Journal of Mathematical Physics

RESEARCH ARTICLE | JANUARY 02 2024

Spacetime quasiperiodic solutions to a nonlinear Schrödinger equation on $\mathbb{Z}$

Special Collection: New Directions in Disordered Systems: In Honor of Abel Klein

Ilya Kachkovskiy ; Wencai Liu ; Wei-Min Wang

J. Math. Phys. 65, 011502 (2024)
https://doi.org/10.1063/5.0166183

Journal of Mathematical Physics ARTICLE pubs.aip.org/aip/jmp

Spacetime quasiperiodic solutions to a nonlinear Schrödinger equation on $\mathbb{Z}$

Cite as: J. Math. Phys. 65, 011502 (2024); doi: 10.1063/5.0166183
Submitted: 3 July 2023 • Accepted: 28 November 2023 •
Published Online: 2 January 2024

Ilya Kachkovskiy,[1,a)] Wencai Liu,[2,b)] and Wei-Min Wang[3,c)]

AFFILIATIONS

[1] Department of Mathematics, Michigan State University, Wells Hall, 619 Red Cedar Rd., East Lansing, Michigan 48824, USA
[2] Department of Mathematics, Texas A & M University, College Station, Texas 77843-3368, USA
[3] CNRS and Département de Mathématique, Cergy Paris Université, 95302 Cergy-Pontoise Cedex, France

Note: Paper published as part of the Special Topic on New Directions in Disordered Systems: In Honor of Abel Klein.
[a)] **Author to whom correspondence should be addressed:** ikachkov@msu.edu
[b)] **Electronic addresses:** liuwencai1226@gmail.com and wencail@tamu.edu
[c)] **E-mail:** wei-min.wang@math.cnrs.fr

ABSTRACT

We consider a discrete non-linear Schrödinger equation on $\mathbb{Z}$ and show that, after adding a small potential localized in the time-frequency space, one can construct a three-parametric family of non-decaying spacetime quasiperiodic solutions to this equation. The proof is based on the Craig–Wayne–Bourgain method combined with recent techniques of dealing with Anderson localization for two-dimensional quasiperiodic operators with degenerate frequencies.

I. INTRODUCTION AND MAIN RESULTS

We study the nonlinear Schrödinger equation on $\mathbb{Z}$:

$$i\frac{\partial}{\partial t}u = \Delta u + Vu + |u|^{2p}u, \quad p \in \mathbb{N}, \tag{1.1}$$

where Δ is the discrete Laplacian:

$$(\Delta\psi)(x) = \psi(x+1) + \psi(x-1), \quad x \in \mathbb{Z},$$

and V is a small non-local potential whose exact form will be described in the end of the section. Consider first the linear equation with zero potential:

$$i\frac{\partial}{\partial t}u = \Delta u. \tag{1.2}$$

Generalized eigenfunctions of the discrete Laplacian Δ can be expressed in the form:

$$u(x) = e^{ij\lambda x}; \quad x \in \mathbb{Z},\ \lambda \in \mathbb{R},\ j \in \mathbb{Z},$$

with eigenvalues $2\cos(j\lambda)$, where $j \in \mathbb{Z}$, $x \in \mathbb{Z}$ and $\lambda \in \mathbb{R}$. The correspondence between pairs (j,λ) and generalized eigenfunctions is not one-to-one: two pairs (j,λ) and (j',λ') correspond to the same eigenfunction if and only if $j\lambda - j'\lambda' \in 2\pi\mathbb{Z}$. This use of two parameters will be convenient, as we later intend to fix λ and restrict ourselves to the sub-family of eigenfunctions parametrized by j.

When λ is an irrational multiple of 2π, the above generalized eigenfunctions are quasiperiodic on $\mathbb{Z}$. These eigenfunctions produce solutions to (1.2):

$$u(x,t) = e^{-2it\cos(j\lambda)}e^{ij\lambda x}.$$

A non-trivial linear combination of such solutions is quasiperiodic both in space and time. Let

$$h_1, h_2 \in \mathbb{Z}\backslash\{0\}; \qquad \omega_1 = 2\cos(h_1\lambda), \qquad \omega_2 = 2\cos(h_2\lambda), \tag{1.3}$$

and assume that $\omega_1 \neq \omega_2$. Then

$$u^{(0)}(t,x) = \sum_{k=1}^{2} u_k e^{-i\omega_k t}e^{ih_k\lambda x} \tag{1.4}$$

is a two-parametric family of solutions to the linear Eq. (1.2). The goal of this paper is to show that there exist solutions to the non-linear Eq. (1.1) which are small perturbations of (1.4), of the form

$$u(t,x) = \sum_{(\mathbf{k},j)\in\mathbb{Z}^2\times\mathbb{Z}} \hat{u}(\mathbf{k},j)e^{i(\mathbf{k}\cdot\omega)t}e^{ij\lambda x}, \tag{1.5}$$

where $\omega = (\omega_1,\omega_2) \in [-3,3]^2, \lambda \in [1,2]$ are frequency parameters,

$$\mathbf{k}\cdot\omega = k_1\omega_1 + k_2\omega_2,$$

$t \in \mathbb{R}_+$ is the time variable, and $x \in \mathbb{Z}$ is the space variable. We will always assume that $\{1,\omega_1,\omega_2\}$ are linearly independent over $\mathbb{Q}$. Most of the analysis will be conducted in the space of variables $(\mathbf{k}, j) \in \mathbb{Z}^3$. It will be convenient to combine them into a single variable

$$\mathbf{n} := (\mathbf{k}, j) \in \mathbb{Z}^3.$$

Denote by $\mathcal{S}(\lambda,\omega_1,\omega_2)$ the class of functions u of the form (1.5) with

$$\sum_{\mathbf{n}\in\mathbb{Z}^3} (1+|\mathbf{n}|)^r|\hat{u}(\mathbf{n})| < +\infty, \qquad \forall r \in \mathbb{N}.$$

Denote also by $\hat{\mathcal{S}}(\lambda,\omega_1,\omega_2) \subset \ell^1(\mathbb{Z}^3)$ the class of the corresponding lattice functions $\hat{u}(\cdot)$on $\mathbb{Z}^3$. The above class is closed under convolution and complex conjugation. As a consequence, for $u \in \mathcal{S}(\lambda,\omega_1,\omega_2)$, we have $|u|^{2p}u \in S(\lambda,\omega_1,\omega_2)$, as well as $\frac{\partial}{\partial t}u$ and Δu. Therefore, this class is a natural space to consider quasiperiodic solutions of the Eq. (1.1) and to define various transformations and functionals.

Ideally, one would like to consider the Eq. (1.1) with $V = 0$. Our current methods fall short of this goal, and, similarly to Refs. 3, 4, and 10, we consider an additional term which will be small in magnitude and have small support in the momentum space. More precisely, let $\mathbf{e}_1 = (1,0)$, $\mathbf{e}_2 = (0,1)$, $h_1 \in \mathbb{Z}\backslash\{0\}, h_2 \in \mathbb{Z}\backslash\{0\}$ and $h_1 \neq \pm h_2$. Denote by $\hat{V}$ the following function on $\mathbb{Z}^2\times\mathbb{Z}$, supported on the set $\{(\mathbf{e}_1,h_1),(\mathbf{e}_2,h_2)\}$:

$$\hat{V}(\mathbf{n}) = m_1\mathbf{1}_{(-\mathbf{e}_1,h_1)}(\mathbf{n}) + m_2\mathbf{1}_{(-\mathbf{e}_2,h_2)}(\mathbf{n}). \tag{1.6}$$

It will be convenient to use the notation $\mathbf{1}_A$ for the indicator function of a subset A. In the case when $A = \{a\}$ is a singleton, we will use $\mathbf{1}_a$ instead, in the situations where it does not lead to a confusion.

Let δ be a small parameter (our results will be asymptotic as $\delta \to 0+$), and $\varkappa$ be a small absolute constant. The potential V will be defined as the following convolution-type operator acting on the space of functions (1.5), or, equivalently, a multiplier acting on the coefficients $\hat{u}(\mathbf{n})$:

$$(Vu)(t,x) := \delta^{\varkappa}\sum_{\mathbf{n}=(\mathbf{k},j)\in\mathbb{Z}^2\times\mathbb{Z}} \hat{V}(\mathbf{n})\hat{u}(\mathbf{n})e^{i(\mathbf{k}\cdot\omega)t}e^{ij\lambda x} = \sum_{s=1,2}\delta^{\varkappa}m_s\hat{u}(-\mathbf{e}_s,h_s)e^{i(-\mathbf{e}_s\cdot\omega)t}e^{ih_s\lambda x}.$$

Clearly, it defines an operator acting on $S(\lambda,\omega_1,\omega_2)$. Rewrite (1.4) as

$$u^{(0)}(t,x) = \sum_{\mathbf{n}\in\mathbb{Z}^2\times\mathbb{Z}} \hat{u}^{(0)}(\mathbf{n})e^{i(\mathbf{k}\cdot\omega^{(0)})t}e^{ij\lambda x}, \tag{1.7}$$

where

$$\hat{u}^{(0)}(\mathbf{n}) = u_1\mathbf{1}_{(-\mathbf{e}_1,h_1)}(\mathbf{n}) + u_2\mathbf{1}_{(-\mathbf{e}_2,h_2)}(\mathbf{n});$$

$$\omega^{(0)} = (\omega_1^{(0)},\omega_2^{(0)}) = (2\cos(h_1\lambda), 2\cos(h_2\lambda)).$$

The presence of the additional parameters m_1, m_2 will allow us to "fine tune" them in order to have the general procedure converge to a solution. The following is our main result.

Theorem 1.1. *For any $\varepsilon > 0$, there exist $\delta_0 = \delta_0(\varepsilon, h_1, h_2) > 0$ such that the following is true for any δ with $0 < \delta < \delta_0$. For $u_i \in (0,1)$, $i = 1, 2$, there exist a closed set $I \subset [1,2]^3$ with*

$$|I| \geq 1 - \varepsilon,$$

and a C^1 map on $(1,2)^3$, $(\lambda, m_1, m_2) \mapsto (\lambda, \omega_1(\lambda, m_1, m_2), \omega_2(\lambda, m_1, m_2))$ such that, for any $(\lambda, m_1, m_2) \in I$, there is a spacetime quasiperiodic solution of the form (1.5) to the nonlinear Schrödinger equation (1.1),

$$u(t,x) = \sum_{\mathbf{n} \in \mathbb{Z}^2 \times \mathbb{Z}} \hat{u}(\mathbf{k}, j) e^{i(\mathbf{k} \cdot \omega)t} e^{ijx\lambda}, \tag{1.8}$$

satisfying:

$$\hat{u}(-\mathbf{e}_1, h_1) = u_1, \qquad \hat{u}(-\mathbf{e}_2, h_2) = u_2,$$

$$|\hat{u}(\mathbf{n})| \leq O(1)\delta^p e^{-|\mathbf{n}|} \text{ for any } \mathbf{n} \in \mathbb{Z}^3 \backslash \{(-\mathbf{e}_1, h_1), (-\mathbf{e}_2, h_2)\},$$

$$\omega_1(\lambda, m_1, m_2) = 2\cos(h_1\lambda) + m_1\delta^{\varkappa} + O(\delta^{2p}); \qquad \omega_2(\lambda, m_1, m_2) = 2\cos(h_2\lambda) + m_2\delta^{\varkappa} + O(\delta^{2p}).$$

A. Motivation

Theorem 1.1 shows the existence of global in time, non-vanishing at infinity and uniformly ℓ^∞ solutions to the nonlinear Eq. (1.1) on $\mathbb{Z}$. Previously, this type of L^∞ solutions was found for the NLS on $\mathbb{R}$.[21] Note that many nonlinear PDE theories address solutions which are localized, in appropriate Sobolev spaces on $\mathbb{R}^d$, for example, or periodic (in space), in Sobolev spaces on the torus $\mathbb{T}^d$, which are compact. On the line $\mathbb{Z}$, the result see Ref. 9, Proposition 2 implies an *a priori* bound on the ℓ^∞-norm, but it does not preclude growth. The results here and in Ref. 21 provide *extended state* solutions to the nonlinear Schrödinger equations, with uniformly bounded ℓ^∞-norms. These are new types of solutions.

The analysis of non-localized solutions without symmetry, such as those in Theorem 1.1 and in Ref. 21 requires different tools. It transforms the extended solution problem (on $\mathbb{Z}$ here) to a *dual* problem in a Fourier space. Establishing a nonlinear version of Anderson localization in the Fourier space (here $\mathbb{Z}^3$) and using duality, lead to extended states solutions.

B. The nonlinear spacetime quasiperiodic problems

When the space direction is periodic, one may work on the torus $\mathbb{T}^d$. The analysis of time quasiperiodic solutions to nonlinear PDEs on the torus has by now become almost a classical subject, see e.g., Refs. 3, 4, and 18, and the review article,[20] see also Refs. 10 and 16. However, most of these techniques are based on the existence of *only one* quasiperiodic direction, namely in time, and the other directions being elliptic, such as in the nonlinear Schrödinger equations. For the nonlinear random Schrödinger equation on the lattice,[6,11,15] the space direction $\mathbb{Z}^d$ is endowed with many parameters. The elliptic directions or the random parameters regularize the problem. The nonlinear problem in this paper has *two* quasiperiodic directions, with *non-decaying* nonlinear term, so most of these techniques are not applicable. We need to develop new tools.

Remark 1.2. The recent work Ref. 17 shows, however, that the space direction may be regularized with *fewer* parameters when the linear problem has Anderson localization and, consequently, the nonlinear term is decaying in space.

C. The new tools

We combine the analysis of linear quasiperiodic problems in higher dimensions[5,8,13] (cf. Ref. 7), with the Lyapunov–Schmidt approach developed for nonlinear time quasiperiodic problems.[4,18] In the space periodic case, the linearized operators are quasiperiodic only in the time direction. In the spacetime quasiperiodic case, however, the linearized operators are quasiperiodic in at least two dimensions. We address this by applying the only known techniques of dealing with multi-dimensional quasiperiodic localization and involving semi-algebraic geometry, the matrix-valued Cartan theorem, and multi-scale procedures, to analyze nonlinear spacetime quasiperiodic problems.

D. Comparison with the space periodic case

As mentioned earlier, the space periodic problem may be reduced to that on the torus. The spectrum of the Laplacian in this case consists of integers, and there are gaps of size at least 1 between non-equal eigenvalues. It is established in Ref. 18 that the NLS:

$$i\frac{\partial}{\partial t}u = \Delta u + |u|^{2p}u, \; p \in \mathbb{N}, \tag{1.9}$$

on $\mathbb{T}^d$ has time quasiperiodic solutions of arbitrary number of frequencies, in any dimension d. The method consists in extracting parameters from the nonlinear term, and subsequently adapting the analysis in Ref. 4, Chap. 19. The uniform spectral gap of *non-equal* eigenvalues permits to control small divisors using the extracted parameters, which are of lower order than $o(1)$, since we seek small, $o(1)$ solutions.

In the spacetime quasiperiodic case, such as here, the spectral gaps are no longer bounded from below. One may still extract parameters, but they are not sufficient for the analysis. The simplest is then to work with related parameter dependent equations. Here we use the parameter V. Note that in problems involving small-divisors, one often starts by studying related parameter dependent equations, such as in the setting of Ref. 4. These parameters are then removed in Ref. 18, yielding solutions to the original NLS in (1.9).

E. On the parameters λ and V

We note that when $V = 0$, λ is the only parameter. From (1.3), the time frequencies $\omega = (\omega_1, \omega_2)$ are functions of λ. Therefore the spacetime frequencies $(\lambda, \omega_1, \omega_2) \in \mathbb{R}^3$ provide only a parameter space of dimension one. In that case, our problem, which is on $\mathbb{Z}^3$, is degenerate.

The recent paper[8] analyzed a degenerate linear quasiperiodic Schrödinger operator on $\mathbb{Z}^2$. The main obstruction to carrying out Green's function estimates in $d = 3$ is the arithmetic lemma. In our case, due to the special structure of the operator, we have a version of the arithmetic lemma, adapted from Ref. 8, that works in $d = 3$. The part that involves the linear analysis does not require the potential V and the parameters (m_1, m_2). However, we had to introduce these parameters in order for the implicit function argument in Sec. VI to be able to run. However, $V = o(1)$ as $\delta \to 0$ is sufficient here, instead of $V = O(1)$, which would have been more typical, cf. e.g., Refs. 3 and 4, Chap. 19. The general scheme of the proof can be considered similar to that in Ref. 21, see also Ref. 19.

F. Structure of the paper

Sections II and III deal with the linearized problem. In Sec. II, we introduce some basic notation for the Green's functions of the linearized operators and state the corresponding arithmetic lemma. In Sec. III, we state a covering lemma (a highly specialized analogue of Ref. 7, Lemma 2.2 and the covering lemma from Ref. 12) and the large deviation theorem which is, again, specialized for the current situation of scale-dependent operators.

In Sec. IV, we state the iteration procedure for the non-linear problem and the role of P- and Q-equations. In Sec. V, we implement the iteration procedure from Sec. IV. This is the most technical section of the paper, since the inductive procedure requires keeping track of regularity and decay estimates of numerous objects at the same time, with some cumbersome relations between these estimates. While these arguments are commonly assumed to be straightforward, we believe that some readers may benefit from the additional level of detail. In Sec. VI, we finish the proof of the main result by using the implicit function theorem and reducing the assumptions on $(\lambda, \omega_1, \omega_2)$ to assumptions on (λ, m_1, m_2). In Sec. VII, we outline the proof of the large deviation theorem (Proposition 3.2) stated in Sec. III.

II. THE GOOD GREEN'S FUNCTIONS AND THE ARITHMETIC LEMMA

Let

$$\mathbb{Z}^3_{\mathrm{pm}} := \mathbb{Z}^3 \times \{+,-\}.$$

A lattice point $\mathbf{n} = (n_1, n_2, n_3) \in \mathbb{Z}^3$ will later be associated with a pair $(\mathbf{k}, j)$ with $\mathbf{k} \in \mathbb{Z}^2$ and $j \in \mathbb{Z}$, as in the original statement of the main theorem. The study of the linearized problem will be centered around operators on $\ell^2(\mathbb{Z}^3_{\mathrm{pm}})$, which can be associated with $\ell^2(\mathbb{Z}^3; \mathbb{C}^2)$. For an operator H, we will denote its matrix elements by

$$H(\mathbf{n}_1, \mathbf{n}_2) \in \mathrm{M}_2(\mathbb{C}), \quad \mathbf{n}_1, \mathbf{n}_2 \in \mathbb{Z}^3.$$

We will use the notation $|H(\mathbf{n}_1, \mathbf{n}_2)|$ to denote the norm of the corresponding (2×2)-matrix; the particular choice of the norm does not matter, and we can for example use the operator norm. We will also use the notation

$$|\mathbf{n}| := |\mathbf{n}|_1 = |n_1| + |n_2| + |n_3|, \qquad |\mathbf{n}|_\infty := \max\{|n_1|, |n_2|, |n_3|\}.$$

An operator H on $\ell^2(\mathbb{Z}^3_{\mathrm{pm}})$ is *Töplitz* if:

$$H(\mathbf{n}_1 + \mathbf{m}, \mathbf{n}_2 + \mathbf{m}) = H(\mathbf{n}_1, \mathbf{n}_2), \quad \forall \mathbf{n}_1, \mathbf{n}_2, \mathbf{m} \in \mathbb{Z}^3.$$

In most cases, we will assume that there exists $\gamma > 0$ such that

$$|H(\mathbf{n}_1, \mathbf{n}_2)| \le e^{-\gamma|\mathbf{n}_1 - \mathbf{n}_2|}. \tag{2.1}$$

Let $D(\theta,\varphi)$ be a family of operators on $\ell^2(\mathbb{Z}^3_{\rm pm})$ parametrized by $(\theta,\varphi)\in\mathbb{R}\times[0,2\pi)$:

$$D(\theta,\varphi)=\begin{bmatrix} D_+ & 0\\ 0 & D_-\end{bmatrix}, \tag{2.2}$$

where $D_\pm$ are diagonal operators on $\ell^2(\mathbb{Z}^3)$ with entries

$$D_\pm(\mathbf{n},\mathbf{n})=\pm(n_1\omega_1+n_2\omega_2+\theta)+2\cos(n_3\lambda+\varphi),\quad \mathbf{n}\in\mathbb{Z}^3.$$

Define also by $T(\theta,\varphi)$ the operator family on $\ell^2(\mathbb{Z}^3_{\rm pm})$:

$$T(\theta,\varphi)=D(\theta,\varphi)+\varepsilon H. \tag{2.3}$$

Note that both D and T are unbounded. However, their domains do not depend on θ and φ.

A. Semi-algebraic sets

A subset $\mathcal{S}\subset\mathbb{R}^n$ is called a (closed) *semi-algebraic set* if it is a finite union of sets defined by a finite number of polynomial inequalities and/or equalities. More precisely, let $\{P_1,\ldots,P_s\}\subset\mathbb{R}[x_1,\ldots,x_n]$ be a family of real polynomials whose degrees are bounded by d. A (closed) semi-algebraic set $\mathcal{S}$ is a subset of the form

$$\mathcal{S}=\bigcup_j\bigcap_{\ell\in\mathcal{L}_j}\{x\in\mathbb{R}^n:\ P_\ell(x)*_{j,\ell}0\}, \tag{2.4}$$

where $\mathcal{L}_j\subset\{1,\ldots,s\}$ and $*_{j,\ell}\in\{\le,=,\ge\}$. In the above case, we say that $\mathcal{S}$ has degree at most sd. In general, the degree of a semi-algebraic set $\mathcal{S}$ is defined to be the smallest possible value sd over all representations, and is denoted by $\deg(\mathcal{S})$.

