathcal{N}, |n| = 2} G_{s+1}(n) \prod_{\text{supp } n} q_j^n \bar{q}_j^{n'},$$

$$Z_{s+1} = Z_s + \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2, n \in \mathcal{N}, |n| \geq 4} G_{s+1}(n) \prod_{\text{supp } n} q_j^n \bar{q}_j^{n'},$$

$$R_{s+1} = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} G_{s+1}(n) \prod_{\text{supp } n} q_j^n \bar{q}_j^{n'}. \quad (3.30)$$

Write

$$R_{s+1} = \mathcal{R}_{s+1} + (R_{s+1} - \mathcal{R}_{s+1}),$$

where

$$\mathcal{R}_{s+1} = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} R_{s+1}(n) \prod_{\text{supp } n \cap A(j_0, N_{s+2}) \neq \emptyset} q_j^n \bar{q}_j^{n'}.$$
By (2.18) and (2.19) in Remark 2.1 and (3.21), one has
\[ ||\mathcal{R}_{s+1}||_{j_0,N,r-s\sigma} \leq \left( \frac{e}{\sigma} \right) ||F_s||_{j_0,N,r-(s-1)\sigma} ||Z_{s+1} + \mathcal{R}_s||_{j_0,N,r-(s-1)\sigma} \]
\[ \leq \left( \frac{(10s)^{10s} \cdot 2^6 e}{\sigma} \cdot N^{3s} r^3 \epsilon^{1-2\alpha} \right)^{s} \cdot 10Nr^3 \epsilon \left( \sum_{i=0}^{s-1} 2^{-i} \right) \]
\[ \leq 10Nr^3 \epsilon \left( \frac{(10(s + 1))^{10(s+1)} \cdot 2^6 e}{\sigma} \cdot N^{3(s+1)} r^3 \epsilon^{1-2\alpha} \right)^{s}, \]
which finishes the proof of (3.24). Similarly, we have
\[ ||\mathcal{R}_{s+1} - \mathcal{R}_s||_{j_0,N,r-s\sigma} \leq 10Nr^3 \epsilon \left( \sum_{i=0}^{s} 2^{-i} \right), \]
where the last inequality is based on (3.15). This finishes the proof of (3.23). Similarly, one has
\[ ||\mathcal{Z}_{s+1}||_{j_0,N,r-s\sigma} \leq 10Nr^3 \epsilon \left( \sum_{i=0}^{s} 2^{-i} \right), \]
which finishes the proof of (3.22).

Finally, the estimate (3.25) follows from the proof of (3.10).

\[
3.3 \text{ The Birkhoff Normal Form Theorem}
\]

In this subsection, we will establish the Birkhoff normal form theorem. Fix
\[ N = \left| \frac{\ln \epsilon}{200 \ln |\ln \epsilon|} \right|^2. \tag{3.31} \]
We begin with a key lemma in dealing with the nonresonant condition (2.21). Denote by \( \text{mes}(\cdot) \) the standard product measure on \([0, 1]^\mathbb{Z}\).

**Lemma 3.3** Fix \( \alpha \in (0, 1/100) \) and \( j_0 \in \mathbb{N} \). Then for \( 0 < \epsilon < \epsilon(\alpha) \ll 1 \), there exists some \( \mathcal{R}(j_0) \subset [0, 1]^\mathbb{Z} \) satisfying
\[ \text{mes}(\mathcal{R}(j_0)) \leq \epsilon^{\alpha/2} \]
such that the following holds: if \( v_1 = (v_{1j})_{j \in \mathbb{Z}} \in [0, 1]^\mathbb{Z} \setminus \mathcal{R}(j_0) \), then all \( v_s = (v_{sj})_{j \in \mathbb{Z}} \) with \( 1 \leq s \leq M \leq \sqrt{N} - 1 \) will satisfy the nonresonant condition (2.21), where \( v_s \) (\( 2 \leq s \leq M \)) are inductively defined in the Iterative Lemma (i.e., Lemma 3.2).

