Localization for quasiperiodic Schrödinger operators with multivariable Gevrey potential functions

Silvius Klein

Abstract. We consider an integer lattice quasiperiodic Schrödinger operator. The underlying dynamics is either the skew-shift or the multi-frequency shift by a Diophantine frequency. We assume that the potential function belongs to a Gevrey class on the multidimensional torus. Moreover, we assume that the potential function satisfies a generic transversality condition, which we show to imply a Łojasiewicz type inequality for smooth functions of several variables. Under these assumptions and for large coupling constant, we prove that the associated Lyapunov exponent is positive for all energies, and continuous as a function of energy, with a certain modulus of continuity. Moreover, in the large coupling constant regime and for an asymptotically large frequency - phase set, we prove that the operator satisfies Anderson localization.

Contents

1. Definitions, notations, statement of main results

In this paper we study the one-dimensional lattice quasiperiodic Schrödinger operator $H(\underline{x})$ acting on $l^2(\mathbb{Z})$ by

$$[H(\underline{x})\,\psi]_n := -\psi_{n+1} - \psi_{n-1} + \lambda\,v(\mathrm{T}^n\underline{x})\,\psi_n \qquad (1.1)$$

where, in equation (1.1),

- $\underline{x} = (x_1, x_2) \in \mathbb{T}^2$ is a parameter that introduces some randomness into the system;

- λ is a real number called the disorder of the system;

- $v(\underline{x})$ is a real valued function on $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$, that is, a real valued 1-periodic function in each variable;

- T is a specific ergodic transformation on $\mathbb{T}^2$, and T^n is its nth iteration.

Some of the questions of interest regarding this, or other related operators, are the spectral types (pure point, absolutely continuous, singularly continuous), the topological structure of the spectrum, the rate of decay of the eigenfunctions, the positivity and regularity of the Lyapunov exponent, the regularity of the integrated density of states.

Due to the ergodicity of the transformation T, the spectrum and the spectral types of the Hamiltonian system $[H(\underline{x})]_{\underline{x}\in\mathbb{T}^2}$ defined by (1.1) are not random - that is, they are independent of $\underline{x}$ almost surely (see [9]).

A stronger property than pure point spectrum is Anderson localization, which for the physical model indicates an insulating behavior, while a purely absolutely continuous spectrum indicates metallic (conductive) behavior.

Let us describe these concepts more formally.

Definition 1.1. An operator satisfies Anderson localization (AL) if it has pure point spectrum with exponentially decaying eigenfunctions.

Consider now the Schrödinger equation

$$H(\underline{x})\psi = E\psi, \qquad (1.2)$$

for $\psi = [\psi_n]_{n\in\mathbb{Z}} \subset \mathbb{R}$ and $E \in \mathbb{R}$.

Due to Schnol–Simon's theorem (see [9]), to prove AL it is enough to show that every extended state is exponentially decaying. In other words, if ψ is a formal solution to the Schrödinger equation (1.2) and if ψ grows at most polynomially $|\psi_n| \lesssim |n|$ then ψ decays exponentially: $|\psi_n| \lesssim e^{-c|n|}$.

The Schrödinger equation (1.2) is a second order finite differences equation,

$$-\psi_{n+1} - \psi_{n-1} + \lambda\, v(\mathrm{T}^n \underline{x})\, \psi_n = E\, \psi_n,$$

which becomes

$$\begin{bmatrix} \psi_{n+1} \\ \psi_n \end{bmatrix} = M_N(\underline{x}, E) \begin{bmatrix} \psi_1 \\ \psi_0 \end{bmatrix},$$

where

$$M_N(\underline{x}, E) = M_N(\underline{x}, \lambda, E) := \prod_{j=N}^{1} \begin{bmatrix} \lambda v(\mathrm{T}^j \underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix}$$

is called the transfer (or fundamental) matrix of (1.1).

Define further the function

$$L_N(\underline{x}, E) = L_N(\underline{x}, \lambda, E) := \frac{1}{N} \log \|M_N(\underline{x}, E)\|$$

and its mean

$$L_N(E) = L_N(\lambda, E) := \int_{\mathbb{T}^2} \frac{1}{N} \log \|M_N(\underline{x}, E)\| \, d\underline{x}$$

Due to sub-additivity, the sequence $L_N(E)$ converges.

Definition 1.2. The limit

$$L(E) := \lim_{N \to \infty} L_N(E)$$

is called *the Lyapunov exponent* of (1.1) and it measures the average exponential growth of the transfer matrices.