The following two facts about semi-algebraic sets will be used during the course of the proofs. See 4 for more applications. The first result is a particular case of:[1]

Proposition 2.1. Let $\mathcal{S}\subset\mathbb{R}^n$ be a semi-algebraic set of degree B. Then the number of connected components of $\mathcal{S}$ does not exceed $(1+B)^{C(n)}$.

The second result is known as the Tarski–Seidenberg principle:

Proposition 2.2. Let $\mathcal{S}\subset\mathbb{R}^{d_1+d_2}$ be a semi-algebraic set of degree B. Then the projection of $\mathcal{S}$ onto $\mathbb{R}^{d_1}$

$$\mathrm{proj}_{x_1}(\mathcal{S})=\{x_1\in\mathbb{R}^{d_1}:(\{x_1\}\times\mathbb{R}^{d_2})\cap\mathcal{S}\neq\emptyset\}$$

is a semi-algebraic set with degree at most $(1+B)^{C(d_1,d_2)}$.

Proposition 2.3. Let $\mathcal{S}\subset[0,1]^{d_1}\times[0,1]^{d_2}=[0,1]^d$ be a semi-algebraic set of degree B and $\mathrm{Leb}_d(S)<\eta,\log B\ll\log 1/\eta$. Denote by $(x,y)\in[0,1]^{d_1}\times[0,1]^{d_2}$ the product variable. Fix $\varepsilon>\eta^{1/d}$. Then there is a decomposition

$$\mathcal{S}=\mathcal{S}_1\bigcup\mathcal{S}_2,$$

with $\mathcal{S}_1$ satisfying

$$\mathrm{Leb}(\mathrm{Proj}_x\mathcal{S}_1)\le B^{C(d)}\varepsilon,$$

and $\mathcal{S}_2$ the transversality property

$$\mathrm{Leb}(\mathcal{S}_2\cap L)\le B^{C(d)}\varepsilon^{-1}\eta^{1/d},$$

for any d_2-dimensional hyperplane L in $[0,1]^{d_1+d_2}$ such that

$$\max_{1\le j\le d_1}|\mathrm{Proj}_L(e_j)|\le\frac{1}{100}\varepsilon,$$

where e_j are the basis vectors for the x-coordinates.

The notation $\log B \ll \log 1/\eta$ means that there exists a (small) constant $c = c(n)$ such that the conclusion of the lemma holds under the assumption $\log B \le c \log 1/\eta$. The above lemma, referred to as the *steep planes lemma*, is the basic tool underlining the semi-algebraic techniques used in the subject. It is stated as (1.5) in Ref. 5, cf., Lemma 9.9[4] and Proposition 5.1 Ref. 8, and relies on the Yomdin–Gromov triangulation theorem. For a complete proof of the latter, see Ref. 2.

B. The arithmetic lemma

The Green's function method of proving Anderson localization involves restrictions of the operator to various boxes. Let $N \in \mathbb{N}$, and T be an operator on $\ell^2(\mathbb{Z}^3_{\mathrm{pm}})$. For a subset $\Lambda \subset \mathbb{Z}^3$, let $\Lambda_{\mathrm{pm}} := \Lambda \times \{-,+\} \subset \mathbb{Z}^3_{\mathrm{pm}}$, and

$$T|_\Lambda := \mathbf{1}_{\Lambda_{\mathrm{pm}}} T \mathbf{1}_{\Lambda_{\mathrm{pm}}}\big|_{\mathrm{Ran}\, \mathbf{1}_{\Lambda_{\mathrm{pm}}}}$$

the block of T corresponding to the subspace $\ell^2(\Lambda_{\mathrm{pm}})$, naturally embedded into $\ell^2(\mathbb{Z}^3_{\mathrm{pm}})$, considered as an operator acting on a finite-dimensional subspace. For boxes centered at the origin, we will use shorter notation

$$T|_N := T|_{[-N,N]^3}.$$

Let $r(\Lambda)$ be the diameter of Λ. Denote by $X^g_{\Lambda,\mu}$ (the "good set") the set of pairs $(\theta,\varphi) \in \mathbb{R} \times [0,2\pi]$such that

$$\|T(\theta,\varphi)|_\Lambda^{-1}\| \le e^{r(\Lambda)^{\frac{6}{7}}}, \tag{2.5}$$

and for any $\mathbf{n},\mathbf{n}' \in \Lambda$ with $|\mathbf{n}-\mathbf{n}'| \ge r(\Lambda)^{999/1000}$, we have the following estimates for the matrix elements:

$$|T(\theta,\varphi)|_\Lambda^{-1}(\mathbf{n},\mathbf{n}')| \le e^{-\mu|\mathbf{n}-\mathbf{n}'|}. \tag{2.6}$$

Let $X^b_{\Lambda,\mu} = \mathbb{R} \times [0,2\pi] \backslash X^g_{\Lambda,\mu}$ be the "bad set," that is, the complement of $X^g_{\Lambda,\mu}$. As always, we will use the notation $X^b_{N,\mu}$ and $X^g_{N,\mu}$ for the case $\Lambda = [-N,N]^3$.

In the multi-scale estimates, it will be important to consider different values of μ for different scales N. For large N, one would expect $\mu < \gamma$, since the factors with decay rate γ will appear in the resolvent identity. On each step of the multi-scale procedure, the rate deteriorates (that is, gets weaker).

In the setting of the present paper, the situation will be complicated by the fact that the operator H itself, as well as the value of γ will also change from scale to scale. One needs to be careful in obtaining a meaningful version of the multi-scale argument for the large deviation theorem. We will formulate such a version in the end of Sec. III.

We will start from some preparations. Compared to the traditional quasiperiodic setting such as Ref. 7, the set of parameters θ,φ under consideration is not compact. However, one can restrict the bad values of the parameters to a bounded N-dependent subset.

Lemma 2.4. Suppose that $|\varepsilon| < e^{-10(1+\gamma)}$ and $|D_\pm(\mathbf{n},\mathbf{n})| \ge r(\Lambda)^3$ for all $\mathbf{n} \in \Lambda$. Then (2.5) and (2.6) are satisfied. As a consequence, we have the following inclusion:

$$X^b_{\Lambda,\mu} \subset [-10N^3, 10N^3] \times [0,2\pi], \quad \forall \mu \in (\gamma/5, 5\gamma).$$

Proof. Both properties follow from considering T being a small perturbation of its diagonal part, and expanding it into the Neumann series. Additionally, (2.5) becomes much stronger with $O(N^{-1})$ in the right hand side. □

As in many results, we will need to introduce Diophantine conditions on the frequency vectors. Fix $C_{\mathrm{dio}}, \delta_{\mathrm{dio}} > 0$. Let

$$\mathrm{DC}_1(C_{\mathrm{dio}},\delta_{\mathrm{dio}},N) = \left\{\alpha \in \mathbb{R} : \mathrm{dist}(k\alpha,\mathbb{Z}) \ge \frac{C_{\mathrm{dio}}}{|k|^{1+\delta_{\mathrm{dio}}}}, \forall k \in \mathbb{Z},\ 1 \le |k| \le N\right\};$$

$$\mathrm{DC}_1(C_{\mathrm{dio}},\delta_{\mathrm{dio}}) := \bigcap_N \mathrm{DC}_1(C_{\mathrm{dio}},\delta_{\mathrm{dio}},N).$$

We will often take $\delta_{\mathrm{dio}} = 1/100$ and denote C_{dio} by τ. To simplify the notations, in this case, we write the Diophantine condition as $\mathrm{DC}_1(\tau)$.

The following result is essentially proven in Ref. 8, Theorem 5.1. The argument in Ref. 8 assumes $\alpha_1 = \alpha_2$, but one can check that this assumption is not necessary for the argument.

Lemma 2.5. Assume $\alpha_1, \alpha_2 \in \mathrm{DC}_1(C_{\mathrm{dio}}, \delta_{\mathrm{dio}})$. Let $\mathcal{A} \subset [0, 2\pi)^2$ be a semi-algebraic subset. Assume that, for any unit line segment $L \subset [0, 2\pi)^2$, we have

$$\mathrm{Leb}_1(\mathcal{A} \cap L) \le \frac{1}{2} \min_{1 \le |\mathbf{k}| \le 2N} \mathrm{dist}(\mathbf{k} \cdot \alpha, \mathbb{Z}).$$

Then

$$\#\{\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2 : |\mathbf{k}| \le N, (k_1\alpha_1, k_2\alpha_2) \mod \mathbb{Z}^2 \in \mathcal{A}\} \le (1 + \deg \mathcal{A})^{C_{\mathrm{abs}}} C_{\mathrm{dio}}^{-C_{\mathrm{abs}}} N^{\frac{3}{4} + 3\delta_{\mathrm{dio}}}. \tag{2.7}$$

for $N \ge N_0(\alpha)$.

Here and in the future, $C_{\mathrm{abs}} > 0$ denotes some positive absolute constant (not necessarily the same in different places).

The following is a modification of the above lemma which is essential for the multi-scale arguments afterwards.

Theorem 2.6. *Let $|\varepsilon| < e^{-10(1+\gamma)}$ and $\gamma/5 < \mu < 5\gamma$. Fix some $\nu_2 > 0$. Let $\omega_1, \omega_2 \in (-3, 3)$ with $|\omega_2| \ge \nu_2$ and $\frac{\omega_1}{\omega_2} \in \mathrm{DC}_1(\tau)$. Let $\lambda \in (1, 2)$ with $\lambda \in \mathrm{DC}_1(C_{\mathrm{dio}}, \delta_{\mathrm{dio}})$.*

Let $\Lambda \subset \mathbb{Z}^3$ with $r(\Lambda) < N_2$. Assume that the set $X^b_{\Lambda,\mu}$ is contained in a semi-algebraic subset $X^{\mathrm{b,alg}}_{\Lambda,\mu}$ with

$$\mathrm{Leb}_1(X^{\mathrm{b,alg}}_{\Lambda,\mu} \cap L) \le e^{-r(\Lambda)^\rho}, \qquad \deg X^{\mathrm{b,alg}}_{\Lambda,\mu} \le r(\Lambda)^C, \tag{2.8}$$

uniformly in all unit line segments $L \subset \mathbb{R} \times [0, 2\pi)$. Then, for any $(\theta, \varphi) \in \mathbb{R} \times [0, 2\pi]$, we have

$$\#\{\mathbf{n} \in \mathbb{Z}^3 : |\mathbf{n}| \le N_2, (\theta + \mathbf{k} \cdot \omega, \varphi + j\lambda) \in X^b_{\Lambda,\mu}\} \le (\nu_2\tau)^{-C_{\mathrm{abs}}} r(\Lambda)^C N_2^{3/4 + 3\delta_{\mathrm{dio}}}, \tag{2.9}$$

assuming N_2 is large enough depending on (ω, λ) and ρ.

Proof. Without loss of generality, assume that $\omega_2 > 0$. Let $\eta = \frac{\omega_1}{\omega_2}$. Denote by Y^b_Λ the subset of $[0, 1) \times [0, 2\pi)$, defined by the following condition:

$$(\theta + k_1\omega_1 + k_2\omega_2, \varphi + j\lambda) \in X^b_{\Lambda,\mu} \Leftrightarrow (\{k_1\eta\}, \{j\lambda\}) \in Y^b_\Lambda.$$

To obtain Y_Λ from $X^b_{\Lambda,\mu}$, one can first stretch the latter by ω_2^{-1} in the first coordinate. The resulting set will be contained in $[-10\, r(\Lambda)^3 \nu_2^{-1}, 10\, r(\Lambda)^3 \nu_2^{-1}] \times [0, 2\pi)$. One can then eliminate the variable k_2 by slicing the set into $O(r(\Lambda)^3 \nu_2^{-1})$ pieces and translating each piece into $[0, 1) \times [0, 2\pi)$. An additional translation, or a slight modification of the previous step, would also eliminate the dependence on (θ, φ).

Since the constants are allowed to depend on ν_2, the new set Y^b_Λ will satisfy (2.8) with, say, $e^{-r(\Lambda)^\rho/2}$ in the right hand side. If $X^b_{\Lambda,\mu}$ was contained in a semi-algebraic subset of degree B, then one can apply the same operations to that set and obtain a semialgebraic approximation for Y^b_Λ of degree at most $10Br(\Lambda)^3\nu_2^{-1}$.

With the above preparations, Lemma 2.5 implies the desired bound for the number of pairs (k_1, j). To complete the proof, recall Lemma 2.4 and note that, once k_1 is fixed, there are at most $O(\nu_2^{-1} r(\Lambda)^3)$ possibilities to choose k_2. The contribution from these possibilities can be absorbed into the front factor. □

Remark 2.7. If $\omega_2 \in \mathrm{DC}_1(\tau)$, then $|\omega_2| \ge \tau$. However, the proof does not require the individual frequencies to be Diophantine. One cannot remove the condition $|\omega_2| \ge \nu_2$: in the extreme case $\omega_2 = 0$ we have $O(k_2)$ possibilities for k_2 alone, which will make $o(k_2)$ impossible to achieve. Since $\frac{\omega_1}{\omega_2} \in \mathrm{DC}_1(\tau)$, we have $|\omega_1| \ge \tau\nu_2$ as part of the assumptions.

In later applications of Theorem 2.6, it will be convenient to have $\nu_2 = \tau$. In this case, the factor $(\tau\nu)^{-C_{\mathrm{abs}}}$ becomes $\tau^{-C_{\mathrm{abs}}}$.

III. COVERING LEMMA AND THE LARGE DEVIATION THEOREM

A. Resolvent identity and the covering lemma

Let $\Lambda = \Lambda_1 \cup \Lambda_2 \subset \mathbb{Z}^3$, and $\Lambda_1 \cap \Lambda_2 = \emptyset$. Let T be an operator on $\ell^2(\Lambda)$. From the resolvent identity, we have

$$T|_\Lambda^{-1} = T|_{\Lambda_1}^{-1} + T|_{\Lambda_2}^{-1} - (T|_{\Lambda_1}^{-1} + T|_{\Lambda_2}^{-1})(T|_\Lambda - T|_{\Lambda_1} - T|_{\Lambda_2})T|_\Lambda^{-1}. \tag{3.1}$$

Here, we assume that $T|_{\Lambda_j}^{-1}$ is extended by zero into $\ell^2((\Lambda \backslash \Lambda_j)_{\mathrm{pm}})$, in order for the addition of operators to make sense. Assume that T satisfies the Töplitz condition:

$$|T(\mathbf{n}_1, \mathbf{n}_2)| \le e^{-\gamma|\mathbf{n}_1 - \mathbf{n}_2|},$$

and let $\mathbf{n}_1 \in \Lambda_1$, $\mathbf{n}_2 \in \Lambda$. Then

$$|T|_\Lambda^{-1}(\mathbf{n}_1,\mathbf{n}_2)| \le |T|_{\Lambda_1}^{-1}(\mathbf{n}_1,\mathbf{n}_2)|\mathbf{1}_{\Lambda_1}(\mathbf{n}_2) + \sum_{\mathbf{n}'\in\Lambda_1,\,\mathbf{n}''\in\Lambda_2} e^{-\gamma|\mathbf{n}'-\mathbf{n}''|}|T|_{\Lambda_1}^{-1}(\mathbf{n}_1,\mathbf{n}')\|T|_\Lambda^{-1}(\mathbf{n}_2,\mathbf{n}'')|. \tag{3.2}$$

In the inductive procedure of constructing solutions to the P-equations, one needs to apply the above resolvent identity several times in a specific way. In one way or another, these ideas have been used in most of the applications of the Craig–Wayne–Bourgain method. A detailed scheme has been recently clarified in Ref. 12. For the convenience of the reader, we describe a version of this argument adapted to our setting. We will start from the "covering lemma." Suppose that T satisfies the following conditions

(cov1) The Töplitz condition $|T(\mathbf{n}_1,\mathbf{n}_2)| \le e^{-\gamma|\mathbf{n}_1-\mathbf{n}_2|}$.

(cov2) There will be boxes of three sizes: the main box $\Lambda = [-N,N]^3$, the central box $\Lambda_c = [-N_1,N_1]^3$, and several "small" boxes of the form $\Lambda_s = \mathbf{n} + [-M,M]^3 \subset \Lambda$. We will assume

$$\sqrt{N} \le N_1 \le N/10; \qquad (\log N)^{100} \le M \le (\log N)^{\log\log N}.$$

All results will hold under the assumption that N is larger than an absolute constant (in particular, all logarithms are well defined).

(cov3) The box Λ_c is good in the following sense:

$$\|T|_{\Lambda_c}^{-1}\| \le B,$$

$$|T|_{\Lambda_c}^{-1}(\mathbf{n}_1,\mathbf{n}_2)| \le e^{-\mu|\mathbf{n}_1-\mathbf{n}_2|}, \quad \text{for } |\mathbf{n}_1-\mathbf{n}_2| \ge R,$$

where $100M \le R \le N_1^{999/1000}$ and $0 < B \le e^{N_1^{999/1000}}$.

(cov4) For each point $\mathbf{n} \in \Lambda\backslash\frac{1}{2}\Lambda_c$, there exists a box $\Lambda_s \subset \Lambda$ of size M, containing $\mathbf{n}$, with $\operatorname{dist}(\mathbf{n},\Lambda\backslash\Lambda_s) \ge M/20$, which is good in the following sense:

$$\|T|_{\Lambda_s}^{-1}\| \le e^{M^{6/7}},$$

$$|T|_{\Lambda_s}^{-1}(\mathbf{n}_1,\mathbf{n}_2)| \le e^{-\mu|\mathbf{n}_1-\mathbf{n}_2|}, \quad \text{for } |\mathbf{n}_1-\mathbf{n}_2| \ge M^{999/1000}.$$

Here, $\gamma,\ \mu \in [1,+\infty)$. For each point $\mathbf{n} \in \Lambda$, the default choice of a "good box" will be Λ_s provided in (4) or Λ_c if $\mathbf{n} \in \frac{1}{2}\Lambda_c$. Sometimes we will need to be more careful and still choose Λ_c even if a legitimate Λ_s exists.

Proposition 3.1. Under the above assumptions and for sufficiently large N, the Green's function in the box Λ is good in the sense of the following two bounds: the norm bound

$$\|T|_\Lambda^{-1}\| \le 2(e^{M^{6/7}} + B)$$

and the exponential decay bound: for $|\mathbf{n}_1-\mathbf{n}_2| \ge R$, we have

$$|T|_\Lambda^{-1}(\mathbf{n}_1,\mathbf{n}_2)| \le e^{-\left(\min\{\gamma,\mu\} - \frac{1000\log N}{M}\right)(|\mathbf{n}_1-\mathbf{n}_2| - M^{999/1000})}.$$

Proof. In multiple aspects, the proof follows standard arguments dating back to Ref. 7. Let us first consider the norm estimate. For each $\mathbf{n}_1 \in \Lambda$, let Λ_1 be the "default good box" around it; that is, Λ' from (3) if it exists, or Λ_c. Apply (3.2) and first take the supremum over $\mathbf{n}_2$; then take a supremum over $\mathbf{n}_1$; note that, since Λ_1 depends on $\mathbf{n}_1$, the right hand side of (3.2) will be different each time. As a consequence, we obtain

$$s(1-t) \le (e^{M^{6/7}} + B),$$

where

$$s = \max_{\mathbf{n}_1,\mathbf{n}_2\in\Lambda} |T|_\Lambda^{-1}(\mathbf{n}_1,\mathbf{n}_2)|,$$

$$t := \sum_{\Lambda_1\subset\Lambda,\mathbf{n}\in\Lambda_1,\mathbf{n}''\in\Lambda\backslash\Lambda_1,\operatorname{dist}(\mathbf{n},\partial\Lambda_1\backslash\partial\Lambda)\ge\operatorname{diam}(\Lambda_1)/40} e^{-|\mathbf{n}-\mathbf{n}''|}|T|_{\Lambda_1}^{-1}(\mathbf{n}_1,\mathbf{n}')|.$$

Here, N is assumed to be large enough so that the exponential factor, which is either bounded by $e^{-M/100}$ or $e^{-N_1/100}$ absorbs the last factor bounded by $e^{M^{6/7}}$ or $e^{N^{6/7}}$, respectively, and all combinatorial factors. We also weakened the estimates by assuming that $\gamma = \mu = 1$. It is clear that we have $t \le 1/2$ for N larger than an absolute constant, which implies the norm estimate.