**Remark 3.1** Let us comment on the definition and nonresonant properties of \( v_s \) first. Assume \( v_1 = (v_{1j})_{j \in \mathbb{Z}} \) satisfies the nonresonant condition (2.21). Then using Lemma 3.2 yields a modulated frequency \( v_2 = (v_{2j})_{j \in \mathbb{Z}} \) which depends on \( v_1 \). At this stage, \( v_2 \) may not satisfy the nonresonant condition (2.21). To propagate the Iterative Lemma, one can make further restrictions on \( v_1 \) so that \( v_2 \) satisfies (2.21). Repeating this procedure and removing more \( v_1 \) can ensure all \( v_s \) (\( 1 \leq s \leq M \)) satisfy (2.21). The detailed proof is postponed to the next section.

Let \( v = v_1 = (v_{1j})_{j \in \mathbb{Z}} \in [0, 1]^\mathbb{Z} \setminus \mathcal{R}(j_0) \). Then applying the Iterative Lemma gives
Theorem 3.4 (Birkhoff Normal Form) Consider the Hamiltonian (3.1) and assume \( v = v_1 = (v_{1j})_{j \in \mathbb{Z}} \in [0, 1]^2 \setminus \mathcal{R}(j_0) \). Given any \( r > 2 \), then there exists an \( \epsilon^*(r, \alpha) > 0 \) such that, for any \( 0 < \epsilon < \epsilon^*(r, \alpha) \) and any \( M \in \mathbb{N} \) with \( M < \sqrt{N} - 1 \), there exists a symplectic transformation \( \Gamma = \Gamma_1 \circ \cdots \circ \Gamma_M \) such that

\[
\tilde{H} = H_1 \circ \Gamma = \tilde{D} + \tilde{Z} + \tilde{R}
\]

where

\[
\begin{align*}
\|\tilde{Z}\|_{j_0, N, r/2} &\leq 20Nr^3\epsilon, \\
\|\tilde{R}\|_{j_0, N, r/2} &\leq 20Nr^3\epsilon,
\end{align*}
\]

and

\[
\|\tilde{R}\|_{j_0, N, r/2} \leq 10Nr^3\epsilon \cdot (10(M + 1))^{10(M+1)} \cdot 2^6 e \cdot N^{3(M+1)+1}r^2\epsilon^{1-2\alpha}M,
\]

with

\[
\tilde{R} = \sum_{n \in \mathbb{N}^2 \times \mathbb{N}^2} \tilde{R}(n) \prod_{\text{supp } n \cap A(j_0, N/2) \neq \emptyset} q_j^n q_j^{n_j}
\]

Furthermore, for any \( A \geq 3 \) the following estimate holds

\[
\left\| \sum_{\Delta(n)+|n|=A} (|\tilde{Z}(n)| + |\tilde{R}(n)|) \prod_{\text{supp } n} q_j^n q_j^{n_j} \right\|_{j_0, N, r/2} \leq 10Nr^3\epsilon \left( (10(M + 1))^{10(M+1)} \cdot 2^6 e \cdot N^{3(M+1)+1}r^2\epsilon^{1-2\alpha} \right)^{A-3}.
\]

Proof First of all, note that the Hamiltonian (3.1) satisfies all assumptions (3.16)–(3.20) for \( s = 1 \), which follows from (3.2).

Secondly, for given \( r > 2 \) and \( 0 < \alpha < 1/100 \), we take \( \epsilon^* = \epsilon^*(r, \alpha) > 0 \) such that

\[
(10N^*)^{20N^*} \cdot 2^{10}er^2 \cdot (\epsilon^*)^{1-2\alpha} \leq \frac{1}{2},
\]

where

\[
N^* = \left\lfloor \frac{\ln \epsilon^*}{200 \ln |\ln \epsilon^*|} \right\rfloor^2.
\]

Then for any \( 0 < \epsilon < \epsilon^*(r, \alpha) \) and any \( 1 \leq s \leq M \), the assumption (3.15) holds with \( \sigma = \frac{r}{2N} \). Moreover, one has

\[
r - s\sigma \geq r - M \cdot \frac{r}{2N} \geq r/2.
\]

In view of (3.3) and for any \( 1 \leq s \leq M \), one has

\[
N_{s+1} = N - 20s \geq N - 20M = N - 20\sqrt{N} \geq \frac{N}{2}.
\]
which implies
\[ \left[ j_0 - \frac{N}{2}, j_0 + \frac{N}{2} \right] \subset A(j_0, N_s) \subset [j_0 - N, j_0 + N]. \]