Ergodicity in fact implies that for a.e. $\underline{x} \in \mathbb{T}^2$,

$$L(E) := \lim_{N \to \infty} \frac{1}{N} \log \|M_N(\underline{x}, E)\|. \tag{1.3}$$

Note that since the transfer matrices have determinant 1, the Lyapunov exponent is always nonnegative. An important question is whether it is in fact (uniformly) bounded away from 0. This would imply, due to Kotani's theorem, absence of absolutely continuous spectrum, and it would represent a strong indication for pure point spectrum. This is also usually the assumption under which strong continuity properties hold and an analysis of the topological structure of the spectrum is more feasible.

In this paper we prove Anderson localization and positivity and continuity of the Lyapunov exponent for large coupling constant, for certain ergodic transformations T on $\mathbb{T}^2$ and under certain regularity and transversality conditions on the potential function $v(\underline{x})$. While all of our results are stated and proven for the two-dimensional torus $\mathbb{T}^2$, their analogues on the higher dimensional torus hold as well.

We describe the assumptions on the transformation and on the potential function.

We start with some notations: for a multi-index $\underline{m} = (m_1, m_2) \in \mathbb{Z}^2$, we write $|\underline{m}| := |m_1| + |m_2|$ and if $\underline{m} = (m_1, m_2) \in \mathbb{N}^2$, then $\underline{m}! := m_1! \cdot m_2!$. Moreover, for $\underline{\alpha} = (\alpha_1, \alpha_2)$ and $\underline{m} = (m_1, m_2) \in \mathbb{N}^2$, we write $\underline{\alpha} \leq \underline{m}$ when $\alpha_1 \leq m_1$ and $\alpha_2 \leq m_2$.

Throughout this paper, the transformation T: $\mathbb{T}^2 \to \mathbb{T}^2$ will represent

■ either the *skew-shift*

$$S_\omega (x_1, x_2) := (x_1 + x_2, x_2 + \omega), \tag{1.4}$$

where $\omega \in \mathbb{T}$ is irrational and its nth iteration is given by

$$S_\omega^n (x_1, x_2) = \left(x_1 + nx_2 + \frac{n(n-1)}{2}\omega, \ x_2 + n\omega \right), \tag{1.5}$$

■ or the *multi-frequency shift*

$$T_{\underline{\omega}} (x_1, x_2) := (x_1 + \omega_1, x_2 + \omega_2), \tag{1.6}$$

where $\underline{\omega} = (\omega_1, \omega_2) \in \mathbb{T}^2$ and ω_1, ω_2 are rationally independent and its nth iteration is given by

$$T_{\underline{\omega}}^n (\underline{x}) = \underline{x} + n \underline{\omega} = (x_1 + n\omega_1, x_2 + n\omega_2). \tag{1.7}$$

The irrationality/rational independence of the frequency ensures that the corresponding transformation is ergodic. However, we need to make a quantitative assumption on this rational independence, for reasons that will be described later.

We say that the frequency $\omega \in \mathbb{T}$ *satisfies a Diophantine condition* DC_κ for some $\kappa > 0$ if

$$\mathrm{dist} \, (l\omega, \mathbb{Z}) =: \|l\, \omega\| > \kappa \cdot \frac{1}{|l|[\log(1 + |l|)]^2}, \quad \text{for all } \ l \in \mathbb{Z}. \setminus \{0\}. \tag{1.8}$$

We say that the (multi)frequency $\underline{\omega} \in \mathbb{T}^2$ *satisfies a Diophantine condition* DC_κ for some $\kappa > 0$ and a fixed constant $A > 2$, if

$$\|\underline{l} \cdot \underline{\omega}\| := \|l_1 \omega_1 + l_2 \omega_2\| > \kappa \cdot \frac{1}{|\underline{l}|^A}, \quad \text{for all } \underline{l} \in \mathbb{Z}^2 \setminus \{(0,0)\}. \tag{1.9}$$

Note that the set of frequencies which satisfy either (1.8) or (1.9) has measure $1 - O(\kappa)$, hence almost every frequency satisfies such a Diophantine condition DC_κ for some $\kappa > 0$.

When necessary, to emphasize the dependence of the operator on the frequency, we will use the notation $H_\omega(\underline{x})$ or $H_{\underline{\omega}}(\underline{x})$ respectively.

Now we describe the assumptions on the potential function $v(\underline{x})$.

- We say that a C^∞ function $v(\underline{x})$ on $\mathbb{T}^2$ *belongs to the Gevrey class* $G^s(\mathbb{T}^2)$ for some $s > 1$ if its partial derivatives have the following bounds:

$$\sup_{\underline{x} \in \mathbb{T}^2} |\partial^{\underline{m}} v(\underline{x})| \le M K^{|\underline{m}|} (\underline{m}!)^s, \quad \text{for all } \underline{m} \in \mathbb{N}^2, \tag{1.10}$$

  for some constants $M, K > 0$.