Let us proceed with the exponential decay estimate. Let $\mathbf{n}_1, \mathbf{n}_2 \in \Lambda$ and $|\mathbf{n}_1 - \mathbf{n}_2| \ge R$. Find a good box Λ_1 around $\mathbf{n}_1$ and apply (3.2). Assume for now that $\mathbf{n}_2 \notin \Lambda_1$. Then we have

$$|T|_\Lambda^{-1}(\mathbf{n}_1, \mathbf{n}_2)| \le \sum_{\mathbf{n} \in \Lambda_1,\, \mathbf{n}'' \in \Lambda \setminus \Lambda_1} e^{-\gamma|\mathbf{n}' - \mathbf{n}''| - \mu|\mathbf{n}_1 - \mathbf{n}'|} |T|_\Lambda^{-1}(\mathbf{n}_2, \mathbf{n}'')|.$$

For $\mathbf{n}$ and $\mathbf{n}''$ that produce the largest possible term in the summation, we have

$$\begin{aligned} |T|_\Lambda^{-1}(\mathbf{n}_1, \mathbf{n}_2)| \le{}& |T|_{\Lambda_1}^{-1}(\mathbf{n}_1, \mathbf{n}_2)| \mathbf{1}_{\Lambda_1} \mathbf{n}_2 \\ &+ |\Lambda_1| |\Lambda| e^{-\min\{\gamma,\mu\}(|\mathbf{n}_1 - \mathbf{n}| + |\mathbf{n} - \mathbf{n}''|)} |T|_\Lambda^{-1}(\mathbf{n}'', \mathbf{n}_2)| \\ \le{}& |T|_{\Lambda_1}^{-1}(\mathbf{n}_1, \mathbf{n}_2)| \mathbf{1}_{\Lambda_1} \mathbf{n}_2 + |\Lambda_1| |\Lambda| e^{-\min\{\gamma,\mu\}|\mathbf{n}_1 - \mathbf{n}''|} |T|_\Lambda^{-1}(\mathbf{n}'', \mathbf{n}_2)|. \end{aligned} \tag{3.3}$$

Let $\mathbf{n}_1'' := \mathbf{n}''$. One can iterate the above inequality by applying it to $\| T|_\Lambda^{-1}(\mathbf{n}_2, \mathbf{n}_1'') \|$ with $\mathbf{n}_1$ replaced by $\mathbf{n}_1''$, and obtain a new "intermediate point" $\mathbf{n}_2''$. The process can be repeated, and we obtain a "path" $\mathbf{n}_1'', \mathbf{n}_2'', \ldots$. For each point $\mathbf{n}_k''$ we find a good box Λ_1 containing $\mathbf{n}_k''$, and the next point $\mathbf{n}_{k+1}''$ will be outside of Λ_1. We will call the jump from $\mathbf{n}_k''$ to $\mathbf{n}_{k+1}''$ *short* or *long* depending on whether Λ_1 is a large or small box. Note that this terminology only reflects a lower bound on the size of the jump; the actual jump can be arbitrarily large within Λ, but cannot be very small since the next point must be outside of Λ_1.

The iterations can be repeated indefinitely. At each step, the number of terms increases at most by 1. Generally, we expect that the first term in the right hand side of (3.2) would only appear at later stages, since $\mathbf{n}_1$ is not very close to $\mathbf{n}_2$. However, one can easily see that eventually it will appear. In that case, we estimate the first term in the right hand side of (3.3) with the best available bound on the matrix element of $T|_{\Lambda_1}^{-1}$, which can either be $e^{M^{5/6}}$, B, or an exponentially decaying bound. Let us terminate the iterations once we meet the condition

$$|\mathbf{n}_1 - \mathbf{n}_1''| + |\mathbf{n}_1'' - \mathbf{n}_2''| + \cdots + |\mathbf{n}_{r-1}'' - \mathbf{n}_r''| \ge 2N_1;$$

recall that N_1 is the size of the central box. In the last step, we estimate the matrix element of $T|_\Lambda^{-1}$ by the (already obtained) norm bound $2(B + e^{M^{5/6}}) \le 4B$.

As a consequence, we have at most, say, $100N_1/M$ terms. Each term corresponds to some sub-path of the initial path. It includes an exponential factor (total length of the sub-path), a combinatorial factor $(|\Lambda_1| |\Lambda|)^r$, where r is the number of steps in the sub-path, and the norm factor, which is either, say $4B$ or $e^{M^{5/6}}$.

To simplify the estimates, first note that the number of terms can be absorbed into the overall bound, and therefore one can deal with the worst possible individual term. Let us consider the combinatorial factor. For a sub-path with r steps, the exponential factor is at most (meaning at least as good as)

$$e^{-\min\{\gamma,\mu\}\frac{Mr}{100}}.$$

The combinatorial factor is bounded by

$$(|\Lambda_1| |\Lambda|)^r \le N^{100r} \le e^{100r \log N}.$$

Since $M \ge (\log N)^{100}$, the combinatorial factor can always be absorbed into the exponential factor for large N, with some deterioration in the factor $\min\{\gamma, \mu\}$ as indicated in the statement of the proposition.

It remains to consider how to deal with the norm factors. Clearly, if the norm factor is $e^{M^{5/6}}$, then the length of the sub-path must be at least $|\mathbf{n}_1 - \mathbf{n}_2| - M^{999/1000}$, and therefore the norm factor is, again, absorbed by the exponential one. If the norm factor is $4B$, it can come either from a sub-path [corresponding to the first term of (3.3) at some iteration step], or from the second term through the length termination condition. In the latter case, since $B \le e^{N_1^{999/1000}}$ and the length of the path is at least $2N_1$, it will again be absorbed by the exponential term.

It remains to consider the situations where the factor B appears in the first term. It can only happen if, at some stage, $\mathbf{n}_k$ is inside of $\Lambda_c/2$ (so that we have to choose Λ_c as a good box), and $|\mathbf{n}_k - \mathbf{n}_2| < R$. In that case, note that $\mathbf{n}_{k-1}$ must be outside of $\Lambda_c/2$. Therefore, the jump from $\mathbf{n}_{k-1}$ to $\mathbf{n}_k$ must have been through a small box (using our terminology, a short jump), but the actual length of the jump, and therefore the length of the path, is at least, say, $N_1/10$. Thus, the exponential factor will also absorb $B \le e^{N_1^{999/1000}}$. □

B. The large deviation theorem

Our next goal is the scale-dependent version of the large deviation theorem for the Green's functions. Let

$$2\gamma_\infty \geq \gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_\infty, \qquad 2\gamma_\infty \geq \gamma_1' \geq \gamma_2' \geq \cdots \geq \gamma_\infty$$

be two non-increasing sequences of real numbers. As above, define a family of operators on $\ell^2(\mathbb{Z}^3_{\mathrm{pm}})$:

$$T_k(\theta,\varphi) = D(\theta,\varphi) + \varepsilon H_k.$$

One can also consider their restrictions into $\mathbb{Z}^d_{\mathrm{pm}}\backslash S$ for a finite subset S in the same way as described above. We omit this straightforward part for simplicity. Assume the following conditions on T_k and H_k.

(ms1) H_k are Töplitz operators with the matrix elements satisfying

$$|H_k(\mathbf{n},\mathbf{n}')| \leq e^{-\gamma_k|\mathbf{n}-\mathbf{n}'|}.$$

(ms2) $\|T_{k+1} - T_k\| \leq e^{-k^2}$.
(ms3) $\gamma'_{k^{C_{\mathrm{ind}}}} \leq \min\{\gamma_k,\gamma_k'\} - k^{-10}$.

The following Proposition 3.2 is a version of the large deviation theorem. There no major difficulties in adapting the known proof from Ref. 7 to the variable decay rate situation and to arbitrary dimension (in our case, $d = 3$), assuming that an appropriate version of the arithmetic lemma holds (Theorem 2.6). However, one also needs to use modified versions of the covering lemmas Ref. 7, Lemmas 2.2, 2.4 which take into account the Töplitz form of the off-diagonal terms, more accurate counting of bad boxes, and stronger off-diagonal decay (that is, $|\mathbf{n}-\mathbf{n}'| \geq N^{999/1000}$ rather than $|\mathbf{n}-\mathbf{n}'| \geq N/100$). Such versions were stated in Refs. 12 and 14, with the latter being the most suitable for our direct applications (the statement in Ref. 14 would have needed a minor modification).

It will be convenient to introduce a new arithmetic condition on the spacetime frequencies. Let

$$\mathrm{DCR}(\tau,N) = \left\{(\lambda,\omega_1,\omega_2) \in (1,2)\times(-3,3)^2 : |\omega_2| \geq \tau,\ \frac{\omega_1}{\omega_2} \in \mathrm{DC}_1(\tau;N),\ \lambda \in \mathrm{DC}_1(\tau;N)\right\};$$

$$\mathrm{DCR}(\tau) = \bigcap_N \mathrm{DCR}(\tau;N).$$

One can check that

$$\mathrm{Leb}_3(\mathrm{DCR}(\tau;N)\backslash\mathrm{DCR}(\tau;N+1)) \leq 10\tau^2 N^{-(2+2/100)}; \tag{3.4}$$

$$\mathrm{Leb}_3((1,2)\times(-3,3)^2\backslash\mathrm{DCR}(\tau)) \leq 10\tau. \tag{3.5}$$

Proposition 3.2. There exists an absolute constant $C_{\mathrm{ind}} \geq 2$ such that, for every $\gamma_\infty \in (1/2,10)$ there exists $\varepsilon_0 > 0$ such that, for any $\varepsilon \in (0,\varepsilon_0)$, we have the following. For every $M \in \mathbb{N}$, and every $(\lambda,\omega_1,\omega_2) \in \mathrm{DCR}(\tau,M^{10})$ there exists a subset $X_M = X_M(\lambda,\omega_1,\omega_2) \subset \mathbb{R}\times\mathbb{T}$ such that, for every $(\theta,\varphi) \in (\mathbb{R}\times\mathbb{T})\backslash X_M$, we have

$$\|(T_k(\theta,\varphi)|_M)^{-1}\| \leq e^{M^{\frac{6}{7}}}, \quad \forall k \geq M. \tag{3.6}$$

and for all $\mathbf{n},\mathbf{n} \in [-M,M]^3$ with $|\mathbf{n}-\mathbf{n}'| \geq N^{999/1000}$ we have

$$|(T_k(\theta,\varphi)|_M)^{-1}(\mathbf{n},\mathbf{n}')| \leq e^{-\gamma_M'|\mathbf{n}-\mathbf{n}'|}, \quad \forall k \geq M. \tag{3.7}$$

For every line segment $L \subset (-3,3)^2$, X_M satisfies

$$\mathrm{Leb}_1(X_M \cap L) \leq e^{-M^{\frac{1}{300}}}. \tag{3.8}$$

Remark 3.3. For the convenience of the reader, the main steps of the proof will be outlined in Sec. VII. As previously, we assume that $\delta_{\mathrm{dio}} = 1/100$ is fixed. An analysis of the proof shows that the choice

$$\varepsilon_0 = \exp\{-\tau^{-C_{\mathrm{abs,dio}}}\}$$

is sufficient. See also Remark 7.6.

Remark 3.4. We will be using the condition (3.8) only for horizontal and vertical line segments. In that way, the steep planes analysis in Sec. V is more similar to Refs. 5 and 6 rather than.[7] Most likely, the Proof of Proposition 3.2 can also be performed along the lines of Ref. 7. However, the latter paper involves a non-explicit arithmetic condition on the frequency vector, which complicates quantitative estimates needed for Remark 3.3. We believe that there is also some value in keeping the conditions on $(\lambda, \omega_1, \omega_2)$ in Preposition 3.2 purely arithmetic, even though they undergo an additional non-arithmetic parameter removal in Step 4 of Sec. V. We would like to thank the anonymous referee whose suggestions lead to this remark.

IV. THE LYAPUNOV–SCHMIDT DECOMPOSITION

Recall that we have $\mathbf{n} = (n_1, n_2, n_3) = (\mathbf{k}, j)$. Write u using the Fourier series (1.5). Let $\bar{u}$ be the complex conjugate of u, and

$$\bar{u}(t,x) = \bar{u}(t,x) = \sum_{\mathbf{n}\in\mathbb{Z}^3} \bar{\hat{u}}(\mathbf{n}) e^{-i(\mathbf{k}\cdot\omega)t} e^{-ijx\lambda} = \sum_{\mathbf{n}\in\mathbb{Z}^3} \bar{\hat{u}}(-\mathbf{n}) e^{i(\mathbf{k}\cdot\omega)t} e^{ijx\lambda}. \tag{4.1}$$

In other words, $\hat{v}(\mathbf{n}) = \bar{\hat{u}}(-\mathbf{n})$. *The function v is only introduced for convenience of notation, in order to avoid multiple stacked conjugation symbols. In the sequel, we will call v and u conjugate.*

A. The functional *F*

Using (1.5) and (4.1), we can rewrite the original NLS Eq. (1.1) in the Fourier representation:

$$(\mathbf{k}\cdot\omega + \delta^{\varkappa}\hat{V}(\mathbf{n}) + 2\cos(j\lambda))\,\hat{u}(\mathbf{n}) + \delta^{2p}((\hat{u} * \hat{v})^{*p} * \hat{u})(\mathbf{n}) = 0, \tag{4.2}$$

$$(-\mathbf{k}\cdot\omega + \delta^{\varkappa}\hat{V}(-\mathbf{n}) + 2\cos(j\lambda))\,\hat{v}(\mathbf{n}) + \delta^{2p}((\hat{u} * \hat{v})^{*p} * \hat{v})(\mathbf{n}) = 0, \tag{4.3}$$

where we recall the definition of $\hat{V}(\mathbf{n})$in (1.6):

$$\hat{V}(\mathbf{n}) = m_1 \mathbf{1}_{\{-\mathbf{e}_1\}}(\mathbf{k})\mathbf{1}_{\{h_1\}}(j) + m_2 \mathbf{1}_{\{-\mathbf{e}_2\}}(\mathbf{k})\mathbf{1}_{\{h_2\}}(j), \quad \mathbf{n} = (\mathbf{k}, j) \in \mathbb{Z}^3.$$

The symbol $*$ denotes the standard convolution:

$$(A * B)(\mathbf{n}) = \sum_{\mathbf{n}'\in\mathbb{Z}^3} A(\mathbf{n} - \mathbf{n}')B(\mathbf{n}'),$$

and $(\hat{u} * \hat{v})^{*p}$ denotes the pth convolution power of $\hat{u} * \hat{v}$.

The expression in the left hand side of (4.2), as well as (4.3), is a function on $\mathbb{Z}^3$ (in the variable $\mathbf{n}$). Let us combine the left hand sides of (4.2) and (4.3) into a single function on $\mathbb{Z}^3_{\mathrm{pm}}$, and denote the resulting function by $F[\hat{u}, \hat{v}]$. In some calculations, it will be convenient to consider $\hat{u}$ and $\hat{v}$ as independent variables, without explicitly using the fact that they are conjugate to each other.

Let $\ell_0(\mathbb{Z}^3)$ denote the space of all finitely supported functions on $\mathbb{Z}^3$. We can consider $(\hat{u}, \hat{v}) \mapsto F[\hat{u}, \hat{v}]$as a non-linear map from $\ell_0(\mathbb{Z}^3) \times \ell_0(\mathbb{Z}^3) = \ell_0(\mathbb{Z}^3_{\mathrm{pm}})$ into itself. Its derivative at the point $(\hat{u}, \hat{v})$is a linear map of the same type of the form

$$F'[\hat{u}, \hat{v}] = D + \delta^{\varkappa} V + \delta^{2p} H[\hat{u}, \hat{v}],$$

where

$$D = \begin{pmatrix} \operatorname{diag}\{\mathbf{k}\cdot\omega + 2\cos(j\lambda)\} & 0 \\ 0 & \operatorname{diag}\{-\mathbf{k}\cdot\omega + 2\cos(j\lambda)\} \end{pmatrix}, \quad \mathbf{n} \in \mathbb{Z}^3, \tag{4.4}$$

and

$$V = \begin{pmatrix} \operatorname{diag}\{\hat{V}(\mathbf{n})\} & 0 \\ 0 & \operatorname{diag}\{\hat{V}(-\mathbf{n})\} \end{pmatrix}$$

are multiplication operators independent of $\hat{u}$, $\hat{v}$, and

$$H = \begin{pmatrix} (p+1)(\hat{u} * \hat{v})^{*p} * & p(\hat{u} * \hat{v})^{*p-1} * \hat{u} * \hat{u} * \\ p(\hat{u} * \hat{v})^{*p-1} * \hat{v} * \hat{v} * & (p+1)(\hat{u} * \hat{v})^{*p} * \end{pmatrix} \tag{4.5}$$

is a matrix of convolution-type operator. For example, $(p+1)(\hat{u} * \hat{v})^{*p} *$ denotes an operator on $\ell_0(\mathbb{Z}^3)$ whose action on an element $w \in \ell_0(\mathbb{Z}^3)$ is given by $(p+1)(\hat{u} * \hat{v})^{*p} * w$.

Note that the non-linear part of F in the original coordinate representation (before Fourier transform) is simply the map

$$f:\mathbb{R}^2\to\mathbb{R}^2,\qquad f(u,v)=\delta^{2p}(uv)^p\begin{pmatrix}u\\ v\end{pmatrix},$$

whose differential is the matrix

$$\delta^{2p}\begin{pmatrix}(p+1)u^pv^p & pu^{p+1}v^{p-1}\\ pu^{p-1}v^{p+1} & (p+1)u^pv^p\end{pmatrix}.$$

We would like to estimate the difference between the increment $f(u+u_{\text{corr}},v+v_{\text{corr}})-f(u,v)$ and its linearization. Clearly, the linear part will completely cancel in this estimate. We can express the difference as follows:

$$\begin{aligned}(u+u_{\text{corr}})^p(v+v_{\text{corr}})^p\begin{pmatrix}u+u_{\text{corr}}\\ v+v_{\text{corr}}\end{pmatrix}-u^pv^p\begin{pmatrix}u\\ v\end{pmatrix}-\begin{pmatrix}(p+1)u^pv^p & pu^{p+1}v^{p-1}\\ pu^{p-1}v^{p+1} & (p+1)u^pv^p\end{pmatrix}\begin{pmatrix}u_{\text{corr}}\\ v_{\text{corr}}\end{pmatrix}\\ =u_{\text{corr}}^2g_1(u,v,u_{\text{corr}},v_{\text{corr}})+v_{\text{corr}}^2g_2(u,v,u_{\text{corr}},v_{\text{corr}})+u_{\text{corr}}v_{\text{corr}}g_3(u,v,u_{\text{corr}},v_{\text{corr}}).\end{aligned}\tag{4.6}$$

One can easily check that g_1, g_2, g_3 are polynomials of degrees at most, say, $10p$ with coefficients bounded by, say, 9^{10p}. As a consequence,

$$\|g_j\|_{C^1[-B,B]^4}\le 10^{10p}(10+B)^{10p},\quad j=1,2,3.$$

In other words, if one assumes that u, v, u_{corr}, v_{corr} are restricted to a bounded domain, then the difference between the increment of f and its linear approximation is, as expected, quadratic in u_{corr} and v_{corr}. Moreover, if u and v additionally depend on a parameter, this estimate can also be differentiated (possibly multiple times) with respect to that parameter, due to smoothness of the right hand side of (4.6). Naturally, the estimate is only meaningful if u_{corr} and v_{corr} are small. Combining with the Fourier representation, we arrive to the following proposition for the original functional F.

Proposition 4.1. Assume that $\|\hat u\|_1,\|\hat v\|_1,\|\hat u_{\text{corr}}\|_1,\|\hat v_{\text{corr}}\|_1\le B$, where $\|\cdot\|_1$ denotes the norm in $\ell^1(\mathbb{Z}^3)$. Assume that $\hat u$ and $\hat v$ are conjugate to each other. Then

$$\left\|F[\hat u+\hat u_{\text{corr}},\hat v+\hat v_{\text{corr}}]-F[\hat u,\hat v]-F'[\hat u,\hat v]\begin{pmatrix}\hat u_{\text{corr}}\\ \hat v_{\text{corr}}\end{pmatrix}\right\|_1\le\delta^{2p}10^{11p}(10+B)^{10p}\|\hat u_{\text{corr}}\|_1^2.\tag{4.7}$$

Suppose, in addition, that $\hat u$, $\hat v$, $\hat u_{\text{corr}}$, $\hat v_{\text{corr}}$ depend on several parameters. Let ∇ denote the derivatives with respect to the above parameters. Then one has

$$\begin{aligned}\left\|\nabla\left(F[\hat u+\hat u_{\text{corr}},\hat v+\hat v_{\text{corr}}]-F[\hat u,\hat v]-F'[\hat u,\hat v]\begin{pmatrix}\hat u_{\text{corr}}\\ \hat v_{\text{corr}}\end{pmatrix}\right)\right\|_1\\ \le\delta^{2p}10^{11p}(10+B)^{10p}\|\hat u_{\text{corr}}\|_1(\|\hat u_{\text{corr}}\|_1+\|\nabla\hat u_{\text{corr}}\|_1).\end{aligned}\tag{4.8}$$

Remark 4.2. One can also restate the above bounds in the form of the following convolution-type relation:

$$\begin{aligned}F[\hat u+\hat u_{\text{corr}},\hat v+\hat v_{\text{corr}}]-F[\hat u,\hat v]-F'[\hat u,\hat v]*\begin{pmatrix}\hat u_{\text{corr}}\\ \hat v_{\text{corr}}\end{pmatrix}\\ =\delta^{2p}\hat u_{\text{corr}}*\hat u_{\text{corr}}*h_1(\hat u,\hat v,\hat u_{\text{corr}},\hat v_{\text{corr}})+\delta^{2p}\hat u_{\text{corr}}*\hat v_{\text{corr}}*h_2(\hat u,\hat v,\hat u_{\text{corr}},\hat v_{\text{corr}})\\ +\delta^{2p}\hat v_{\text{corr}}*\hat v_{\text{corr}}*h_3(\hat u,\hat v,\hat u_{\text{corr}},\hat v_{\text{corr}}),\end{aligned}$$

where h_1, h_2, h_3, similarly to the above, are convolution polynomials of degrees at most $10p$ with coefficients bounded by, say, 9^{10p}.