Finally, it follows from Lemma 3.3 that all \( v_s \) \((1 \leq s \leq M)\) satisfy the \((\epsilon, \alpha, N)\)-nonresonant condition (2.21). Then by using Iterative Lemma, one can find a symplectic transformation \( \Gamma = \Gamma_1 \circ \cdots \circ \Gamma_M \) such that
\[ \tilde{H} := H_{M+1} = H_1 \circ \Gamma, \]
which satisfies (3.32), (3.33) and (3.34). \( \square \)

4 Estimate on the Measure

In this section, we complete the proof of Lemma 3.3

Proof of Lemma 3.3 Given \( N > 0, j_0 \in \mathbb{N}, n \in \mathbb{N}_Z \) and \( 1 \leq s \leq \left\lfloor \sqrt{N} \right\rfloor - 1 \), define the resonant set \( R_s(n) \) by
\[ R_s(n) = \left\{ v_1 = (v_{1j})_{j \in \mathbb{Z}} \in [0, 1]^Z : \left| \sum_{j \in \mathbb{Z}} (n_j - n'_j) v_{sj} \right| < \frac{\epsilon}{N \Delta^2(n)|n|^\Delta(n)+1} \right\}. \]

Let
\[ R(j_0) = \bigcup_{s=1}^{\left\lfloor \sqrt{N} \right\rfloor - 1} R_s(j_0), \]
where \( R_s(j_0) = \bigcup^{x(s)} R_s(n) \) and the union \( \bigcup^{x(s)} \) is taken for \( n \) satisfying \( \text{supp } n \cap A(j_0, N_{s+1}) \neq \emptyset \) and \( \Delta(n) + |n| \leq s + 2 \). Obviously, we have the counting bound
\[ \#\{ n : \text{supp } n \cap A(j_0, N_{s+1}) \neq \emptyset, \Delta(n) = a, |n| = b \} \leq C (b + N_{s+1}) b^a, \tag{4.1} \]
where \( C > 0 \) is some absolute constant.

It is easy to see that
\[ \text{mes} \left( R'_s(n) \right) \leq \frac{C \epsilon^a}{N \Delta^2(n)|n|^\Delta(n)+1}, \tag{4.2} \]
where
\[ R'_s(n) = \left\{ v_s : \left| \sum_{j \in \mathbb{Z}} (n_j - n'_j) v_{sj} \right| < \frac{\epsilon}{N \Delta^2(n)|n|^\Delta(n)+1} \right\}. \]

Consider first the case \( s = 1 \). Then we have by (4.1) and (4.2)
\[ \text{mes}(R_1(j_0)) \leq \sum_{\text{supp } n \cap A(j_0, N_2) \neq \emptyset} \text{mes}(R_1(n)) \]
\[ \leq C \epsilon^a \sum_{\text{supp } n \cap A(j_0, N_2) \neq \emptyset} \frac{1}{N \Delta^2(n)|n|^\Delta(n)+1} \]
\[ \leq C \frac{N_2}{N} \epsilon^\alpha. \quad (4.3) \]

For \(2 \leq s \leq M\), let \(w^{(s)} = \left( w_j^{(s)} \right)_{j \in \mathbb{Z}}\) with \(w_j^{(s)} = v_{sj} - v_{1j}\) and \(W_s = \sum_{j \in \mathbb{Z}} w_j^{(s)} q_j \tilde{q}_j\). One sees that \(v_{sj} - v_{1j} = 0\) unless \(|j| - j_0| \leq N + 1\). Moreover, in view of (3.18), one has
\[ |||W_s|||_{j_0, N, r - (s - 1)\sigma} \leq 20Nr^3 \epsilon. \]

From Schur’s test,
\[ \left| \frac{\partial w^{(s)}}{\partial v_1} \right|_{\ell^2 \rightarrow \ell^2} \leq 40Nr^3 \epsilon (s + 2) \leq 40N^2r^3 \epsilon, \quad (4.4) \]
as \(\Delta(n) \leq s + 2\) and \(s \leq M < N\). Moreover, (4.4) implies that the frequency modulation map \(v_1 \rightarrow v_s = v_1 + w^{(s)}\) satisfies
\[ e^{-1} \leq (1 - 40N^2r^3 \epsilon)^{2N+2} \leq \det \left| \frac{\partial v_s}{\partial v_1} \right| \leq (1 + 40N^2r^3 \epsilon)^{2N+2} \leq e. \]