This condition is equivalent (see the exercises from Chapter IV in [15]) to the following exponential-type decay of the Fourier coefficients of v:

$$|\hat{v}(\underline{l})| \le M e^{-\rho|\underline{l}|^{1/s}} \quad \text{for all } \underline{l} \in \mathbb{Z}^2 \tag{1.11}$$

for some constants $M, \rho > 0$, where

$$v(\underline{x}) = \sum_{\underline{l} \in \mathbb{Z}^2} \hat{v}(\underline{l}) \, e^{2\pi i \, \underline{l} \cdot \underline{x}}.$$

Note from (1.10) or (1.11) with $s = 1$ that the Gevrey class $G^1(\mathbb{T}^2)$ is the class of real analytic functions on $\mathbb{T}^2$.

Note also that $s_1 < s_2 \implies G^{s_1}(\mathbb{T}^2) \subset G^{s_2}(\mathbb{T}^2)$, so the greater the order of the Gevrey class, the larger the class.

The Gevrey-class of any order $s > 1$ is an intermediate Carleman class of functions between analytic functions and C^∞ functions. They are not, however, quasi-analytic (one can easily construct examples or use a general test for quasi-analyticity of Carleman classes, as in Chapter V.2 in [15]).

We will then impose on our potential function v the following generic transversality condition (TC).

- We say that a function $v(\underline{x})$ is *transversal* if v is not flat at any point:

  for any $\underline{x} \in \mathbb{T}^2$ there is $\underline{m} \in \mathbb{N}^2$, $|\underline{m}| \neq 0$ such that $\partial^{\underline{m}} v(\underline{x}) \neq 0$.　　(1.12)

Non-constant analytic functions automatically satisfy (1.12). A Schrödinger operator with potential given by a function which satisfies the Gevrey regularity condition (1.10) and the transversality condition (1.12) is a natural extension of the non constant analytic case considered in [5] and [6].

We are ready to formulate the main result of this paper.

Theorem 1.1. *Consider the Schrödinger operator* (1.1) *on* $l^2(\mathbb{Z})$,

$$[H(\underline{x})\,\psi]_n := -\psi_{n+1} - \psi_{n-1} + \lambda\,v(\mathrm{T}^n \underline{x})\,\psi_n,$$

where the transformation T *is either the skew-shift* (1.4) *or the multi-frequency shift* (1.6). *Assume that for some* $\kappa > 0$ *the underlying frequency satisfies the Diophantine condition* DC_κ *described in* (1.8) *or* (1.9) *respectively.*

Assume moreover that the potential function $v(\underline{x})$ *belongs to a Gevrey class* $G^s(\mathbb{T}^2)$ *and that it is transversal as in* (1.12).

There is $\lambda_0 = \lambda_0(v, \kappa)$ *such that the following hold.*

- *If* $|\lambda| \geq \lambda_0$, *the Lyapunov exponent is positive for all energies* $E \in \mathbb{R}$:

$$L(E) \geq \frac{1}{4}\log|\lambda| > 0.　　(1.13)$$

- *If* $|\lambda| \geq \lambda_0$, *the Lyapunov exponent* $L(E)$ *is a continuous functions of the energy* E, *with modulus of continuity on any compact interval* $\mathcal{E}$ *at least,*

$$h(t) = C\,e^{-c|\log t|^\eta},　　(1.14)$$

where $C = C(\mathcal{E}, \lambda, v, \kappa, s)$ *and* c, η *are some positive absolute constants.*

- *Let* $\mathrm{T} = \mathrm{S}_\omega$ *be the skew-shift. For every* λ *with* $|\lambda| \geq \lambda_0$, *there is an exceptional set* $\mathcal{B} = \mathcal{B}_\lambda \subset \mathbb{T}^3$, *with* mes $\mathcal{B} < \kappa$, *such that for all* $(\omega, \underline{x}) \notin \mathcal{B}$, *the operator* $H_\omega(\underline{x})$ *satisfies Anderson localization.*

- *Let* $\mathrm{T} = \mathrm{T}_{\underline{\omega}}$ *be the multi-frequency shift. Fix* $\underline{x_0} \in \mathbb{T}^2$ *and* λ *with* $|\lambda| \geq \lambda_0$. *Then for a.e. multi-frequency* $\underline{\omega} \in \mathrm{DC}_\kappa$, *the operator* $H_{\underline{\omega}}(\underline{x_0})$ *satisfies Anderson localization.*

2. Summary of related results, general strategy

The results in this paper extend the ones in [6] and [5] (see also J. Bourgain's monograph [4]) from non-constant real analytic potential functions, to the more general class of Gevrey potential functions satisfying a transversality condition. They also mirror similar results obtained for the one-frequency shift model on the torus $\mathbb{T}$ (see [16]).