B. The Lyapunov–Schmidt decomposition

Let

$$S_+ = \{(-\mathbf{e}_1, h_1), (-\mathbf{e}_2, h_2)\}, \qquad S_- = -S_+ = \{(\mathbf{e}_1, -h_1), (\mathbf{e}_2, -h_2)\},$$

and

$$S := (S_+ \times \{+\}) \cup (S_- \times \{-\}) \subset \mathbb{Z}^3_{\text{pm}}. \tag{4.9}$$

The Lyapunov–Schmidt decomposition is a decomposition of the system of Eqs. (4.2) and (4.3) $F[\hat{u}, \hat{v}] = 0$ into two systems, by restricting it into S and $\mathbb{Z}^3_{\text{pm}}\backslash S$, respectively. The first system

$$F[\hat{u}, \hat{v}]|_{\mathbb{Z}^3_{\text{pm}}\backslash S} = F[\hat{u}, \hat{v}]|_{S^c} = 0, \tag{4.10}$$

namely, for any $\mathbf{n} \in \mathbb{Z}^3\backslash S_+$,

$$(\mathbf{k}\cdot\omega + 2\cos(j\lambda))\hat{u}(\mathbf{n}) + \delta^{2p}((\hat{u} * \hat{v})^{*p} * \hat{u})(\mathbf{n}) = 0, \tag{4.11}$$

and for any $\mathbf{n} \in \mathbb{Z}^3\backslash S_-$,

$$(-\mathbf{k}\cdot\omega + 2\cos(j\lambda))\hat{v}(\mathbf{n}) + \delta^{2p}((\hat{u} * \hat{v})^{*p} * \hat{v})(\mathbf{n}) = 0 \tag{4.12}$$

is usually referred to as *P-equations*. The restriction into S is called the *Q-equations*:

$$F[\hat{u}, \hat{v}]|_S = 0. \tag{4.13}$$

We will impose $\hat{u}(-\mathbf{e}_k, h_k) = u_k$ for $k = 1, 2$. With this assumption and using the conjugacy relation between $\hat{u}$ and $\hat{v}$, we can see that the Q-equations reduce to the following system:

$$\omega_k = 2\cos(h_k\lambda) + m_k\delta^{\varkappa} + \frac{\delta^{2p}}{u_k}((\hat{u} * \hat{v})^{*p} * \hat{u})(-\mathbf{e}_k, h_k), \quad k = 1, 2. \tag{4.14}$$

We remark that only the Q-equations depend *explicitly* on m, and the P-equations (4.11) and (4.12) do not. The latter enables us to solve the P-equations in the (ω, λ) variables and is a simplifying feature.

C. The P-equations and the multi-scale Newton scheme

We use a Newton scheme to solve the P-equations. Fix a conjugate pair (u, v), and let T_u be the linearized operator on $\ell^2(\mathbb{Z}^3_{\text{pm}}\backslash S)$:

$$T_u = F'[\hat{u}, \hat{v}]|_{S^c} = (D + \delta^{2p}H[\hat{u}, \hat{v}])|_{S^c}. \tag{4.15}$$

Recall the formal Newton scheme:

$$\begin{pmatrix} \hat{u}_{\text{corr}} \\ \hat{v}_{\text{corr}} \end{pmatrix} = -(T_u)^{-1}F[\hat{u}, \hat{v}]|_{S^c}. \tag{4.16}$$

Here, $F[\hat{u}, \hat{v}]|_{S^c}$ is an element of $\ell^2(S^c)$, and T_u is an operator acting on the same space. Therefore, both $\hat{u}_{\text{corr}}$ and $\hat{v}_{\text{corr}}$ are elements of $\ell^2(\mathbb{Z}^3\backslash S_+)$ and $\ell^2(\mathbb{Z}^3\backslash S_-)$, respectively. One can easily check that $\hat{u}_{\text{corr}}$ and $\hat{v}_{\text{corr}}$ are conjugate, that is,

$$\hat{u}_{\text{corr}}(\mathbf{n}) = \bar{\hat{v}}_{\text{corr}}(-\mathbf{n}).$$

We adopt a *multiscale Newton scheme* as follows. Start from the zero iteration

$$\hat{u}^{(0)}(\mathbf{n}) = \begin{cases} u_1, & \mathbf{n} = (-\mathbf{e}_1, h_1); \\ u_2, & \mathbf{n} = (-\mathbf{e}_2, h_2); \\ 0, & \mathbf{n} \in \mathbb{Z}^3\backslash S_+. \end{cases}$$

Consider a sequence of scales $N^{(0)} \le N^{(1)} \le \ldots$. At the iteration step $(r+1)$, consider

$$T_u|_N = T_u|_{[-N,N]^3_{\text{pm}}\backslash S}.$$

Journal of
Mathematical Physics
ARTICLE
pubs.aip.org/aip/jmp

Note that, since T_u is already restricted to $\mathbb{Z}^3_{\mathrm{pm}}\backslash S$ in the original definition, this is the same notation as in Sec. II. Define the $(r+1)$-th correction to be:

$$\begin{pmatrix} \hat{u}^{(r+1)}_{\mathrm{corr}} \\ \hat{v}^{(r+1)}_{\mathrm{corr}} \end{pmatrix} = -(T_{u^{(r)}}|_{N^{(r+1)}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{N^{(r+1)},S^c} \tag{4.17}$$

and obtain the next iterations by

$$\hat{u}^{(r+1)} := \hat{u}^{(r)} + \hat{u}^{(r)}_{\mathrm{corr}}; \qquad \hat{v}^{(r+1)} := \hat{v}^{(r)} + \hat{v}^{(r)}_{\mathrm{corr}}; \qquad \hat{v}^{(r+1)}(\mathbf{n}) = \bar{\hat{u}}^{(r+1)}(-\mathbf{n}), \tag{4.18}$$

where one can easily check that the last two equalities imply one another.

We will use the terminology *finite volume* for the above iterated equations and their solutions $u^{(r)}$, $v^{(r)}$, $\hat{u}^{(r)}$, $\hat{v}^{(r)}$, and *infinite volume* for the original non-linear equations and their solutions.

D. The Q-equations

At each iteration step, the values of $\hat{u}^{(r)}$and $\hat{v}^{(r)}$on S do not change, and are always equal to those of $\hat{u}^{(0)}$, $\hat{v}^{(0)}$. As a consequence, the iteration process for the P-equations is consistent. In other words, the outcome of the Newton iterations for the P-equations is sufficient to determine the input for the next iterations. As a consequence, finite volume versions of the Q equations can be essentially ignored. If the above Newton iterations for the P equations converge to an infinite volume solution, one can interpret the Q-equations (4.14) as the final relation between $\lambda, m_1, m_2, \omega_1, \omega_2$. During the iteration, we will require $(\lambda, \omega_1, \omega_2)$ to belong avoid certain "bad" subsets, from the application of Proposition 2.3 ("steep planes lemma"). In Sec. VI, we will restate this as a condition completely in terms of (λ, m_1, m_2), as required by the statement of the main result, Theorem 1.1.

V. INDUCTIVE PROCEDURE FOR THE P-EQUATIONS

In this section, we will implement the multi-scale Newton scheme described above, for the P-equations. This is the most technical section of the paper.

As discussed above, we start from the initial iteration

$$\hat{u}^{(0)}(\mathbf{n}) = \begin{cases} u_1, & \mathbf{n} = (-\mathbf{e}_1, h_1); \\ u_2, & \mathbf{n} = (-\mathbf{e}_2, h_2); \\ 0, & \mathbf{n} \in \mathbb{Z}^3\backslash S_+, \end{cases} \tag{5.1}$$

and $\hat{v}^{(0)}$being the conjugate of $\hat{u}^{(0)}$. For a sequence of scales $N^{(r)} = A^{r+r_0}$, define

$$\begin{pmatrix} \hat{u}^{(r+1)}_{\mathrm{corr}} \\ \hat{v}^{(r+1)}_{\mathrm{corr}} \end{pmatrix} = -(T_{u^{(r)}}|_{N^{(r+1)}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{N^{(r+1)},S^c},$$

$$\hat{u}^{(r+1)} := \hat{u}^{(r)} + \hat{u}^{(r+1)}_{\mathrm{corr}}; \qquad \hat{v}^{(r+1)} := \hat{v}^{(r)} + \hat{v}^{(r+1)}_{\mathrm{corr}}; \qquad \hat{v}^{(r+1)}(\mathbf{n}) = \overline{\hat{u}^{(r+1)}(-\mathbf{n})}.$$

The goal of this section is to establish several inductive estimates on $\hat{u}^{(r)}$and $\hat{v}^{(r)}$, each of which will be supported on $[-A^{r+r_0}, A^{r+r_0}]$. Recall

$$\mathrm{DCR}(\tau, N) = \left\{(\lambda, \omega_1, \omega_2) \in (1,2) \times (-3,3)^2 : |\omega_2| \geq \tau,\ \frac{\omega_1}{\omega_2} \in \mathrm{DC}_1(\tau; N),\ \lambda \in \mathrm{DC}_1(\tau; N)\right\}.$$

Let

$$I_0' := \{(\omega, \lambda) \in \mathrm{DCR}(\tau, A^{10r_0}) : |D(\mathbf{n})| \geq \delta_{\mathrm{res}} \quad \text{for all } \mathbf{n} \in [-A^{2r_0}, A^{2r_0}]^3\}. \tag{5.2}$$

Let I_0 be the union of all dyadic cubes contained in I_0' with side at least δ^2_{res}. It is easy to see that, say

$$\mathrm{Leb}_3\big(\mathrm{DCR}(\tau, A^{10r_0})\backslash I_0\big) \leq A^{100r_0}\delta_{\mathrm{res}}. \tag{5.3}$$

We will show that there exist subsets $I_0 = I_1 = I_2 = I_3 = \cdots = I_{r_0} \supset I_{r_0+1} \supset \ldots$ satisfying:

(i1) I_r can be split into a disjoint union of $\exp\{((r+r_0)\log A)^{C_{\text{abs},1}}\}$ closed rectangles $I_{r,j}$, each of which is of diameter at most $\exp\{-((r+r_0)\log A)^{C_{\text{abs},2}}\}$ and has edges parallel to the coordinate axes.

(i2) The rectangles in $I_{r,j}$ are nested: for each $I_{r,j}$ there exists $I_{r-1,j'}$ such that $I_{r,j} \subset I_{r-1,j'}$.

(i3) We have $\text{Leb}_3(I_j \backslash I_{j+1}) \le e^{-C_{\text{abs},6}(r+r_0)\log A}$.

For $(\omega,\lambda) \in I_r$, we will show:

(sol1) $\hat{u}^{(r)}$ and $\hat{v}^{(r)}$ are supported on $[-A^{r+r_0}, A^{r+r_0}]^3$ for $r \ge 1$.

(sol2) $|\hat{u}^{(r)}(\mathbf{n})| \le \left(1 - 2^{-(r+1)}\right)e^{-\mu_r|\mathbf{n}|}$, for $|\mathbf{n}| \in \mathbb{Z}^3\backslash S_+$.

(sol3) $|\hat{u}^{(r+1)}_{\text{corr}}(\mathbf{n})| \le \delta_{r+1}e^{-\mu_r^{(3)}|\mathbf{n}|} \le \delta_{r+1}e^{-\mu_{r+1}|\mathbf{n}|}$, see (5.6) for details on notation.

(sol4) $|F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c}(\mathbf{n})| \le \kappa_r e^{-\mu_r|\mathbf{n}|}$.

(sol5) Let

$$G_r := \begin{cases} \exp\left(C'_{\text{abs},3}(r+r_0)\log A\right), & 0 \le r \le r_0; \\ \exp\left(((r+r_0)\log A)^{C_{\text{abs},3}}\right), & r > r_0. \end{cases}$$

$$D_r := \begin{cases} 1, & 0 \le r \le r_0; \\ ((r+r_0)\log A)^{C_{\text{abs},4}}, & r > r_0. \end{cases}$$

We have the following Green's function bounds:

$$\|(T_{u^{(r)}}|_{A^{r+r_0}})^{-1}\| \le \delta_{\text{res}}^{-1}G_r, \tag{5.4}$$

and for $|\mathbf{n} - \mathbf{n}'| \ge D_r$, we have

$$|(T_{u^{(r)}}|_{A^{r+r_0}})^{-1}(\mathbf{n},\mathbf{n}')| \le \delta^{p-1/2}e^{-\mu_r|\mathbf{n}-\mathbf{n}'|}. \tag{5.5}$$

(sol6) The matrix elements of the operator $T_{u^{(r)}}$ satisfy

$$|T_{u^{(r)}}(\mathbf{n},\mathbf{n}')| \le \delta^{2p-1}e^{-\mu_r^{(1)}|\mathbf{n}-\mathbf{n}'|}, \quad \mathbf{n} \ne \mathbf{n}'.$$

$$\|\nabla T_{u^{(r)}}(\mathbf{n},\mathbf{n}')\| \le \delta^{2p-1}e^{-\mu_r^{(1)}|\mathbf{n}-\mathbf{n}'|}, \quad \mathbf{n} \ne \mathbf{n}'.$$

The operator $T_{u^{(r)}}$ satisfies the conclusion of Proposition 3.2 at the scale

$$M = M(r) = \lfloor((r+r_0)\log A)^{C_L}\rfloor,$$

where C_L is a large absolute constant (say, $C_L = 300$), and $\gamma'_{M(r)} = \mu_r - \frac{1}{50(r+r_0)^{99}}$.

In addition to the above, we will establish the following regularity properties.

(reg1) The map $(\omega,\lambda) \mapsto \hat{u}^{(r)}$ can be extended to a map of the class

$$C^1((1,2)\times(-3,3);\ell^1([-A^{r+r_0}, A^{r+r_0}]^3)).$$

(reg2) On $(-3,3)^2 \times (1,2)$, we have $|\nabla\hat{u}^{(r)}(\mathbf{n})| \le (1 - 2^{-r-1})\delta^{1+3p/8}e^{-\mu_r|\mathbf{n}|}$.

(reg3) For $(\omega,\lambda) \in I_r$, we have $|\nabla\hat{u}^{(r+1)}_{\text{corr}}(\mathbf{n})| \le \delta_{r+1}e^{-\mu_r^{(4)}|\mathbf{n}|}$, where ∇ denotes the gradient with respect to $\omega_1, \omega_2, \lambda$.

(reg4) For $(\omega,\lambda) \in I_r$, we have $|\nabla F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c}(\mathbf{n})| \le \kappa_r e^{-\mu_r|\mathbf{n}|}$.

(reg5) For $(\omega,\lambda) \in I_{r,j}$, the components of $u^{(r)}(\omega,\lambda)$ are rational functions of $\omega_1, \omega_2, \cos\lambda$ of degree at most $e^{((r+r_0)\log A)^2}$.

(reg6) The small parameters in the above bounds satisfy the estimate

$$\kappa_r = \delta^{p(3/2)^r} = e^{-pC_{\text{abs},7}r_0(3/2)^r\log A}, \qquad \delta_{r+1} = G_r^2\delta_{\text{res}}^{-1}\kappa_r.$$

We will also have

$$e^{-C_{\text{abs},8}r_0\log A} \le \delta \le e^{-C_{\text{abs},7}r_0\log A} \le \delta_{\text{res}} \le e^{-C_{\text{abs},9}r_0\log A}.$$

(reg7) The bounds in (sol5) for $r \ge r_0$ are stable under perturbations of $u^{(r)}$ of size e^{-M^2} in the ℓ^∞ norm (see the notation of Steps 3 and 4). Since $\delta_r \ll e^{-M^2}$, they are also stable under perturbations of size $10\delta_r$. For $r \le r_0$, the bounds are stable under perturbations of size δ_{res}^2. The same is true for perturbations of the parameters ω, λ.

A. Some conventions about the estimates

The below estimates are quite technical and involve multiple r-dependent parameters. We will try to introduce some conventions in order to have more structure and lesser chance of errors.

(1) The exponential decay rate of several involved quantities is denoted by μ_r. In the transition from r to $r+1$, we will have to introduce several auxiliary decay rates:

$$\mu_r^{(0)} = \mu_r, \qquad \mu_r^{(j+1)} = \mu_r^{(j)} - \frac{1}{100(r+r_0)^{99}}, \quad j = 1,2,\ldots,4. \tag{5.6}$$

In the end, we will take $\mu_{r+1} := \mu_r^{(5)}$. Each sub-step will slightly decrease the rate of decay. This, together with the choice of small δ, will guarantee

$$\mu_r \in (\gamma/2, \gamma)$$

for some $3 \le \gamma \le 10$ and all $r \ge 0$.

(2) For a given scale r, in most of the calculations one can assume the constant prefactor to be 1 in (sol2) and (reg2), and then subsequently use (sol2) and (reg2) to confirm that one can use $(1-2^{-r-2})$ for the scale $r+1$.

B. The induction step: Plan

The induction procedure involves multiple estimates related to each other. Here, we outline the order in which we prove them. We assume that the induction assumptions are obtained at step r.

(1) Check all estimates at the initial scale for $u^{(0)}$ (for small δ using direct perturbation arguments).
(2) Using the induction assumptions on $u^{(r)}$, construct the operator $T_{u^{(r)}}$ and check the first two estimates in (sol6). The off-diagonal decay rate for matrix elements of $T_{u^{(r)}}$ will be $\mu_r^{(1)}$.
(3) Check the assumptions of the large deviation theorem (Proposition 3.2) on the relevant scales, thus confirming the remaining claims in (sol6). The decay rate in the output of LDT will be $\mu_r^{(2)}$. This step will only be required only for $r \ge r_0$ and will use the corresponding scale-dependent Diophantine condition. Quantitative aspects of the Diophantine condition will be discussed during Step 9.
(4) Obtain estimates (sol5) of $\left(T_{u^{(r)}}|_{A^{r+r_0+1}}\right)^{-1}$, using the estimates on $T_{u^{(r)}}|_{N^{(r)}}$ and the large deviation theorem. The exponential decay rate in this version of (sol5) will be $\mu_r^{(3)}$. During this process, a small subset of the parameters (ω,λ) will be removed through an application of the steep planes lemma, in addition to the Diophantine condition. This removal of the parameters (ω,λ) is not optimal for some smaller scales (depending on δ_{res}), compared to a direct perturbation argument. We will also determine the range of scales for which a conclusion of this step can be achieved by a direct perturbation approach with better measure estimates, and combine them into one estimate which will have power law decay as $\delta \to 0$.
(5) Construct $\hat{u}_{\mathrm{corr}}^{(r+1)}$ and estimate it using the previous step and the induction assumptions on $F[\hat{u}^{(r)}, \hat{v}^{(r)}]$, obtaining (sol3) and (reg3). The decay rate in these estimates will be $\mu_r^{(4)}$. This also implies (sol5) at the scale $r+1$ with decay rate $\mu_r^{(3)}$.
(6) Estimate $F[\hat{u}^{(r+1)}, \hat{v}^{(r+1)}]|_{N^{(r+1)},S^c}$ using the results of the previous steps. The decay rate will be $\mu_r^{(5)}$. At this step, we will take advantage of the Newton approximation argument in order to get a quadratic improvement. The rate $\mu_r^{(5)}$ will be the weakest possible among those obtained at step r, and we declare $\mu_{r+1} := \mu_r^{(5)}$.
(7) Summarize the inductive assumptions on smallness of various constants and justify (reg6).
(8) Use the definition $\hat{u}^{(r+1)} = \hat{u}^{(r)} + \hat{u}_{\mathrm{corr}}^{(r+1)}$ to estimate $\hat{u}^{(r+1)}$ and establish (sol3), (reg3) with the decay rate $\mu_r^{(4)}$.
(9) Complete the semi-algebraic bounds (degree and rectangular structure) and extensions to the full parameter space.

C. Some bounds on convolution-type operators

We will need some elementary bounds on convolutions and convolution-type operators.