Hence, one has by (4.2) and (4.4)
\[ \mes(\mathcal{R}_s(n)) \leq e \cdot \mes(\mathcal{R}_s'(n)) \leq \frac{Ce^\alpha}{N\Delta^2(n)|n|^{\Delta(n)+1}}. \]

Similar to the proof of (4.3), we have
\[ \mes(\mathcal{R}_s(j_0)) \leq Cs^4 \frac{N_{s+1}}{N} \epsilon^\alpha. \]

Finally, by recalling (3.31), we obtain
\[ \mes(\mathcal{R}(j_0)) \leq \sum_{s=1}^{\left[ \sqrt{N} \right]-1} \mes(\mathcal{R}_s(j_0)) \leq C |\log \epsilon|^5 \epsilon^\alpha \leq \epsilon^\alpha/2. \]

This finishes the proof of Lemma 3.3. \(\square\)

### 5 Proof of Main Theorem

Now we are in a position to complete the proof of Theorem 1.2.

**Proof of Theorem 1.2** In view of Theorem 3.4, one obtains the \(\tilde{H}(\tilde{q}, \tilde{\tilde{q}})\) in new coordinates. Then the new Hamiltonian equation is given by
\[ i \tilde{q} = 2 \frac{\partial \tilde{H}}{\partial \tilde{q}}. \quad (5.1) \]

We get by using (5.1) that
\[
\frac{d}{dt} \left| \tilde{q}_j(t) \right|^2 = \left\{ \sum_{|j| > j_0} \left| \tilde{q}_j(t) \right|^2, \tilde{D} + \tilde{Z} + \tilde{K} \right\} = \left\{ \sum_{|j| > j_0} \left| \tilde{q}_j(t) \right|^2, \tilde{K} \right\}.
\]
\[ = 4 \text{Im} \sum_{|j| > j_0} \tilde{q}_j(t) \frac{\partial \tilde{R}}{\partial \tilde{q}} \]

\[ = \sum_{n \in \mathbb{N}^2 \times \mathbb{Z}} \tilde{R}(n) \sum_{|j| > j_0} (n_j - n'_j) \prod_{\text{supp } n} \tilde{q}_j \tilde{q}'_j. \]

In view of (3.35), we decompose \( \tilde{R} \) into three parts:

\[ \tilde{R} = \tilde{R}^{(1)} + \tilde{R}^{(2)} + \tilde{R}^{(3)}, \]

where

\[ \tilde{R}^{(1)} = \tilde{R}, \]

\[ \tilde{R}^{(2)} = \sum_{n \in \mathbb{N}^2 \times \mathbb{Z}} \tilde{R}(n) \sum_{|j| > j_0} (n_j - n'_j) \prod_{\text{supp } n \cap A(j_0, N/2) = \emptyset, \Delta(n) \geq M+4} \tilde{q}_j \tilde{q}'_j, \]

\[ \tilde{R}^{(3)} = \sum_{n \in \mathbb{N}^2 \times \mathbb{Z}} \tilde{R}(n) \sum_{|j| > j_0} (n_j - n'_j) \prod_{\text{supp } n \cap A(j_0, N/2) = \emptyset, \Delta(n) \leq M+3} \tilde{q}_j \tilde{q}'_j. \] (5.2)

Using (3.34) and (3.36) implies

\[ \left\| \tilde{R}^{(1)} + \tilde{R}^{(2)} \right\|_{j_0, N, r/2} \leq 20Nr^3 \epsilon \cdot \left( (10(M + 1))^{10(M+1)} \cdot 2^6 e \cdot N^{3(M+1)+1} r^2 \epsilon^{1-2\alpha} \right)^M. \]

Take

\[ M = \left[ \sqrt{N} \right] - 1 \approx \frac{\ln \epsilon}{200 \ln |\ln \epsilon|}. \]

Then one has

\[ \left\| \tilde{R}^{(1)} + \tilde{R}^{(2)} \right\|_{j_0, N, r/2} \leq \epsilon \cdot \exp \left( -\frac{|\ln \epsilon|^2}{200 \ln |\ln \epsilon|} \right). \] (5.3)

where we use \( 0 < \alpha < \frac{1}{100} \) and \( \epsilon \ll 1. \)