Lemma 5.1. Suppose that

$$0 \le f_j(\mathbf{n}) \le e^{-\mu|\mathbf{n}|}, \quad 1 \le j \le p.$$

Then

$$f(\mathbf{n}) := (f_1 * f_2 * \cdots * f_p)(\mathbf{n}) \le 2^{6p-1} e^{\mu} \left(\frac{|\mathbf{n}|^{3p}}{\mu} + \frac{(3p)!}{\mu^{3p+1}} \right) e^{-\mu|\mathbf{n}|}.$$

For $R \ge 1$, we also have

$$f(\mathbf{n}) \le 2^{6p-1} e^{\mu} \left(\frac{(10pR)^{6p}}{\mu} + \frac{(3p)!}{\mu^{3p+1}} \right) e^{-(\mu - R^{-1})|\mathbf{n}|}.$$

Proof. We have

$$\begin{aligned} f(\mathbf{n}) &\le \sum_{k=|\mathbf{n}|}^{+\infty} e^{-\mu k} \#\{(\mathbf{n}_1, \ldots, \mathbf{n}_p) : |\mathbf{n}_1| + \cdots + |\mathbf{n}_p| = k\} \le 2^{3p} \sum_{k=|\mathbf{n}|}^{+\infty} k^{3p} e^{-\mu k} \\ &\le 2^{3p} e^{\mu} \int_{|\mathbf{n}|}^{+\infty} s^{3p} e^{-\mu s}\, ds = 2^{3p} e^{\mu} e^{-\mu|\mathbf{n}|} \int_0^{+\infty} (s + |\mathbf{n}|)^{3p} e^{-\mu s}\, ds \\ &\le 2^{6p-1} e^{\mu} e^{-\mu|\mathbf{n}|} \int_0^{+\infty} (s^{3p} + |\mathbf{n}|^{3p}) e^{-\mu s}\, ds \le 2^{6p-1} e^{\mu} \left(\frac{|\mathbf{n}|^{3p}}{\mu} + \frac{(3p)!}{\mu^{3p+1}} \right) e^{-\mu|\mathbf{n}|}. \end{aligned}$$

The second estimate follows from the fact that, for $|\mathbf{n}| \ge 100p^2R^2$, the extra factor $e^{-R^{-1}|\mathbf{n}|}$ will absorb $|\mathbf{n}|^{3p}$. □

We will also need another more specific convolution-type estimate.

Corollary 5.2. Suppose that $g_j, f : \mathbb{Z}^3 \to \mathbb{R}_+$ and for $j = 1, 2, \ldots, p-1$ we have

$$g_j(\mathbf{n}) \le \begin{cases} e^{-\mu|\mathbf{n}|}, & |\mathbf{n}| \ge D; \\ G, & |\mathbf{n}| < D. \end{cases}$$

with $G > 1$. Suppose also that $f(\mathbf{n}) \le e^{-\mu|\mathbf{n}|}$ for all $\mathbf{n} \in \mathbb{Z}^3$. Then

$$(g_1 * g_2 * \cdots * g_{p-1} * f)(\mathbf{n}) \le 2^{7p-1} e^{\mu} \left(\frac{|\mathbf{n}|^{3p}}{\mu} + \frac{(3p)!}{\mu^{3p+1}} + (3p)! \right) e^{-\mu|\mathbf{n}|} + G^{p-1} (4D)^{3p} e^{2(p-1)\mu D} e^{-\mu|\mathbf{n}|}.$$

For $R \ge 1$, we also have

$$\begin{aligned} &(g_1 * g_2 * \cdots * g_{p-1} * f)(\mathbf{n}) \le \\ &\quad \le 2^{7p-1} e^{\mu} \left(\frac{(10pR)^{6p}}{\mu} + \frac{(3p)!}{\mu^{3p+1}} + (3p)! \right) e^{-(\mu - R^{-1})|\mathbf{n}|} + G^{p-1} (4D)^{3p} e^{2(p-1)\mu D} e^{-\mu|\mathbf{n}|}. \end{aligned}$$

Proof. Since $g_i(\mathbf{n}) \le e^{-\mu|\mathbf{n}|} \mathbf{1}_{|\mathbf{n}| \ge D} + G \mathbf{1}_{|\mathbf{n}| < D}$, one can expand the left hand side into 2^{p-1} terms, by considering two possible choices for each g_j. By adding an extra factor 2^p and after a small weakening of the estimates to ensure monotonicity in p, one can restrict themselves to two extreme cases, with either first or second choice made for all g_j at the same time. We leave the details to the reader. □

Remark 5.3. The bounds in Lemma 5.1 and Corollary 5.2 are of pointwise type. As a consequence, one can replace convolution-type operators by more general operators with matrix elements satisfying $|g_j(\mathbf{m}, \mathbf{n})| \le g_j(\mathbf{n} - \mathbf{m})$.

D. Step 1: The initial scale and some perturbative estimates

While we will construct $u^{(1)}, u^{(2)}, \ldots$ inductively, it is convenient to obtain bounds (in particular Step 4) "in advance," since the bounds obtained as a results of the inductive procedure would be too weak on small scales. Suppose that $(\omega, \lambda) \in I_0$. Recall that, from (sol6), we have (after simplifying some notation)

$$T := T_{u^{(r)}} = (D + \delta^{2p} H[\hat{u}^{(r)}, \hat{v}^{(r)}])|_{S^c} =: D + \delta^{2p-1} H, \qquad |H(\mathbf{n}_1, \mathbf{n}_2)| \le e^{-\mu_r|\mathbf{n}_1 - \mathbf{n}_2|},$$

For $0 \le r \le r_0$, we have $\|D^{-1}\| \le \delta_{\mathrm{res}}^{-1}$. We assume that T is restricted to an appropriate box $[-A^{r+r_0}, A^{r+r_0}]$, although it does not affect the argument much (however, a larger box will require stronger assumptions to maintatin the bound $\|D^{-1}\| \le \delta_{\mathrm{res}}^{-1}$). Assume that δ is small enough so that $\|\delta^{2p-1}H\| \le \frac{1}{2}\delta_{\mathrm{res}}$. One has the Neumann series expansion:

$$T^{-1} = D^{-1} + \delta^{2p-1}D^{-1}HD^{-1} + \delta^{4p-2}D^{-1}HD^{-1}HD^{-1} + \cdots$$

Assume also that $|\delta| < A^{-100r_0p}$. Then the factors $\delta^{(2p-1)k}$ will absorb all factors of the form D^{-1} and all polynomial/combinatorial factors, leaving with exponential factors from convolutions of several copies of H. As a consequence, in the above range we have

$$\|T^{-1}\| \le \frac{1}{2}\delta_{\mathrm{res}}^{-1};$$

$$\|T^{-1}(\mathbf{m},\mathbf{n})\| \le \delta^{p-1/2}e^{-\mu_{r+1}|\mathbf{n}-\mathbf{m}|}, \quad \mathbf{n} \ne \mathbf{m}.$$

Summarizing the above, the estimates (sol5) are obtained in the range $0 \le r \le r_0$ assuming the estimates from (sol6).

E. Step 2: The operator $T_{u^{(r)}}$

We recall that

$$T_u = (D + \delta^{2p}H[\hat{u}, \hat{v}])|_{S^c},$$

where

$$D = \begin{pmatrix} \mathrm{diag}\ \{\mathbf{k}\cdot\omega + 2\ \cos(j\lambda)\} & 0 \\ 0 & \mathrm{diag}\ \{-\mathbf{k}\cdot\omega + 2\ \cos(j\lambda)\} \end{pmatrix}, \quad (\mathbf{k}, j) = \mathbf{n} \in \mathbb{Z}^3,$$

and H is a (2×2)-matrix of convolution-type operators:

$$H = \begin{pmatrix} (p+1)(\hat{u}*\hat{v})^{*p}* & p(\hat{u}*\hat{v})^{*p-1}*\hat{u}*\hat{u}* \\ p(\hat{u}*\hat{v})^{*p-1}*\hat{v}*\hat{v}* & (p+1)(\hat{u}*\hat{v})^{*p}* \end{pmatrix}.$$

Assume the choice of δ made at Step 1. Note that, since $\hat{u}$ contains the values of $\hat{u}$ on S_+ which are not necessarily small, they may spread to the translations of the points of S_+ after taking the convolution. Assume, in addition to the former, that *the choice of large A, r_0 absorbs these values*. For example one can choose them so that

$$\delta u_j \le e^{-100p|h_j|}, \quad j = 1, 2. \tag{5.7}$$

With the above choices, one can see that both bounds hold for $0 \le r \le r_0$, assuming (sol3) and (reg3). Indeed, polynomial factors from the convolution can be absorbed into the upper bound on δ.

For the larger scales, it will be convenient to prove a stronger bound

$$|T_{u^{(r)}}(\mathbf{n},\mathbf{n}')| \le \delta^{2p-1}(1-2^{-r-1})e^{-\mu_r^{(1)}|\mathbf{n}-\mathbf{n}'|}, \quad \mathbf{n} \ne \mathbf{n}'$$

inductively, by estimating the difference

$$T_{u^{(r-1)}}(\mathbf{n},\mathbf{n}') - T_{u^{(r)}}(\mathbf{n},\mathbf{n}'). \tag{5.8}$$

Indeed, the above difference can be expressed as a sum of $2p$ convolution-type operators, each of which contains $\hat{u}_{\mathrm{corr}}^{(r)}$or $\hat{v}_{\mathrm{corr}}^{(r)}$, and the remaining factors are of the form $\hat{u}^{(r-1)}$, $\hat{u}^{(r)}$, $\hat{v}^{(r-1)}$, $\hat{v}^{(r)}$. As an example, for the top left matrix element we have

$$\begin{aligned} &(p+1)(\hat{u}^{(r)}*\hat{v}^{(r)})^{*p} - (p+1)(\hat{u}^{(r-1)}*\hat{v}^{(r-1)})^{*p} \\ &\quad = (p+1)\left\{(\hat{u}^{(r-1)} + \hat{u}_{\mathrm{corr}}^{(r)})^{*p}*(\hat{v}^{(r-1)} + \hat{v}_{\mathrm{corr}}^{(r)})^{*p} - (\hat{u}^{(r-1)}*\hat{v}^{(r-1)})^{*p}\right\} \\ &\quad = \hat{u}_{\mathrm{corr}}^{(r)}*h_p^{(1)}(\hat{u}^{(r)}, \hat{v}^{(r)}, \hat{u}^{(r-1)}, \hat{v}^{(r-1)}) + \hat{v}_{\mathrm{corr}}^{(r)}*h_p^{(2)}(\hat{u}^{(r)}, \hat{v}^{(r)}, \hat{u}^{(r-1)}, \hat{v}^{(r-1)}), \end{aligned} \tag{5.9}$$

where $h_p^{(1,2)}$are convolution polynomials of degrees at most $2p-1$ in $\hat{u}^{(r)}$, $\hat{v}^{(r)}$, $\hat{u}^{(r-1)}$, $\hat{v}^{(r-1)}$(one has to take the value of the function at $|\mathbf{n}-\mathbf{n}'|$ to obtain the matrix element between $\mathbf{n}$ and $\mathbf{n}'$).

Recall that $u^{(r)}$ has decay rate $\mu^{(r)}$. Apply Lemma 5.1. After some simplifications, will have the right hand side of (5.9) will be bounded by

$$\delta_r\left(\frac{1000(r+r_0)p}{\mu}\right)^{C(p)} e^{-\mu_r^{(1)}|\mathbf{n}-\mathbf{n}'|} \le 2^{-r-1}e^{-\mu_r^{(1)}|\mathbf{n}-\mathbf{n}'|}.$$

Since $r \ge r_0$ and due to (reg6), the factor δ_r will absorb any power of $(r+r_0)$, thanks to the choice of r_0 or A being allowed to depend on p.

A similar calculation can be done to estimate the derivatives, assuming (reg2) and (reg3). We leave it to the reader.

Note that one can also estimate the off-diagonal entries of $T_{u^{(r)}}$ directly by combining Lemma 5.1 with (sol2). However, this will produce some extra polynomial factors that are cumbersome to deal with.

F. Step 3: Verification of the assumptions of the large deviation theorem

In this subsection, we will verify the assumptions of Proposition 3.2. A somewhat careful analysis is required, since the rate of decay of the off-diagonal entries changes from scale to scale, and the rate of the change decreases as r gets larger. At the induction step $r \ge r_0$, we will be applying Proposition 3.2 to the operator $T_{u^{(r)}}$ at the scale

$$M = M(r) := \lfloor(\log N)^{C_L}\rfloor = \lfloor((r+r_0+1)\log A)^{C_L}\rfloor \sim A^{r_1}, \tag{5.10}$$

where

$$r_1 \sim \frac{C_L\,\log\log N}{\log A} \sim C_L\frac{\log(r_0+r+1)+\log\log A}{\log A}$$

and C_L is a large constant (say, $C_L \ge 400$). Here, $\sim$ denotes the nearest possible value assuming r_1 is integer.

From Step 2, the off-diagonal entries of $T_{u^{(r)}}$ have the Töplitz decay rate is $\mu_r^{(1)}$. Since Proposition 3.2 is applied on the scale M, we would have $\gamma_M = \mu_r^{(1)}$, where M and r are related through (5.10). We would like the decay rate in the conclusion of 3.2 to be $\mu_r^{(2)}$, which corresponds to $\gamma'_M = \mu_r^{(2)}$. Let s be an integer such that $M(s)$ is closest possible to M^2. The assumptions of Proposition 3.2 require $\gamma_{M^2} \le \gamma'_M - M^{-10}$. On the other hand, the scale of the size close to M^2 is used on the step s, and therefor $\gamma_{M^2} \sim \mu_1^{(s)}$. In order for these inequalities to be consistent with each other, it is sufficient to enforce

$$\mu_s^{(1)} \le \mu_r^{(2)} - M^{-10}.$$

From the definitions of the above objects, it is sufficient to have

$$s \ge r+1; \qquad 100(r+r_0)^{99} \le ((r+r_0+1)\log A)^{2C_L}.$$

Clearly, both inequalities are achievable with, say, $C_L \ge 10$, $r_0 \ge (\log A)^2$, with margins sufficient to overcome rounding errors. Note that one does not need to be particularly careful about these relations on scales smaller than $M(r_0)$, since one does not need to match the exponents γ with μ, and can start with a stronger decay in the initial scale of Proposition 3.2 which will gradually decrease to $\mu_{r_0}^{(1)}$ (or better) as the scales reach $M(r_0)$. For example, the choice made in (5.7) is by far sufficient.

Let us also discuss the choice of small δ. In view of Remarks 3.3 and 7.6, we would need

$$|\delta| \le e^{-\tau^{-C_{\mathrm{abs,dio}}}}$$

or, equivalently,

$$\frac{1}{(\log(1/\delta))^{1/C_{\mathrm{abs,dio}}}} \le \tau.$$

In view of (reg6), this also converts into

$$C_{\mathrm{abs},7}r_0\,\log A \ge \tau^{-C_{\mathrm{abs,dio}}}.$$

This choice also achievable by choosing large r_0 and/or A.

This completes verification of the assumptions of Proposition 3.2. As a consequence, we obtained a subset $X''_M \subset \mathbb{R}\times[0,2\pi)$such that, for $(\theta,\varphi) \in \mathbb{R}\times[0,2\pi)\backslash X''_M$, we have

$$\|(T_{u^{(r)}}(\theta,\varphi)|_M)^{-1}\| \le e^{M^{\frac{6}{7}}};$$

for any $\mathbf{n},\mathbf{n}' \in [-M,M]^3$ with $|\mathbf{n}-\mathbf{n}| \ge M^{999/1000}$ we have

$$|(T_{u^{(r)}}(\theta,\varphi)|_M)^{-1}(\mathbf{n},\mathbf{n}')| \le e^{-\mu_r^{(2)}|\mathbf{n}-\mathbf{n}'|}.$$

We also have that, for every line segment $L \subset \mathbb{R} \times [0, 2\pi)$, we have

$$\mathrm{Leb}_1(L \cap X''_M) \le e^{-M^{1/300}}. \tag{5.11}$$

G. Step 4: Estimates of $(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}$

This step is the central part of the argument. We would like to advance (sol5) from the scale A^{r+r_0} to the scale $A^{r+r_0+1} = N^{(r+1)}$. At this stage, we do not change $T_{u^{(r)}}$ to $T_{u^{(r+1)}}$ just yet. This change will be performed once we obtain estimates for $\hat{u}^{(r+1)}_{\mathrm{corr}}$ in later steps.

We will also be assuming that $r \ge r_0$. While the induction steps will be carried over starting from $r = 0$, the estimates in (sol5) at small scales, as explained during Step 1, are obtained directly from perturbation arguments. They are also not sensitive to the change between $T_{u^{(r)}}$ and $T_{u^{(r+1)}}$.

Our goal is to apply Proposition 3.1, with the following data:

- The central box will be $[-A^{r+r_0}, A^{r+r_0}]$, and the operator will be $T_{u^{(r)}}$. This comes from the induction assumptions.
- The size of small boxes will be, as in the previous calculations,

$$M := \lfloor (\log N)^{C_L} \rfloor = \lfloor ((r + r_0 + 1) \log A)^{C_L} \rfloor \sim A^{r_1},$$

$$r_1 \sim \frac{C_L \log\log N}{\log A} \sim C_L \frac{\log(r_0 + r + 1) + \log\log A}{\log A}.$$

  Here, $\sim$ denotes the nearest possible value assuming r_1 is integer, and C_L is a large constant.
- We will construct a semi-algebraic subset of parameters (ω, λ) so that each M-box away from the center of the large box (as required for Proposition 3.1) is good for the operator $T_{u^{(r_2)}}$ with $r_2 = (\log A)^2 r_1$.
- Assuming (reg6) and $(\omega, \lambda) \in I_{r_2}$, we have

$$\|\hat{u}^{(r_2)} - \hat{u}^{(r)}\| \le 2\hat{u}^{(r_2)}_{\mathrm{corr}} \ll e^{-M^2},$$

  which implies the complete assumptions of Proposition 3.1 for the operators $T_{u^{(r)}}$ (in other words, the bounds for M-boxes will remain the same, up to a factor of 2, when we replace r_2 by r).

At each step, we are essentially dealing with semi-algebraic subsets associated to the operator on smaller boxes, and not with the central box (the central box estimate is contained in the induction assumptions). The fact that the boxes are of small size allows us to consider $\hat{u}^{(r_2)}_{\mathrm{corr}}$ instead of $\hat{u}^{(r)}_{\mathrm{corr}}$. Both of these considerations are important in order to have the degrees of the corresponding subsets not too large.

Let

$$X'_M := \{(\theta, \varphi) : \; T_{u^{(r_2)}}(\theta, \varphi)|_M \text{ is not good}\},$$

where "good" will mean slightly modified usual conditions

$$\|(T_{u^{(r_2)}}(\theta, \varphi)|_M)^{-1}\| \le 2e^{M^{\frac{6}{7}}};$$

for any $\mathbf{n}, \mathbf{n}' \in [-M, M]^3$ with $|\mathbf{n} - \mathbf{n}| \ge M^{999/1000}$ we have

$$|(T_{u^{(r_2)}}(\theta, \varphi)|_M)^{-1}(\mathbf{n}, \mathbf{n}')| \le 2e^{-\mu^{(2)}_r |\mathbf{n} - \mathbf{n}'|}.$$

Let also X''_M be the set defined in Step 3, that is, the same set with $u^{(r_2)}$ replaced by $u^{(r)}$ and the factor 2 removed. We have

$$X'_M \cap I_{r_2} \subset X''_M \cap I_{r_2}.$$

Indeed, if $(\theta, \varphi) \in I_{r_2, j} \backslash X''_M$, then the $T_{u^{(r)}}(\theta, \varphi)|_M$ is good in the usual sense, which will be preserved, up to a factor 2, after switching $u^{(r)}$ by $u^{(r_2)}$ due to (5.7). Therefore $(\theta, \varphi) \notin X'_M$. Note that (5.7) may not hold outside of I_{r_2}.

As mentioned above, the reason to consider r_2 instead of r is the degree. Define $X_M := X'_M \cap I_{r_2}$. Since $u^{(r_2)}$ is a rational function of degree $e^{(r_2 \log A)^2}$ (possibly different on each rectangle $I_{r,j}$), we can assume without loss of generality that

$$\deg X_M \le \exp\{(r_2 \log A)^{C_{\mathrm{abs},1}}\} e^{(r_2 \log A)^2} \le \exp\{(\log(r + r_0 + 1) \log A)^{2C_{\mathrm{abs},1}}\}; \tag{5.12}$$

here we assume A is large enough so that the extra $\log A$ absorbs C_L and $\log\log A$. Note that the first factor, which provides a large power of $\log(r_0 + r + 1)\log A$, comes from the fact that the construction needs to be performed on each rectangle of $I_{r_2, j}$ separately, and the number of the rectangles $I_{r_2,j}$ appears as an extra complexity factor. The degree argument would have also worked for r_1 instead of r_2, but the existing bound on $\|u^{(r_1)}-u^{(r)}\|$ is not good enough to conclude goodness of Green's functions for $T_{u^{(r_1)}}$ from that of $T_{u^{(r)}}$.

In order to apply Proposition 3.1, we would like to make all M-boxes under consideration good. In other words, to have

$$\mathbf{n}\cdot(\omega,\lambda)\notin X_M,\quad \text{for } \mathbf{n}\in[-A^{r+r_0+1},A^{r+r_0+1}]^3\backslash[-A^{r+r_0+1}/10,A^{r+r_0+1}/10]^3.$$

Let

$$\mathcal{X}_M:=\{(\omega,\lambda,\theta_1,\theta_2,\varphi):T_{u^{(r_2)}}(\theta_1+\theta_2,\varphi)|_M \text{ is not good}\}.$$

Note that the "not good" property, in general, depends on ω and λ. The set X_M defined above is a section of $\mathcal{X}_M$ with (ω,λ) fixed. Due to relations between X_M, X_M', X_M'', and $\mathcal{X}_M$, we have

$$\mathrm{Leb}_5(\mathcal{X}_M)\le 100e^{-M^{1/300}},$$

$$\mathrm{Leb}_4(\mathcal{X}_M(\theta_1))+\mathrm{Leb}_4(\mathcal{X}_M(\theta_2))+\mathrm{Leb}_4(\mathcal{X}_M(\varphi))\le 100e^{-M^{1/300}},$$

$$\mathrm{Leb}_3(\mathcal{X}_M(\theta_1,\theta_2))+\mathrm{Leb}_3(\mathcal{X}_M(\theta_2,\varphi))+\mathrm{Leb}_3(\mathcal{X}_M(\theta_1,\varphi))\le 100e^{-M^{1/300}},$$

where, for example, $\mathcal{X}_M(\theta_1,\theta_2)$ denotes the intersection of $\mathcal{X}_M$ with a hyperplane defined by θ_1, θ_2 fixed. We will follow the same idea as in (3.26) of Ref. 5. Let

$$\varepsilon_1=A^{-\frac{r+r_0+1}{4}},\qquad \varepsilon_2=A^{-\frac{r+r_0+1}{2}},\qquad \varepsilon_3=10A^{-(r+r_0+1)}.$$

Apply Proposition 2.3 with $\varepsilon=\varepsilon_1$ and $S=\mathcal{X}_M$, with $x=(\omega,\lambda)$ and $y=(\theta_1,\theta_2,\varphi)$, and construct the sets S_1 and S_2. We will have

$$\mathrm{Leb}(\mathrm{Proj}_x S_1)\le\exp\{(\log(r+r_0+1)\log A)^{C_{\mathrm{abs}}C_{\mathrm{abs},1}}\}\varepsilon_1\le A^{-(r+r_0+1)/5}.$$

The additional absolute constant appeared from applying Proposition 2.3. Note that $r+r_0$ large enough implies that $(r+r_0)\log A$ dominates $(\log(r+r_0)\log A)^C$. This is where we used the fact that r_0 needs to be chosen large depending on A. Getting back to the second conclusion of Proposition 2.3, for any (fixed) $\mathbf{n}=(\mathbf{k},j)$ with $|k_1|,|k_2|,|j|\ge 100\varepsilon_1^{-1}$, we have the following estimate on the three-dimensional measure:

$$\mathrm{Leb}_3\{(\omega,\lambda):(\omega,\lambda,k_1\omega_1,k_2\omega_2,\lambda j)\in S_2\}\le\exp\{(\log(r+r_0+1)\log A)^{C_{\mathrm{abs}}C_{\mathrm{abs},1}}\}\varepsilon_1^{-1}e^{-\frac{1}{6}M^{1/300}}.$$

Recall that $M\ge(\log N)^{C_L}$, which implies $\varepsilon_1^{-1}\le e^{M^{1/C_L}}$. By choosing a large C_L (say, $C_L\ge 400$) we can sum over all $|k_1|,|k_2|,|j|\ge 100\varepsilon_1^{-1}$ under consideration and, ultimately, guarantee that one can avoid S_2 after removing a subset of (ω,λ) of measure at most $e^{-M^{1/400}}$.

Just as in Ref. 5, the argument is not yet complete, since the centers of the boxes under consideration do not necessarily have all individual coordinates large. In order to consider the remaining cases, assume first that $|k_1|\le A^{r/4}$ and, for fixed N_1, apply Proposition 2.3 with $d_1=d_2=2$ and $\varepsilon=\varepsilon_2$, removing a small measure subset of (ω_2,λ). Large C_L, again, would imply that one can absorb the factors appearing from conditioning on such k_1. After repeating for all three coordinates, we can guarantee that the box $\mathbf{n}+[-M,M]^3$ will be good assuming that at least two numbers among $(|k_1|,|k_2|,|j|)$ are larger than $A^{r/2}$ (either from conditioning, or from the previous case). Finally, in the case where only one coordinate is larger than $A^{r/10}$, we can assume that the remaining two coordinates do not exceed $A^{r/2}$ and apply Proposition 2.3 with $\varepsilon=\varepsilon_3$, $d_1=d_2=1$, with both remaining coordinates being fixed.

Let us summarize what we have obtained so far. Every box $Q=\mathbf{n}+[-M,M]^3$ with $|\mathbf{n}|\ge A^{r/10}$ satisfies

$$\|(T_{u^{(r)}}|_Q)^{-1}\|\le 2e^{M^{\frac{6}{7}}},$$

and for any $\mathbf{n},\mathbf{n}'\in\mathbb{Z}^3$ with $|\mathbf{n}-\mathbf{n}'|\ge M^{999/1000}$, we have

$$|(T_{u^{(r)}}|_Q)^{-1}(\mathbf{n},\mathbf{n}')|\le 2e^{-\mu_r^{(2)}|\mathbf{n}-\mathbf{n}'|}.$$

Additionally, the central box $[-A^{r+r_0},A^{r+r_0}]^3$ satisfies, by the induction assumptions,

$$\|(T_{u^{(r)}}|_{A^{r+r_0}})^{-1}\|\le\delta_{\mathrm{res}}^{-1}\exp\{((r+r_0)\log A)^{C_{\mathrm{abs},3}}\},$$

and for any $|\mathbf{n}-\mathbf{n}'|>r^{C_{\mathrm{abs},4}}$,

$$|(T_{u^{(r)}}|_{A^{r+r_0}})^{-1}(\mathbf{n},\mathbf{n}')|\le\delta_{\mathrm{res}}^{-1}e^{-\mu_r|\mathbf{n}-\mathbf{n}'|}.$$

As a consequence, for each $\mathbf{n} \in [-A^{r+r_0+1}, A^{r+r_0+1}]$, there exists a good box

$$\Lambda_1 \subset [-A^{r+r_0+1}, A^{r+r_0+1}], \quad \mathbf{n} \in \Lambda_1,$$

of size M or A^{r+r_0}, such that $\operatorname{dist}(\mathbf{n}, [-A^{r+r_0+1}, A^{r+r_0+1}]\backslash\Lambda_1)$ is at least $M/2$ or $A^{r+r_0}/10$, respectively.

The set of (ω, λ) removed to achieve this has measure bounded by $e^{-M^{1/400}}$ and has degree bounded by

$$\exp\{(\log(r + r_0 + 1)\log A)^{2C_{\mathrm{abs},1}}\}.$$

Note that we are also assuming that $(\omega, \lambda) \in I_r$, so that the above measure estimate is on how much needs to be removed from I_r. While the assumptions on small boxes start from $(\omega, \lambda) \in I_{r_2}$, we also need the induction assumption for the central box to hold, requiring $(\omega, \lambda) \in I_r$. The set I_{r+1} will be constructed during Step 9 in order to preserve nesting properties and rectangular structure.

Apply Proposition 3.1. Note that during the iterations we are using only off-diagonal entries of $T_{u^{(r)}}$ which contain δ^{2p-1}. Therefore, $\delta_{\mathrm{res}}^{-1}$ can be absorbed into it and we do not have any accumulation of negative powers of δ as the iterations proceed. We thus obtain, for the box $\Lambda = [-A^{r+r_0+1}, A^{r+r_0+1}]^3$:

$$\|(Tu^{(r)}|A^{r+1})^{-1}\| \le 2\delta_{\mathrm{res}}^{-1}(\exp\{((r + r_0)\log A)^{C_{\mathrm{abs},3}} + e^{M^{6/7}}). \tag{5.13}$$

Note that

$$e^{M^{6/7}} = e^{(\log A^{r+r_0})^{C_L}} = A^{((r+r_0)\log A)^{C_L-1}},$$

which implies the norm estimate in (sol5) with $u^{(r)}$ on the box of the size A^{r+r_0+1} and with the exponential decay rate $\mu_{r+1}^{(3)}$. To summarize, we have

$$\|(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}\| \le 2\delta_{\mathrm{res}}^{-1}G_r, \tag{5.14}$$

and for $|\mathbf{n} - \mathbf{n}'| \ge D_r$, we have

$$|(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}(\mathbf{n}, \mathbf{n}')| \le \delta^{p-1/2}e^{-\mu_r^{(3)}|\mathbf{n}-\mathbf{n}'|}. \tag{5.15}$$

We can now address the stability of the obtained estimates. Note that the estimates in the definition of good M-boxes are stable under perturbations of size e^{-M^2}, and $\delta_r \ll e^{-M^2}$ (here, we allow an extra 2 factor which can be absorbed into various constants appearing along the way). The estimate for the central box, from the induction assumptions, is stable under perturbations of size $10\delta_r$. As a consequence, provided (sol3), we can replace $\hat{u}^{(r)}$by $\hat{u}^{(r+1)}$, and the resulting bounds will be stable under perturbations of size $9\delta_r$. As long is δ_{r+1} is much smaller, this implies (reg7) on the next scale.

As discussed in the plan of the induction step, very small δ may lead to an improvement in the measure bounds, since one can use direct perturbation arguments instead of the steep planes lemma. It is easy to see that all the conclusions of the above considerations will be satisfied on the set defined by (5.2) with r_0 replaced by $r_0 + r + 1$. Therefore, it is more beneficial to use (5.2) as long as the measure estimates (5.3) are better than $e^{-M^{1/400}}$. As a consequence, the actual measure estimate that one needs to remove is, say,

$$\min\{A^{100(r+r_0+1)}\delta_{\mathrm{res}}, e^{-M^{1/400}}\} \le \delta_{\mathrm{res}}^{1/2}e^{-M^{1/500}} = \delta_{\mathrm{res}}^{1/2}\exp\{-((r + r_0 + 1)\log A)^{C_L/500}\}. \tag{5.16}$$

The factor δ_{res} in front will be important during the final measure estimates.

H. Step 5: Construction of $\hat{u}_{\mathrm{corr}}^{(r+1)}$, $\hat{v}_{\mathrm{corr}}^{(r+1)}$

We have the induction assumptions

$$F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c} \le \kappa_r e^{-\mu_r|\mathbf{n}|}, \qquad \nabla F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c} \le \kappa_r e^{-\mu_r|\mathbf{n}|}. \tag{5.17}$$

The next step corrections are defined by

$$\begin{pmatrix} \hat{u}_{\mathrm{corr}}^{(r+1)} \\ \hat{v}_{\mathrm{corr}}^{(r+1)} \end{pmatrix} = -(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c}. \tag{5.18}$$

We will apply Corollary 5.2, using (5.17) and the conclusions of Step 4 (5.14) and (5.15). For $r \le r_0$, we have $D = 1$, $R = 100(r + r_0)^{99}$ in the notation of Corollary 5.2, which implies

$$|\hat{u}_{\text{corr}}^{(r+1)}(\mathbf{n})| \le \kappa_r \left(\frac{10^{20}(r + r_0)^{600}}{\mu^3} + 16\delta_{\text{res}}^{-1} G_r e^{2\mu} \right) e^{-\mu_r^{(4)}|\mathbf{n}|} \le \kappa_r \delta_{\text{res}}^{-1} G_r^2 e^{-\mu_r^{(4)}|\mathbf{n}|}. \tag{5.19}$$

Note that the extra factor of G_r would absorb all other contributions. In order to estimate the gradient, one can use the formula $\nabla(T^{-1}) = T^{-1}(\nabla T)T^{-1}$. All factors in this formula have decay rate $\mu_r^{(3)}$ or better. Note that, while (sol6) only estimates off-diagonal entries of ∇T, the diagonal entries have explicit form and can be estimated by $A^{C(r+r_0)}$. Thus, a similar estimate follows from Corollary 5.2, with more iterations of the convolution. However, one can easily check that the extra factors can also be absorbed into an extra G_r. Thus,

$$|(\nabla \hat{u}_{\text{corr}}^{(r+1)})(\mathbf{n})| \le \kappa_r \delta_{\text{res}}^{-1} G_r^2 e^{-\mu_r^{(4)}|\mathbf{n}|}. \tag{5.20}$$

In the case $r \ge r_0$, we now have a non-trivial factor $e^{2\ \mu D} = e^{2\ \mu D_r}$ in the application of Corollary 5.2. However, G_r is also larger. By choosing $C_{\text{abs},3} \gg C_{\text{abs},4}$, one can absorb it into an extra factor G_r in a similar way. As a consequence, (5.19) and (5.20) *are valid in the full range of values of* r.

Note that, since (sol3) has been obtained, one can use (reg7) and state that (sol5) now holds at the scale $r + 1$, with the decay rate $\mu_r^{(3)}$ (note that the application of (reg7) does not cause deterioration in the exponential decay rate).

I. Step 6: Estimates of $F[\hat{u}^{(r+1)}, \hat{v}^{(r+1)}]|_{S^c}$

Recall that, from Remark 4.2, we have

$$F[\hat{u} + \hat{u}_{\text{corr}}, \hat{v} + \hat{v}_{\text{corr}}] - F[\hat{u}, \hat{v}] - F'[\hat{u}, \hat{v}] * \begin{pmatrix} \hat{u}_{\text{corr}} \\ \hat{v}_{\text{corr}} \end{pmatrix} = \delta^{2p} R(\hat{u}, \hat{u}_{\text{corr}}),$$

where the remainder is a convolution polynomial in $\hat{u}$ and $\hat{u}_{\text{corr}}$ and its conjugates, at least quadratic in $\hat{u}_{\text{corr}}$ and $\hat{v}_{\text{corr}}$. Apply the above equality to $\hat{u}^{(r+1)} = \hat{u}^{(r)} + \hat{u}_{\text{corr}}^{(r+1)}$ and restrict everything to S^c. By construction,

$$\begin{pmatrix} \hat{u}_{\text{corr}}^{(r+1)} \\ \hat{v}_{\text{corr}}^{(r+1)} \end{pmatrix} = -(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c},$$

and in particular the left hand side is supported on S^c. Therefore, we have

$$F[\hat{u}^{(r+1)}, \hat{v}^{(r+1)}]|_{S^c} = F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c} + T_{u^{(r)}} \begin{pmatrix} \hat{u}_{\text{corr}}^{(r+1)} \\ \hat{v}_{\text{corr}}^{(r+1)} \end{pmatrix} + \delta^{2p} R(\hat{u}^{(r)}, \hat{u}_{\text{corr}}^{(r+1)}),$$

which implies after a substitution

$$\begin{aligned} F[\hat{u}^{(r+1)}, \hat{v}^{(r+1)}]|_{S^c} &= (T_{u^{(r)}}|_{A^{r+r_0+1}} - T_{u^{(r)}})(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c} + \delta^{2p} R(\hat{u}^{(r)}, \hat{u}_{\text{corr}}^{(r+1)}) \\ &= -(1 - \mathbf{1}_{A^{r+r_0+1}}) T_{u^{(r)}} \mathbf{1}_{A^{r+r_0+1}} (T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c} + R(\hat{u}^{(r)}, \hat{u}_{\text{corr}}^{(r+1)}). \end{aligned}$$

Due to the construction of $\hat{u}^{(r)}$, it is supported on $[-A^{r+r_0}, A^{r+r_0}]^3$, and therefore $F[\hat{u}^{(r)}, \hat{v}^{(r)}]$ is supported on, say, $[-6pA^{r+r_0}, 6pA^{r+r_0}]^3$. Assume that A is large enough so that $3p < A/100$. Then one can further expand

$$\begin{aligned} &F[\hat{u}^{(r+1)}, \hat{v}^{(r+1)}] = \\ &= -(1 - \mathbf{1}_{A^{r+r_0+1}}) T_{u^{(r)}} (\mathbf{1}_{A^{r+r_0+1}} - \mathbf{1}_{A^{r+r_0+1} - 6pA^{r+r_0}}) (T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1} \mathbf{1}_{6pA^{r+r_0}} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{S^c} \\ &\qquad + \delta^{2p} R(\hat{u}^{(r)}, \hat{u}_{\text{corr}}^{(r+1)}). \end{aligned} \tag{5.21}$$

In the above expression, we used the fact that $T_{u^{(r)}}(\mathbf{m},\mathbf{n}) = 0$ for $|\mathbf{m}-\mathbf{n}| \geq 6pA^{r+r_0}$, since its off-diagonal part contains at most p copies of convolution with $\hat{u}^{(r)}$. In other words,

$$(\mathbf{1}_{A^{r+r_0+1}} - 1)T_{u^{(r)}}\mathbf{1}_{A^{r+r_0+1}} = (\mathbf{1}_{A^{r+r_0+1}} - 1)T_{u^{(r)}}\mathbf{1}_{A^{r+r_0+1}}(1 - \mathbf{1}_{A^{r+r_0+1}-6pA^{r+r_0}}).$$

Let us estimate the first term in (5.21). From the outcome of Step 4, we have

$$\left|\left((1 - \mathbf{1}_{A^{r+r_0+1}-3pA^{r+r_0}})(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}\mathbf{1}_{3pA^{r+r_0}}\right)(\mathbf{n},\mathbf{n}')\right| \leq \delta^{p-1/2}e^{-\mu_r^{(3)}|\mathbf{n}-\mathbf{n}'|}, \tag{5.22}$$

with a similar estimate for the matrix elements of $T_{u^{(r)}}$. Applying Corollary 5.2, we obtain that the first term is bounded in absolute value by, say

$$\left(\frac{100|\mathbf{n}|}{1+\mu}\right)^{100}(1 - \mathbf{1}_{A^{r+r_0+1}})\kappa_r e^{-\mu_r^{(3)}|\mathbf{n}|} \leq \frac{1}{2}\kappa_{r+1}(1 - \mathbf{1}_{A^{r+r_0+1}})e^{-\mu_r^{(4)}|\mathbf{n}|}. \tag{5.23}$$

Here we use that fact that, by (reg6), κ_r grows fast (super-exponentially in r), but still slow enough so that one can gain an extra factor of κ_{r+1} by deteriorating the exponent from $-\mu_r^{(3)}$ to $-\mu_r^{(4)}$ (the first factor is also absorbed in the deterioration). It is important here that $\mathbf{n}$ is outside the box $[-A^{r+r_0+1}, A^{r+r_0+1}]^3$. See (5.26).

An estimate of the gradient is similar, with potentially more iterations of convolution due to the use of the formula $\nabla(T^{-1}) = T^{-1}(\nabla T)T^{-1}$ which may result in the larger power of $|\mathbf{n}|$ in the front factor (see Corollary 5.2).

We can now estimate the remaining "quadratic" term $\delta^{2p}R(\hat{u}^{(r)}, \hat{u}_{\mathrm{corr}}^{(r+1)})$. Recall that R is a convolution polynomial, which contains at least two factors $\hat{u}_{\mathrm{corr}}^{(r+1)}$ or $\hat{v}_{\mathrm{corr}}^{(r+1)}$. The polynomial has at most (say) 10^p monomials. Each factor in each monomial has decay rate $\mu_r^{(4)}$ or better. At least two factors in each monomial will additionally have δ_r in front. Thus, after combining all bounds and applying Lemma 5.1, we obtain

$$|\delta^{2p}R(\hat{u}^{(r)}, \hat{u}_{\mathrm{corr}}^{(r+1)})(\mathbf{n})| \leq \frac{1}{2}\delta^{2p}\delta_r^2 G_r e^{-\mu_r^{(5)}|\mathbf{n}|}.$$

Here, we chose for simplicity to absorb various polynomial and combinatorial factors into an extra factor G_r as we did before. A similar estimate holds for the gradient.

J. Step 7: Decay of small inductive parameters

Recall that

$$\hat{u}^{(0)} = u_1\mathbf{1}_{\{(-\mathbf{e}_1,h_1)\}} + u_2\mathbf{1}_{\{(-\mathbf{e}_2,h_2)\}}, \qquad \hat{v}^{(0)}(\mathbf{n}) = \overline{\hat{u}^{(0)}(-\mathbf{n})},$$

and

$$|F[\hat{u}^{(0)}, \hat{v}^{(0)}]|_{S^c}(\mathbf{n})| \leq \delta^{2p}\max(|u_1|,|u_2|)^p C(p).$$

Note that, due to the presence of the diagonal part, $F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_S$ is not necessarily small. We also have that

$$\operatorname{supp} F[\hat{u}^{(0)}, \hat{v}^{(0)}] \subset [-3p\max(|h_1|,|h_2|), 3p\max(|h_1|,|h_2|)]^3.$$

We will be choosing a large A depending on S and p, then a large r_0 depending on A. It will be convenient to have the choice of δ being

$$e^{-C_{\mathrm{abs},8}r_0 \log A} \leq \delta \leq e^{-C_{\mathrm{abs},7}r_0 \log A}. \tag{5.24}$$

As of now, a smaller δ is only better. However, avoiding too small values of δ will make certain estimates more simple.

With all the above preparations, one can clearly choose r_0 and A in a way that $\kappa_0 = \delta^p$. The estimates in Steps 5 and 6 force us to choose

$$\delta_{r+1} \geq \kappa_r\delta_{\mathrm{res}}^{-1}G_r^2, \quad \kappa_{r+1} \geq \delta^{2p}G_r\delta_{r+1}^2.$$

As a consequence, we have the following recursion inequality for κ_r:

$$\kappa_0 = \delta^p, \quad \kappa_{r+1} \geq \delta^{2p}\delta_{\mathrm{res}}^{-2}G_r^5\kappa_r^2.$$

We will choose a specific solution to this recursion by assuming

$$\delta^{2p}\delta_{\rm res}^{-2}G_r^5\kappa_r^{1/2} \le 1 \tag{5.25}$$

and checking that it holds retroactively. The solution of the recursion

$$\kappa_0 = \delta^p, \quad \kappa_{r+1} = \kappa_r^{3/2}$$

is

$$\kappa_r = \delta^{p(3/2)^r} = e^{-pC_{\rm abs,7}r_0(3/2)^r \log A}. \tag{5.26}$$

It is easy to see that, assuming that $C_{\rm abs,7}$ is large enough, it satisfies (5.25), and therefore is a valid choice for the original recurrence. We have also retroactively confirmed (5.23). Note that, if δ is too small (say, $\delta = e^{-A^{r_0}}$), the arguments that lead to (5.23) prevent κ_r from decaying as fast as in the first equality in (5.26), and the recursive estimates become more cumbersome.