Now consider the monomials in \( \tilde{R}^{(3)} \). Recalling that

\[ \Delta(n) \leq M + 3 < 2\sqrt{N}, \]

if \( \text{supp } n \cap A(j_0, N/2) = \emptyset \), then

\[ \text{supp } n \subset (-\infty, -j_0) \cup (j_0, \infty). \]

Hence the terms in (5.2) satisfy

\[ \sum_{|j| > j_0} (n_j - n'_j) = 0 \] (5.4)

Using (5.3) and (5.4), one has

\[ \frac{d}{dt} \sum_{|j| > j_0} |\tilde{q}_j(t)|^2 \leq \epsilon \cdot \exp \left( -\frac{|\ln \epsilon|^2}{200 \ln |\ln \epsilon|} \right). \]
Integrating in $t$, we obtain
\begin{equation}
\sum_{|j|> j_0} |\tilde{q}_j(t)|^2 \leq \sum_{|j|> j_0} |\tilde{q}_j(0)|^2 + \epsilon \cdot \exp \left( -\frac{\ln \epsilon^2}{200 \ln \ln \epsilon} \right) t. \tag{5.5}
\end{equation}

Note that the symplectic transformation only acts on the $N$-neighborhood of $\pm j_0$. We obtain
\begin{equation}
\sum_{|j|> j_0} |q_j(t)|^2 \leq \sum_{|j|> j_0} |\tilde{q}_j(t)|^2,
\end{equation}

which together with (5.5) gives
\begin{equation}
\sum_{|j|> j_0+N} |q_j(t)|^2 \leq \sum_{|j|> j_0} |\tilde{q}_j(0)|^2 + \epsilon \cdot \exp \left( -\frac{\ln \epsilon^2}{200 \ln \ln \epsilon} \right) t.
\end{equation}

On the other hand, the Hamiltonian preserves the $\ell^2$-norm. So we have
\begin{equation}
\sum_{|j|> j_0} |\tilde{q}_j(0)|^2 = \sum_{j \in \mathbb{Z}} |q_j(0)|^2 - \sum_{|j| \leq j_0} |\tilde{q}_j(0)|^2 < \sum_{|j| \leq j_0-N} |q_j(0)|^2.
\end{equation}

Choosing $\tilde{j}_0$ large enough and letting $j_0 \in [\tilde{j}_0, 2\tilde{j}_0]$ such that
\begin{equation}
\sum_{|j|> j_0-N} |q_j(0)|^2 < \delta,
\end{equation}

then for
\begin{equation}
|t| \leq \delta \cdot \exp \left( \frac{\ln \epsilon^2}{200 \ln \ln \epsilon} \right),
\end{equation}

one has
\begin{equation}
\sum_{|j|> j_0+N} |q_j(0)|^2 \leq 2\delta.
\end{equation}

\[\square\]

**Acknowledgements** H.C. was supported by NNSFC No. 11671066, No. 11401041 and NSFSP No. ZR2019MA062. Y.S. was supported by NNSFC No. 11901010 and Z.Z. was supported by NNSFC No. 11425103. The authors are very grateful to the anonymous referees for valuable suggestions.