K. Step 8: Estimates of $\hat{u}^{(r+1)}$

So far, we have

$$|\hat{u}^{(r+1)}(\mathbf{n})| \le |\hat{u}^{(r)}(\mathbf{n})| + |\hat{u}_{\rm corr}^{(r+1)}(\mathbf{n})| \le (1-2^{-(r+1)})e^{-\mu_r|\mathbf{n}|} + \delta_{r+1}e^{-\mu_r^{(4)}|\mathbf{n}|} \le (1-2^{-(r+2)})e^{-\mu_r^{(4)}|\mathbf{n}|},$$

since $\delta_r \le 2^{r+2}$. Gradient estimates are similar, in view of $\nabla\hat{u}^{(0)} = 0$.

L. Step 9: Structure of I_j and extension into the full parameter space

In this subsection, we finalize the proofs of the inductive statements. In particular, we discuss the definition of the sets I_r in order to satisfy (i1)–(i3), degree bounds on $\hat{u}^{(r)}$ in order to establish (reg5), estimates of the measure of the bad set in terms of δ, and extension of the functions $\hat{u}^{(r)}$, $\hat{v}^{(r)}$ into the full parameter space.

1. The sets I_r and the dyadic structure

Recall that the set $I_0 = I_1 = \cdots = I_{r_0}$ is a union of at most $A^{100r_0}\delta_{\rm res}^{-10}$ dyadic cubes with sides at least $\delta_{\rm res}^2$. For $r \ge r_0+1$, we will now describe more explicitly the removed subsets of parameters. Let

$$\Gamma_{\rm DC}^{(r)} := {\rm DCR}(\tau, A^{10(r+r_0)})\backslash{\rm DCR}(\tau, A^{10(r+r_0+1)})$$

be the set of Diophantine frequencies one needs to remove to enforce the Diophantine conditions on the next scale. We have

$${\rm Leb}_3(\Gamma_{\rm DC}^{(r)}) \le 10\tau^2 A^{-9(r+r_0+1)}.$$

Let also $\Gamma_{\rm SP}^{(r)}$ be the set of parameters removed in the end of Step 4. We have

$${\rm Leb}_3(\Gamma_{\rm SP}^{(r)}) \le \delta_{\rm res}^{1/2}e^{-M^{1/500}} = \delta_{\rm res}^{1/2}\,\exp\{-((r+r_0+1)\log A)^{C_L/500}\}.$$

Let $\Gamma^{(r)}$ be the union of all dyadic cubes with sides at least e^{-M^2}, contained in $\left(\Gamma_{\rm DC}^{(r)} \cup \Gamma_{\rm SP}^{(r)}\right) \cap I_r$. Define for $r \ge r_0$:

$$I_{r+1} := I_r\backslash\Gamma^{(r+1)}.$$

From (reg7) and the choice of the sizes of the dyadic cubes, one can see that the conclusions of Step 4 will still be valid for the sets I_r. The dyadic structure implies the nesting properties, as well as the bounds on the number of cubes. Therefore, we have obtained (i1)–(i3). The sets $I_{r,j}$ can be simply chosen to be the corresponding dyadic cubes. On each step, we are dividing some cubes into smaller cubes and remove some of the new cubes.

2. Degree bounds on $\hat{u}^{(r)}$

We will prove the degree bounds by induction. Assume $\deg \hat{u}^{(r)}(\mathbf{n}) \le A^{((r+r_0)\log A)^2}$. Recall

$$\begin{pmatrix} \hat{u}^{(r+1)}_{\mathrm{corr}} \\ \hat{v}^{(r+1)}_{\mathrm{corr}} \end{pmatrix} = -(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1} F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c}.$$

Then we have, due to the convolution bounds,

$$\deg F[\hat{u}^{(r)}, \hat{v}^{(r)}]|_{A^{r+r_0+1}, S^c} \le 100pe^{((r+r_0)\log A)^2}, \qquad \deg T_{u^{(r)}}|_{A^{r+r_0+1}}(\mathbf{m},\mathbf{n}) \le 100pe^{((r+r_0)\log A)^2}.$$

Due to Cramer's rule, we also have

$$(T_{u^{(r)}}|_{A^{r+r_0+1}})^{-1}(\mathbf{m},\mathbf{n}) \le A^{20(r+r_0)} \cdot 100pe^{((r+r_0)\log A)^2}.$$

Overall, assuming A is large depending on p,

$$\begin{aligned} \deg \hat{u}^{(r+1)}_{\mathrm{corr}} &\le A^{40(r+r_0)} e^{((r+r_0)\log A)^2} \\ &= \exp\{40(r+r_0)\log A + ((r+r_0)\log A)^2\} \le \exp\{((r+r_0+1)\log A)^2\}, \end{aligned}$$

provided $\log A \ge 20$. This implies the degree bounds on the next step.

3. Estimates of the measure of the bad set in terms of δ

Recall from Step 3 that the choice of δ is made based on the Diophantine parameter τ:

$$|\delta| \le e^{-\tau^{-C_{\mathrm{abs,dio}}}}$$

or, equivalently,

$$\frac{1}{(\log(1/\delta))^{1/C_{\mathrm{abs,dio}}}} \le \tau.$$

In view of (reg6), this implies

$$\begin{aligned} \mathrm{Leb}_3(\Gamma^{(r)}) &\le 10\tau^2 A^{-9(r+r_0+1)} + \delta_{\mathrm{res}}^{1/2} \exp\{-((r+r_0+1)\log A)^{C_L/500}\} \\ &\le \frac{10\delta^{1/C_{\mathrm{abs,7}}}}{(\log(1/\delta))^{2/C_{\mathrm{abs,dio}}}} A^{-8(r+r_0+1)} + \delta^{C_{\mathrm{abs,9}}/C_{\mathrm{abs,7}}} \exp\{-((r+r_0+1)\log A)^{C_L/500}\}. \end{aligned} \tag{5.27}$$

Note that the actual constants are not particularly important. However, the estimate

$$\sum_{r\ge r_0} \mathrm{Leb}_3(\Gamma^{(r)}) = O(\delta^{c_{\mathrm{abs}}}),$$

which follows from (5.27), is of crucial importance, since the measure will be multiplied by $\delta^{-\varkappa}$ in the end of Sec. VI.

4. Extension to the whole parameter space

For each dyadic cube $Q \subset I_r$, consider a smaller (not necessarily dyadic) cube $\widetilde{Q}$ with the same center and

$$\mathrm{diam}(\widetilde{Q}) = \mathrm{diam}(Q) - \mathrm{diam}(Q)^2.$$

For each such Q, consider a smooth cutoff function η_Q that is equal to 1 on $\widetilde{Q}$ and zero on $\mathbb{R}^3 \setminus Q$. Let $\eta^{(r)} := \sum_Q \eta_Q$. Note that, by construction, $\eta^{(0)} = \cdots = \eta^{(r_0)}$, since the set I_r does not change. Let

$$\Xi^{(r)} := \begin{cases} \eta^{(r)}, & 0 \le r \le r_0 \\ \prod_{j=r_0}^{r} \eta^{(j)}, & r > r_0. \end{cases}$$

Replace $\hat{u}^{(r)}_{\mathrm{corr}}$ by $\Xi^{(r)} \hat{u}^{(r)}_{\mathrm{corr}}$. Note that, since $\Xi^{(r)} \subset I_r = \mathrm{dom}\, \hat{u}^{(r)}_{\mathrm{corr}}$, this defines a smooth extension of $\hat{u}^{(r)}_{\mathrm{corr}}$ into the whole space $(1,2)\times(-3,3)^2$ (or into $\mathbb{R}^3$ if necessary).

One can easily check that, within the above constraints, we have

$$|\nabla \Xi^{(r)}| \le \begin{cases} A^{100r_0}\delta_{\mathrm{res}}^{-3}, & 0 \le r \le r_0; \\ e^{10M^2}, & r > r_0. \end{cases}$$

In both cases, it follows from (reg6) that one can absorb this additional correction to the gradient into δ_r. As a consequence, the sequence $\hat{u}^{(r)}$ converges to a function u in $C^1((1,2)\times(-3,3))$. For the "good" values of parameters

$$(\lambda,\omega_1,\omega_2) \in (1,2)\times(-3,3)^2 \setminus \bigcup_r \Gamma^{(r)}$$

u becomes a solution to the non-linear equation (m_1 and m_2 are obtained from the Q-equations on each scale, and also clearly converge). For general values of $(\lambda,\omega_1,\omega_2)$, one can still perform some of the induction steps, and then the sequence will "freeze," thus providing an approximate solution to the non-linear equation, with the error κ_r.

Remark 5.4. The reason for multiplying the cut-off functions in the expression $\Xi^{(r)} = \prod_{j=r_0}^{r} \eta^{(j)}$ for $r \ge r_0$ is that, without the multiplication, the supports of the cut-off functions do not necessarily decrease, since larger cubes may have larger neighborhood of their boundary removed.

VI. Q-EQUATIONS USING PARAMETERS m AND λ: COMPLETION OF THE PROOF

As discussed in Sec. V, we have constructed solutions $u^{(r)}$ and $v^{(r)}$ which converge, respectively, to functions u and v for $(\lambda,\omega_1,\omega_2) \in (1,2)\times(-3,3)\setminus\bigcup_r \Gamma_r$, and obtained all stated properties of those solutions. It remains to consider the Q-equations

$$\omega_k = 2\cos(h_k\lambda) + m_k\delta^{\varkappa} + \frac{\delta^{2p}}{u_k}((\hat{u}*\hat{v})^{*p}*\hat{u})(-\mathbf{e}_k,h_k), \quad k=1,2,$$

where u_k are the values of the initial iteration $\hat{u}^{(0)}$ on S:

$$\hat{u}^{(0)}(\mathbf{n}) = \begin{cases} u_1, & \mathbf{n} = (-\mathbf{e}_1,h_1); \\ u_2, & \mathbf{n} = (-\mathbf{e}_2,h_2); \\ 0, & \mathbf{n} \in \mathbb{Z}^3\setminus S_+. \end{cases}$$

Note that, for all $r \ge 0$, we have

$$\hat{u}^{(r)}(-\mathbf{e}_k,j_k) = u_k, \quad k=1,2,$$

so that the iterations only change the values of $\hat{u}^{(r)}$ outside S_+.

Fix $(\mathbf{e}_1,j_1)$, $(\mathbf{e}_2,j_2)$, u_1, u_2, and let

$$f_k^{(r)}(\lambda,\omega_1,\omega_2) := \frac{((\hat{u}^{(r)}*\hat{v}^{(r)})^{*p}*\hat{u}^{(r)})(-\mathbf{e}_k,h_k)}{u_k}, \quad k=1,2. \tag{6.1}$$

Then, one can write a finite volume version of the Q-equations as follows:

$$\omega_k^{(r)} = 2\cos(h_k\lambda) + m_k^{(r)}\delta^{\varkappa} + \delta^{2p} f_k^{(r)}(\lambda,\omega_1,\omega_2), \quad k=1,2. \tag{6.2}$$

Note that we do not actually enforce these equations for finite r. Due to the results of Sec. V L 4, the right hand side of (6.1) converges in C^1 to some function f. Thus, we can rewrite the infinite volume Q-equations as

$$\omega_k = 2\cos(h_k\lambda) + m_k\delta^{\varkappa} + \delta^{2p} f(\lambda,\omega_1,\omega_2),$$

or, equivalently,

$$m_k(\lambda,\omega_1,\omega_2) = \delta^{-\varkappa}\omega_k + \delta^{-\varkappa}\cos(h_k\lambda) - \delta^{2p-\varkappa} f_k(\lambda,\omega_1,\omega_2). \tag{6.3}$$

Recall that we were requiring

$$(\lambda,\omega_1,\omega_2) \in (1,2)\times(-3,3)\backslash\bigcup_r \Gamma^{(r)}.$$

Recall (5.27). Since the differential of the map (6.3) is bounded by, say, $10\delta^{-\varkappa}(|h_1|+|h_2|)$, we can ultimately conclude that the set of bad parameters (λ,m_1,m_2) that one needs to exclude will have measure bounded by at most

$$\delta^{-\varkappa}\sum_{r\geq r_0}\left(\frac{10\delta^{1/C_{\text{abs},7}}}{(\log(1/\delta))^{2/C_{\text{abs,dio}}}}A^{-8(r+r_0+1)} + \delta^{C_{\text{abs},9}/C_{\text{abs},7}}\exp\{-((r+r_0+1)\log A)^{C_L/500}\}\right), \tag{6.4}$$

which can be made arbitrarily small with small $\varkappa$ depending on the various absolute constants chosen in the course of the inductive procedure. Note that the logarithmic factor can be absorbed by the power of δ in the numerator.

Remark 6.1. Instead of just enforcing the infinite volume Q-equation, one can enforce a stronger condition: that is, that all modulated frequencies $\omega_k^{(r)}$ in the left hand side of (6.2) actually belong to the good set. Indeed, in that case one needs perform a similar removal with $f_k(\lambda,\omega_1,\omega_2)$ in (6.3) replaced by $f_k^{(r)}(\lambda,\omega_1,\omega_2)$, and take the union over all r. One can check that $\|f_k - f_k^{(r)}\|_{C_1} \lesssim \tilde{\delta}_r$. As a consequence, the removed set of parameters (λ,m_1,m_2) will expand by $O(\tilde{\delta}_r)$ which does not affect smallness of the expression (6.4). However, this additional requirement does not appear to be necessary.

VII. ON THE LARGE DEVIATION THEOREM

In this section, we will discuss the Proof of Proposition 3.2, which is a large deviation theorem for Green's function for scale-dependent operators H_k. The proof is standard and is based on the techniques that date back to Ref. 7, and we assume that the reader has some familiarity with these techniques. The main difference of the current setting from the one in Refs. 7 and 8 is the presence of non-local operators, with the exponential decay rate of non-local terms varying between scales. A version of the LDT in the non-local setting has been developed in Ref. 14, however, it needs to be modified in order to prevent too fast deterioration of the decay rate. A detailed argument with full control of the exponents has been developed in Ref. 12 in the setting of Ref. 4, Chap. 18. In the latter setting, the operator families are one-parametric, and the (much stronger version of) the arithmetic lemma follows from the Diophantine condition on the frequency vector. In our setting, we need to deal with two-parametric families and two-dimensional semi-algebraic subsets; however, the main ingredient (the arithmetic Lemma 2.6) is already at our disposal. Combining the ideas from[6,7,12,14] is relatively straightforward. For the convenience of the reader, we would like to provide more detail on such combination. We will take advantage of several lemmas that can be used without any changes, and try to provide the main steps of the argument, as streamlined as possible, for the remaining proof.

A. The initial scale

Recall the definition of the operator T:

$$T(\theta,\varphi) = D(\theta,\varphi) + \varepsilon H = \begin{bmatrix} D_+(\theta,\varphi) & 0 \\ 0 & D_-(\theta,\varphi), \end{bmatrix} + \varepsilon H,$$

where

$$D_\pm(\mathbf{n},\mathbf{n}) = \pm(n_1\omega_1 + n_2\omega_2 + \theta) + 2\cos(n_3\lambda + \varphi), \quad \mathbf{n}\in\mathbb{Z}^3;$$

$$H(\mathbf{n}_1+\mathbf{m},\mathbf{n}_2+\mathbf{m}) = H(\mathbf{n}_1,\mathbf{n}_2), \quad \forall\mathbf{n}_1,\mathbf{n}_2,\mathbf{m}\in\mathbb{Z}^3;$$

$$|H(\mathbf{n}_1,\mathbf{n}_2)| \leq e^{-\gamma|\mathbf{n}_1-\mathbf{n}_2|}.$$

An *elementary region* in $\mathbb{Z}^3$ is a subset of the form $R\backslash(R+\mathbf{n})$, where R is a rectangle in $\mathbb{Z}^3$ and $\mathbf{n}\in\mathbb{Z}^3$ is a non-zero lattice vector. The class of elementary regions of diameter M is denoted by $\mathcal{ER}(M)$.

For the purpose of the following proposition, fix γ, $0 < b < 1$, $0 < \beta < 1$. We will say that a region $\Lambda \in \mathcal{ER}(M)$ is *good* if

$$\|T(\theta,\varphi)|_{\Lambda}^{-1}\| \leq e^{M^b}, \tag{7.1}$$

and for all $\mathbf{n}, \mathbf{n}' \in \Lambda$ with $|\mathbf{n} - \mathbf{n}'| \geq M^{\beta}$, we have the following estimates for the matrix elements:

$$|T(\theta,\varphi)|_{\Lambda}^{-1}(\mathbf{n},\mathbf{n}')| \leq e^{-\gamma|\mathbf{n}-\mathbf{n}'|}. \tag{7.2}$$

Proposition 7.1. Suppose that $T = T(\theta,\varphi)$ is defined as above. For every $M_0 \in \mathbb{N}$, $0 < \gamma_0 < \gamma_1$, $0 < \rho < b < \beta < 1$, and $(\lambda, \omega_1, \omega_2) \in \mathbb{R}^3$, there exists $\varepsilon_0 = \varepsilon_0(M_0, \gamma_0, \rho, b, \beta)$ and $M_0' = M_0'(\gamma_0, \gamma_1, \rho, b, \beta)$ such that, for every $\Lambda \in \mathcal{ER}(M)$ with $M_0' \leq M \leq M_0$ and $\gamma_0 \leq \gamma \leq \gamma_1$, the set

$$X_M = \{(\theta,\varphi) \in \mathbb{R}\times\mathbb{T} : \Lambda \text{ is not good}\}$$

satisfies

$$\mathrm{Leb}_1(X_M \cap L) \leq e^{-M^{\rho}}$$

for every unit line segment $L \subset \mathbb{R}\times\mathbb{T}$.

Proof. Fix $a \in \mathbb{R}$, and let $f(\theta,\varphi) := \theta + 2\cos(a+\varphi)$. It is easy to see that, for every $b \in \mathbb{R}$ and every unit line segment $L \in \mathbb{R}\times\mathbb{T}$, we have

$$\mathrm{Leb}_1(L \cap \{(\theta,\varphi) : |f(\theta,\varphi) - b| \geq \zeta\}) \leq c_{\mathrm{abs}}\zeta^{1/2}, \quad \forall \zeta \in (0,1). \tag{7.3}$$

This follows from the fact that no more than two derivatives of f can vanish simultaneously at a given point. One can also consider it as an elementary case of the Łojasiewicz inequality.

Let $0 < \rho < \rho_1 < b_1 < b < 1$. Let $\Lambda \in \mathcal{ER}(M)$, and

$$X := \{(\theta,\varphi) \in \mathbb{R}\times\mathbb{T} : |D_+(\mathbf{n},\mathbf{n})| + |D_-(\mathbf{n},\mathbf{n})| \geq e^{-M^{b_1}} \quad \text{for some } \mathbf{n} \in \Lambda\}.$$

Using (7.3), we have for every unit line segment $L \subset \mathbb{R}\times\mathbb{T}$:

$$\mathrm{Leb}_1(X \cap L) \leq 2M^3 c_{\mathrm{abs}} e^{-\frac{1}{2}M^{\rho_1}} \ll e^{-M^{\rho}}.$$

Suppose that $(\theta,\varphi) \in \mathbb{R}\times\mathbb{T}\backslash X$ and $\varepsilon < M^{-10}e^{-M^b}$. Then

$$|D_{\pm}(\mathbf{n},\mathbf{n})| \geq e^{-M^{b_1}}, \ \ \mathbf{n} \in \Lambda; \quad \|\varepsilon H\| \leq \frac{1}{2}e^{-M^b}.$$

Using the Neumann series argument, we obtain the first estimate

$$\|T(\theta,\varphi)|_{\Lambda}^{-1}\| \leq 2e^{M^{b_1}} \ll e^{M^b}.$$

In order to obtain the exponential decay estimate, let us use the shorter notation $T(\theta,\varphi)|_{\Lambda} = T_{\Lambda} = D_{\Lambda} + \varepsilon H_{\Lambda}$. We have

$$T_{\Lambda}^{-1} = D_{\Lambda}^{-1} - \varepsilon D_{\Lambda}^{-1}H_{\Lambda}D_{\Lambda}^{-1} + \varepsilon^2 D_{\Lambda}^{-1}H_{\Lambda}D_{\Lambda}^{-1}H_{\Lambda}D_{\Lambda}^{-1} - \cdots$$

Let us estimate individually the contributions of each term of the above series. The first term does not have any off-diagonal component. In the next terms, one can estimate each matrix element using matrix multiplication. The combinatorial factors and the products of D_{Λ}^{-1} can be absorbed into the powers of ε. After the matrix multiplication, the total exponential factor contributing into $T_{\Lambda}^{-1}(\mathbf{n}_1,\mathbf{n}_2)$ will be at least $\gamma|\mathbf{n}_1 - \mathbf{n}_2|$. After the summation of the series, we can obtain, say

$$|T_{\Lambda}^{-1}(\mathbf{n}_1,\mathbf{n}_2)| \leq e^{M^{b_1}} + \varepsilon^{1/2}e^{-\gamma|\mathbf{n}_1-\mathbf{n}_2|},$$

which implies (7.2) after a suitable choice of M_0'. □

B. Covering lemma, coupling lemma, and Cartan lemma

The following lemma will be referred to as the covering lemma. In this form, it is stated in Ref. 12. The proof is very similar to the one of Proposition 3.1.