**References**

1. Aizenman, M.: Localization at weak disorder: some elementary bounds. Rev. Math. Phys. 6(5A), 1163–1182 (1994). Special issue dedicated to Elliott H. Lieb
2. Aizenman, M., Molchanov, S.: Localization at large disorder and at extreme energies: an elementary derivation. Commun. Math. Phys. 157(2), 245–278 (1993)
3. Anderson, P.W.: Absence of diffusion in certain random lattices. Phys. Rev. 109(5), 1492 (1958)
4. Aizenman, M., Warzel, S.: Random Operators, volume 168 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI (2015). Disorder effects on quantum spectra and dynamics
5. Benettin, G., Fröhlich, J., Giorgilli, A.: A Nekhoroshev-type theorem for Hamiltonian systems with infinitely many degrees of freedom. Commun. Math. Phys. 119(1), 95–108 (1988)
6. Bodyfelt, J.D., Laptyeva, T.V., Skokos, Ch., Krimer, D.O., Flach, S.: Nonlinear waves in disordered chains: probing the limits of chaos and spreading. Phys. Rev. E 84(1), 016205 (2011)
7. Bourgain, J., Wang, W.-M.: Diffusion bound for a nonlinear Schrödinger equation. In: Mathematical Aspects of Nonlinear Dispersive Equations, volume 163 of Ann. of Math. Stud., pp. 21–42. Princeton Univ. Press, Princeton, NJ (2007)
8. Bourgain, J., Wang, W.-M.: Quasi-periodic solutions of nonlinear random Schrödinger equations. J. Eur. Math. Soc. (JEMS) 10(1), 1–45 (2008)
9. Dalfóvo, F., Giorgini, S., Pitaevskii, L.P., Stringari, S.: Theory of Bose-Einstein condensation in trapped gases. Rev. Mod. Phys. 71(3), 463 (1999)
10. Delyon, F., Lévy, Y., Souillard, B.: Anderson localization for multidimensional systems at large disorder or large energy. Comm. Math. Phys. 100(4), 463–470 (1985)
11. Fishman, S., Krivolapov, Y., Soffer, A.: On the problem of dynamical localization in the nonlinear Schrödinger equation with a random potential. J. Stat. Phys. 131(5), 843–865 (2008)
12. Fishman, S., Krivolapov, Y., Soffer, A.: Perturbation theory for the nonlinear Schrödinger equation with a random potential. Nonlinearity 22(12), 2861–2887 (2009)
13. Flach, S., Krimer, D.O., Skokos, Ch.: Universal spreading of wave packets in disordered nonlinear systems. Phys. Rev. Lett. 102(2), 024101 (2009)
14. Fishman, S., Krivolapov, Y., Soffer, A.: The nonlinear Schrödinger equation with a random potential: results and puzzles. Nonlinearity 25(4), R53–R72 (2012)
15. Fröhlich, J., Martinelli, F., Scoppola, E., Spencer, T.: Constructive proof of localization in the Anderson tight binding model. Commun. Math. Phys. 101(1), 21–46 (1985)
16. Fröhlich, J., Spencer, T.: Absence of diffusion in the Anderson tight binding model for large disorder or low energy. Commun. Math. Phys. 88(2), 151–184 (1983)
17. Fröhlich, J., Spencer, T., Wayne, C.E.: Localization in disordered, nonlinear dynamical systems. J. Stat. Phys. 42(3–4), 247–274 (1986)
18. Goldseid, I., Molchanov, S., Pastur, L.: A random homogeneous Schrödinger operator has a pure point spectrum. Funct. Anal. 11(1), 1–10, 96 (1977)
19. Geng, J., You, J., Zhao, Z.: Localization in one-dimensional quasi-periodic nonlinear systems. Geom. Funct. Anal. 24(1), 116–158 (2014)
20. Kirsch, W.: An invitation to random Schrödinger operators. In: Random Schrödinger Operators, volume 25 of Panor. Synthèses, pp. 1–119. Soc. Math. France, Paris, 2008. With an appendix by Frédéric Klopp
21. Pikovsky, A.S., Shepelyansky, D.L.: Destruction of Anderson localization by a weak nonlinearity. Phys. Rev. Lett. 100(9), 094101 (2008)
22. Skokos, Ch., Flach, S.: Spreading of wave packets in disordered systems with tunable nonlinearity. Phys. Rev. E 82(1), 016208 (2010)
23. Skokos, Ch., Krimer, D.O., Komineas, S., Flach, S.: Delocalization of wave packets in disordered nonlinear chains. Phys. Rev. E 82(1), 016208 (2010)
24. Simon, B., Wolff, T.: Singular continuous spectrum under rank one perturbations and localization for random Hamiltonians. Commun. Pure Appl. Math. 39(1), 75–90 (1986)
25. Wang, W.-M.: Logarithmic bounds on Sobolev norms for time dependent linear Schrödinger equations. Commun. Partial Differ. Equ. 33(10–12), 2164–2179 (2008)
26. Wang, W.-M., Zhang, Z.: Long time Anderson localization for the nonlinear random Schrödinger equation. J. Stat. Phys. 134(5–6), 953–968 (2009)
27. Yuan, X.: Construction of quasi-periodic breathers via KAM technique. Commun. Math. Phys. 226(1), 61–100 (2002)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.