Lemma 7.2. Let $\Lambda \subset \mathbb{Z}^d$ be a finite subset with $|\Lambda| = N$ and T be an operator on $\ell^2(\Lambda)$. Let

$$0 < \theta < 1, \qquad 0 < b < 1, \qquad 0 < \tau < 1, \qquad b\tau < \theta, \qquad \alpha > 0, \qquad \rho > 0.$$

Let also

$$(\log N)^{10/b} < M < N^{\tau}.$$

$$|T(\mathbf{m},\mathbf{n})| \le e^{-\rho|\mathbf{m}-\mathbf{n}|}, \quad \mathbf{m} \ne \mathbf{n}, \quad \mathbf{m},\mathbf{n} \in \Lambda,$$

and for every $\mathbf{m} \in \Lambda$ there exists an elementary region $U(m) \subset \Lambda$ containing $\mathbf{m}$ such that

$$|U(\mathbf{m})| = M, \quad \operatorname{dist}(\mathbf{m}, \Lambda \backslash U(\mathbf{m})) > M/2,$$

such that

$$\| T|_{U(\mathbf{m})}^{-1} \| < e^{M^b},$$

and

$$T|_{U(\mathbf{m})}^{-1} \langle e^{-\rho|\mathbf{m}-\mathbf{n}|}, \qquad \textit{for } \mathbf{m},\mathbf{n} \in U(\mathbf{m}),\ |\mathbf{m}-\mathbf{n}| \rangle M^{\theta}.$$

Then, we have

$$\| T^{-1} \| < 2N^d e^{M^b} < e^{N^b},$$

$$|T^{-1}(\mathbf{m},\mathbf{n})| \langle e^{-(\rho-(\log N)^{-50})|\mathbf{m}-\mathbf{n}|}, \quad |\mathbf{m}-\mathbf{n}| \rangle N^{\theta},$$

assuming $N \ge N_0(\alpha, b, d, \rho, \theta, \tau)$.

The next lemma will be referred as the coupling lemma. Its original form dates back to Ref. 7, Lemma 2.4 for $d = 2$. A refined version for arbitrary d was obtained in Ref. 14. In the following form, it is proved in Ref. 12.

Lemma 7.3. Let T be an operator on $\ell^2(\Lambda_0)$, where $\Lambda_0 \in \mathcal{ER}(N) \subset \mathbb{Z}^d$. Let

$$0 < \tau \le b \le \theta < 1, \quad \beta \ge \frac{1-2\tau}{1-\tau},$$

and let

$$N^{\tau} < M_0 < 2N^{\tau}.$$

Assume that Assume that the matrix elements of T satisfy

$$|T(\mathbf{m},\mathbf{n})| \le e^{-\rho|\mathbf{m}-\mathbf{n}|}, \quad \mathbf{m} \ne \mathbf{n}, \quad \mathbf{m},\mathbf{n} \in \Lambda_0, \tag{7.4}$$

and for any $\Lambda \in \mathcal{ER}(L)$with any $L \in (N^{\tau}, N)$, we have

$$\| T|_{\Lambda}^{-1} \| \le e^{L^b}.$$

We will say that an elementary region $\Lambda \in \mathcal{ER}(L)$, $\Lambda \subset \Lambda_0$ is good if we have, in addition to the above,

$$\| T|_{\Lambda}^{-1}(\mathbf{m},\mathbf{n}) \| < e^{-\rho|\mathbf{m}-\mathbf{n}|}, \qquad \textit{for } \mathbf{m},\mathbf{n} \in \Lambda,\ |\mathbf{m}-\mathbf{n}| \ge L^{\beta}.$$

Otherwise Λ is called bad. Assume that for any family $\mathcal{F}$ of pairwise disjoint bad M'-regions in Λ_0 with $M_0+1 \le M' \le 2M_0+1$, we have

$$\#\mathcal{F} < \frac{N^b}{M_0}.$$

Then, we have

$$|T^{-1}(\mathbf{m},\mathbf{n})|\langle e^{-(\rho - N^{-\delta})|\mathbf{m}-\mathbf{n}|}, \qquad \text{for } \mathbf{m},\mathbf{n} \in \Lambda_0,\ |\mathbf{m}-\mathbf{n}|\rangle N^{\theta},$$

where $N \ge N_0(b,d,\tau,\beta)$ and $\delta = \delta(b,d,\tau,\beta) > 0$.

Remark 7.4. In both results, the Töplitz decay condition (7.4) can be replaced by a similar condition with a polynomial factor in front:

$$|T(\mathbf{m},\mathbf{n})| \le (1+|\mathbf{m}-\mathbf{n}|)^p e^{-\rho|\mathbf{m}-\mathbf{n}|}, \quad \mathbf{m} \ne \mathbf{n}, \quad \mathbf{m},\mathbf{n} \in \Lambda_0.$$

In this case, the assumption on largeness of the initial scale will depend on p. The reasoning is based on the fact that the calculations in the proof already involve various (polynomial) combinatorial factors of similar kind which are absorbed into the exponential decay estimates.

Finally, we will need the following matrix-valued Cartan theorem ("Cartan lemma"), see Refs. 4 and 5.

Lemma 7.5. Let $A(\sigma)$ be a self-adjoint $D \times D$ matrix function of a real parameter $\sigma \in [-\delta,\delta]$, satisfying the following conditions:

(1) $A(\sigma)$ is real analytic in σ, and admits a holomorphic extension to the strip $(-\delta_1,\delta_1) + i(-\delta_2,\delta_2)$, satisfying in that strip

$$\|A(z)\| \le B_1.$$

(2) For each $\sigma \in [-\delta_1,\delta_1]$, there is a subset $\Lambda \subset [1,D]$, such that

$$|\Lambda| < D_0, \quad \|\left(R_{[1,N]\setminus\Lambda} A(\sigma) R_{[1,N]\setminus\Lambda}\right)^{-1}\| < B_2.$$

(3) $|\{\sigma \in [-\delta_1,\delta_1] : \|A(\sigma)^{-1}\| > B_3\} < 10^{-3}\delta_2(1+B_1)^{-1}(1+B_2)^{-1}|$.

Then, for any $\varkappa < (1+B_1+B_2)^{-10D_0}$, we have

$$|\{\sigma \in (-\delta_1/2, \delta_1/2) : \|A(\sigma)^{-1}\| > \zeta^{-1}\}| < \exp \frac{c \log \zeta}{D_0 \log (D_0 + B_1 + B_2 + B_3)}. \tag{7.5}$$

C. Proof of proposition 3.2

One can easily check that all the estimates hold for any finite number of scales by choosing a small ε. Therefore, we can assume that the scales are large enough (depending on τ and ν_2). Note also that the estimates are stable under perturbations of size e^{-k^2}, and therefore one can essentially ignore the differences between k and $k+1$.

While the main result is stated with particular numerical values of the parameters, it will be convenient to denote them by letters:

$$\beta = 999/1000, \quad b = 6/7, \quad \rho = 1/300.$$

Assume that the theorem has been proven for the scales $[0,M]$ with the exponent γ. Let C_1 be a large constant chosen later. Let $M_0 \in [0,M]$ and $M_1 = M^{C_1}$. The first step is to "propagate" (3.6) into the scales up to $M_0^{C_1}$. In the process, we would like to obtain stronger bounds: say, with $e^{-M_1^{1000\rho}}$ in the right hand side. Later, these stronger bounds will be needed in order to satisfy the assumptions of Lemma 7.3.

Let $\Lambda \in \mathcal{ER}(M_1)$, and let

$$\Lambda = \cup_\alpha \Lambda_\alpha, \quad \Lambda_\alpha = \Lambda \cap Q_\alpha,$$

$$Q_\alpha \in [-M_0, M_0] + 2M_0\mathbb{Z}^3.$$

One can check that $\Lambda_\alpha \in \mathcal{ER}(M')$ with $M_0 \le M' \le 2M_0$, except for finitely many (bounded by an absolute constant) values of α.

From the induction assumptions, there exists a semi-algebraic subset $\mathcal{A} \subset \mathbb{R} \times [0, 2\pi)$such that, for $(\theta, \varphi) \in \mathbb{R} \times [0, 2\pi) \setminus \mathcal{A}$, we have (3.6) and (3.7) for all elementary regions Λ containing the origin with $\operatorname{diam}(\Lambda) \in [M_0, 2M_0]$. The set $\mathcal{A}$ is of degree at most $M_0^{C_{\text{abs}}}$ and intersects any unit line segment in $\mathbb{R} \times [0, 2\pi)$in a subset of one-dimensional measure at most $M_1^{C_{\text{abs}}} e^{-M_0^{\rho}}$.

By Theorem 2.6 ("arithmetic lemma"), and assuming that M_0 is large enough (depending on the Diophantine constants of (ω, λ)), we have

$$\#\{\mathbf{n} \in \mathbb{Z}^3 : (\theta + \mathbf{n} \cdot \omega, \varphi + n_3\lambda) \in \mathcal{A}\} \le M_0^{C_{\text{abs}}} M_1^{3/4+\delta}, \tag{7.6}$$

where δ can be made arbitrary small depending on the Diophantine constants of (ω, λ). Let us call an elementary region Λ_α good if all points from Λ_α are outside of the above set. Let

$$\Lambda_{\text{good}} = \Lambda_{\text{good}}(\theta, \varphi) = \bigcup \{\Lambda_\alpha \subset \Lambda : \Lambda_\alpha \text{ is good; } \operatorname{diam}(\Lambda_\alpha \in [M_0, 2M_0])\}.$$

Standard application of the covering Lemma 7.2 implies, after absorbing the constants,

$$\|T(\theta, \varphi)|_{\Lambda_{\text{good}}}^{-1}\| \le e^{(3M_0)^b},$$

and we have, from the definition of Λ_{good},

$$|\Lambda \setminus \Lambda_{\text{good}}| \le M_0^{C_{\text{abs}}} M_1^{3/4+\delta}.$$

We will also need to introduce an additional scale

$$M_0' = \lfloor (10M_0)^{1/\rho} \rfloor.$$

Similar considerations with boxes of size M_0' imply that, outside of a semi-algebraic subset of $[0, 2\pi) \times \mathbb{R}$ with sectional measures bounded by $e^{-5M_0'}$, we have

$$\|T(\theta, \varphi)|_{\Lambda}^{-1}\| \le e^{(3M_0')^b}.$$

Let L be a unit line segment in $\mathbb{R} \times [0, 2\pi)$, and $\theta(t), \varphi(t)$ be the natural parametrization of L (by length). Let $T(t) := T(\theta(t), \varphi(t))$. Apply Cartan's lemma to the family $T(t)$, with:

(1) $B_1 = 10M_1$.
(2) $D_0 = M_0^{C_{\text{abs}}} M_1^{3/4+\delta}$ (the cardinality of the bad set).
(3) $B_2 = e^{(3M_0)^b}$ (upper bound on the inversion of restriction to the good set).
(4) $B_3 = e^{(3M_0')^b}$. This way, Assumption (3) of the Cartan's lemma reduces to $e^{-(M_0')^{\rho}} \sim e^{-10M_0} \ll e^{-(4M_0)^b}$ and therefore holds for large scales.

Then, for

$$\varkappa < e^{-M_0^{C_{\text{abs}}} M_1^{3/4+\delta}},$$

the measure estimate in (7.5) becomes

$$\exp \frac{c \log \zeta}{M_0^{C_{\text{abs}}+b/\rho} M_1^{3/4+\delta}}.$$

In order to obtain the inductive conclusion, we choose $\zeta = e^{-M_1^b}$. This is possible assuming

$$\frac{C_{\text{abs}}}{C_1} + \frac{3}{4} + \delta < b.$$

This reduces the estimate of the measure in (7.5) to

$$\exp\{-cM_1^{b - \frac{3}{4} - \delta - \frac{b}{\rho C_1} - \frac{C_{\text{abs}}}{C_1}}\}.$$

To summarize, we start from assuming the proposition for an interval of scales

$$[M_0, (10M_0)^{1/\rho}]$$

and conclude that, for the scale $M_1 = M_0^{C_1}$, the first estimate (3.6) holds outside of a subset of section measure estimated by

$$\exp\{-cM_1^{b-\frac{3}{4}-\delta-\frac{b}{\rho C_1}-\frac{C_{\mathrm{abs}}}{C_1}}\}.$$

We would like $M_1 > M_0^{C_1}$, so that the conclusion is meaningful, and also get $e^{-M_1^{1000\rho}}$ in the last estimate. Overall, this leads to the following assumptions:

$$b - \frac{3}{4} - \delta - \frac{b}{\rho C_1} - \frac{C_{\mathrm{abs}}}{C_1} > 1000\rho; \qquad \frac{C_{\mathrm{abs}}}{C_1} + \frac{3}{4} + \delta < b; \quad C_1\rho > 10.$$

These assumptions can be satisfied as long as $b > 3/4$, by first requiring the Diophantine properties to imply $0 < \delta < b - 3/4$, then choosing a large C_1 so that the left hand side of the first inequality becomes positive (say, for $\rho = 1$), and then choosing a large ρ to satisfy the inequality.

Once the above parameters are fixed, one needs to additionally assume that the scales are large enough (depending on these parameters and the Diophantine constants of the frequency vector). This is achievable by taking a small ε. Note that one can also achieve same results for range of constants C_1 (say, $[C_1, C_2]$) instead of an individual value. The choice of large M_0 and then ε will then depend on C_1 and C_2.

The next step is to propagate the exponential decay bounds. Let us start from the same assumption as before: that LDT is proven in the range of scales $[M_0, (10M_0)^{1/\rho}]$ with the exponent γ. Let $M_1 = M_0^{C_1}$, where C_1 satisfies the constraints from the previous part (we may impose more conditions on it later). Our goal is to prove an exponential decay bound on the scale M_1. Due to application of Lemma 7.3 ("coupling lemma"), the exponent γ will change, and we will need to keep track of that. We will use an auxiliary scale $M \sim M_1^{\tau} \sim M_0^{C_1\tau}$ (here $\sim$ denotes nearest integer) and apply Lemma 7.3 with $N = M_1$.

Similarly to the previous step, define $\mathcal{A}$ to be a semi-algebraic subset with sectional measures at most $e^{-M_0^{\rho}/2}$, such that, outside of $\mathcal{A}$ we have all elementary regions containing the origin of sizes between M_0 and $2M_0$ satisfy (3.6) and (3.7). Note that the combinatorial factors are absorbed into the deterioration of exponent from $-M_0^{\rho}$ to $-M_0^{\rho}/2$. Consider now elementary regions of size M inside of the M_1-region. If an M-region does not have any bad M_0-regions inside of it, one can apply Lemma 7.2 and conclude that the M-box satisfies the exponential decay estimates with the exponent $\gamma - (\log M)^{-50}$. The number of disjoint bad M-regions inside of M_1-region is bounded by

$$M_0^{C_{\mathrm{abs}}} M_1^{3/4+\delta} \sim \frac{M_1^{3/4+\delta+C_{\mathrm{abs}}/C_1+\tau}}{M}.$$

Let us assume

$$3/4 + \delta + C_{\mathrm{abs}}/C_1 + \tau \le b.$$

This will satisfy the sub-linear bound in the assumptions of Lemma 7.3. To satisfy the remaining assumption, we need to enforce (3.6) on all scales between M and M_1 using the previous part. This will require removing a subset of sectional measure at most

$$e^{-M^{999\rho}} \sim e^{-M_1^{999\rho\tau}}.$$

In order to be consistent with the conclusion of LDT, we would thus like to have $\tau > 1/999$. As long as $\delta + 3/4 + 1/999 < 1$, this is possible to achieve by choosing b and τ.

Remark 7.6. We can now provide more details on the relation between ε_0 and τ. In the induction step, the choice of the large scale in relation to τ was made in (7.6), in order for the factor M^{δ} to absorb $C(\omega, \lambda) = \tau^{-c_{\mathrm{abs}}}$ in Theorem 2.6 (see also Remark 2.7). Once the choice of the parameters ρ, b, β above is fixed and an appropriate C_1 is selected, it ultimately forces the initial scale M_0 to satisfy $M_0 \ge \tau^{-C_{\mathrm{abs}}}$. This choice of initial scale forces $\varepsilon_0 < M^{-10} e^{-M_0^b}$, which implies the claim in Remark 3.3. Here, C_{abs} and c_{abs} are absolute constants.

ACKNOWLEDGMENTS

I.K. was supported by NSF Grant Nos. DMS–1846114, DMS–2052519, and the 2022 Sloan Research Fellowship. W.L. was supported by NSF Grant Nos. DMS–2000345, DMS–2052572 and DMS–2246031. W.-M.W acknowledges the support from the CY Initiative of Excellence, "Investissements d'Avenir" Grant No. ANR–16–IDEX–0008.

The authors would like to thank the anonymous referee for important remarks which, hopefully, led to significant improvements in the readability of the text (in particular, see Remark 3.4).

During their careers, the authors were inspired by the works of Abel Klein, in particular, his deep contributions to the multi-scale analysis. This paper is dedicated to Professor Klein's Birthday.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ilya Kachkovskiy: Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Wencai Liu**: Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Wei-Min Wang**: Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1] Basu, S., "On bounding the Betti numbers and computing the Euler characteristic of semi-algebraic sets," Discrete Comput. Geom. **22**(1), 1–18 (1999).
[2] Binyamini, G. and Novikov, D., "Complex cellular structures," Ann. Math. **190**(1), 145–248 (2019).
[3] Bourgain, J., "Quasi-periodic solutions of Hamiltonian perturbations of 2d linear Schrödinger equations," Ann. Math. **148**, 363–439 (1998).
[4] Bourgain, J., in *Green's Function Estimates for Lattice Schrödinger Operators and Applications*, *Annals of Mathematics Studies* (Princeton University Press, Princeton, NJ, 2005), Vol. 158.
[5] Bourgain, J., "Anderson localization for quasi-periodic lattice Schrödinger operators on $\mathbb{Z}^d$, d arbitrary," Geom. Funct. Anal. **17**(3), 682–706 (2007).
[6] Bourgain, J. and Kachkovskiy, I., "Anderson localization for two interacting quasiperiodic particles," Geom. Funct. Anal. **29**(1), 3–43 (2019).
[7] Bourgain, J. and Wang, W.-M., "Quasi-periodic solutions of nonlinear random Schrödinger equations," J. Eur. Math. Soc. **10**, 1–45 (2008).
[8] Bourgain, J., Goldstein, M., and Schlag, W., "Anderson localization for Schrödinger operators on $\mathbb{Z}^2$ with quasi-periodic potential," Acta Math. **188**, 41–86 (2002).
[9] Dodson, B., Soffer, A., and Spencer, T., "The nonlinear Schrödinger equation on $\mathbb{Z}$ and $\mathbb{R}$ with bounded initial data: Examples and conjectures," J. Stat. Phys. **180**, 910–934 (2020).
[10] Eliasson, L. H. and Kuksin, S., "KAM for the nonlinear Schrödinger equation," Ann. Math. **172**(1), 371–435 (2010).
[11] Geng, J., Sun, Y., and Wang, W.-M., "Quasiperiodic solutions to nonlinear random Schrödinger equations at fixed potential realizations," J. Math. Phys. **64**(3), 032701 (2023).
[12] He, C., Shi, J., Shi, Y., and Yuan, X., "On linear stability of KAM tori via the Craig–Wayne–Bourgain method," arXiv:2003.01487 (2022).
[13] Jitomyrskaya, S., Liu, W., and Shi, Y., "Anderson localization for multi-frequency quasi-periodic operators on $\mathbb{Z}^d$," Geom. Funct. Anal. **30**(2), 457–481 (2020).
[14] Liu, W., "Quantitative inductive estimates for Green's functions of non-self-adjoint matrices," Anal. PDE **15**(8), 2061–2108 (2022).
[15] Liu, W. and Wang, W.-M., "Nonlinear Anderson localized states at arbitrary disorder," arXiv:2201.00173 (2022).
[16] Procesi, C. and Procesi, M., "A KAM algorithm for the resonant non-linear Schrödinger equation," Adv. Math. **272**, 399–470 (2015).
[17] Shi, Y. and Wang, W.-M., "Anderson localized states for the quasi-periodic nonlinear wave equation on $\mathbb{Z}^d$," arXiv:2306.00513 (2023).
[18] Wang, W.-M., "Energy supercritical nonlinear Schrödinger equations: Quasiperiodic solutions," Duke Math. J. **165**(6), 1129–1192 (2016).
[19] Wang, W.-M., "Space quasi-periodic standing waves for nonlinear Schrödinger equations," Commun. Math. Phys. **378**, 783–806 (2020).
[20] Wang, W.-M., "Semi-algebraic sets method in PDE and mathematical physics," J. Math. Phys. **62**, 021506 (2021).
[21] Wang, W.-M., "Infinite energy quasi-periodic solutions to the nonlinear Schrödinger equations on $\mathbb{R}$," Int. Math. Res. Not. **2023**, 7201.