It should be noted, however, that unlike the one or multi-frequency shift, the skew-shift, due to its weekly mixing properties, is expected to behave more like the random model (presumably regardless of the regularity of the potential). In other words, for the skew-shift, these results are expected to be independent of the size of the disorder λ. Hence one expects that if $\lambda \neq 0$, the Lyapunov exponent is positive and Anderson localization holds for all energies. Moreover, one expects no gaps in the spectrum (unlike in the one-frequency shift case, when the spectrum is a Cantor set) - see the comments at the end of Chapter 15 in [4]. Some results on these very challenging problems have been obtained in [2], [3], [18], and [19].

Localization results for these types of operators defined by skew-shift dynamics have applications to quantum chaos problems. More specifically, they imply existence of almost periodic solutions to the quantum kicked rotator equation. However, one has to establish (dynamical) localization for a more general, long range operator, one where the discrete Laplacian is replaced by a Toeplitz operator with fast off-diagonal decay of its monodromy matrix entries. This was already established for analytic potential functions (see Chapter 15 and 16 in [4]), but we will not address this problem for Gevrey potential functions in this paper.

Most of the results on localization for discrete quasiperiodic Schrödinger operators (with either shift or skew-shift dynamics) have been obtained under the assumption that the potential function is the cosine function, or a trigonometric polynomial or a real analytic and non-constant function (see J. Bourgain's monograph [4]).

Assuming Gevrey regularity and a transversality condition, there are localization results for the shift model that closely resemble the ones in the analytic case (see [11], [16]). It should be noted, however, that they are usually perturbative and that more subtle results regarding fine continuity properties of the integrated density of states (as in [12], [13]) or the topological structure of the spectrum (as in [14]) are not available in this context.

For potential functions that are more general than Gevrey, namely C^{α}, the results available now (on localization and positivity of the Lyapunov exponent) require that a(n asymptotically small relative to the size λ of the disorder but)

positive set of energies be excluded or that the potential function be replaced by some generic variations of itself (see [1], [7], [8]).

To prove Theorem 1.1 we will follow the same strategy used in [16] for the single frequency shift model: at each scale, substitute the potential function by an appropriate polynomial approximation (see Section 3). This in turn will allow the use of subharmonic functions techniques (see Section 4) developed in [5], [6], [4]. An additional challenge is describing the transversality condition (1.12) for multi-variable smooth functions in a quantitative way. We derive (see Section 5) a Łojasiewicz type inequality for such functions, of the kind previously available for non-constant trigonometric polynomials (see [17], [10]) or analytic functions (see [21], [12]).

The main technical result of this paper, from which all statements in Theorem 1.1 follow, is a large deviation theorem (LDT) for logarithmic averages of transfer matrices (see Section 6).

According to (1.3), due to ergodicity, for a.e. $\underline{x} \in \mathbb{T}^2$,

$$\frac{1}{N} \log \|M_N(\underline{x}, E)\| \to L(E) \quad \text{as } N \to \infty.$$

The LDT provides a quantitative version of this convergence:

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 : \left|\frac{1}{N} \log \|M_N(\underline{x}, E)\| - L_N(E)\right| > \epsilon\right] < \delta(N, \epsilon), \qquad (2.1)$$

where $\epsilon = o(1)$ and $\delta(N, \epsilon) \to 0$ as $N \to \infty$

The size of the deviation ϵ and the measure of the exceptional set $\delta(N, \epsilon)$ are very important. The sharpest such estimate (see Theorem 7.1 in [12]), available for the single-frequency shift model with analytic potential, holds for any $\epsilon > 0$ and exponentially small measure $\delta(N, \epsilon) \approx e^{-c\delta N}$, thus morally matching the large deviation estimates for random variables that these deterministic quantities mimic here. Having such sharp estimates leads to a sharper modulus of continuity of the Lyapunov exponent (see [12]).

For the multi-frequency shift and the skew-shift models, even with analytic potentials, the available estimates are not as sharp. In this paper, for Gevrey potential functions, we will get $\epsilon \approx N^{-\tau}$ and $\delta \approx e^{-N^{\sigma}}$ for some absolute constants $\tau, \sigma \in (0, 1)$.

Following the approach in [4], [6], a large deviation estimate like (2.1) will allow us to obtain a lower (positive) bound and continuity of the Lyapunov exponent, once these properties are established at an initial scale N_0 for $L_{N_0}(E)$. It will also allow us (the reader will be referred to [4], [6] for details) to establish estimates on

the Green's functions associated with the operator (1.1), more specifically the fact that double resonances for Green's functions occur with small probability, which leads to Anderson localization.

Most of the paper will then be devoted to proving a LDT like (2.1):

$$\text{mes} \left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{N} \log \| M_N(\underline{x}, E) \| - L_N(E) \right| > N^{-\tau} \right] < e^{-N^\sigma} \qquad (2.2)$$

through an inductive process on the scale N.

The *base step* of the inductive process for proving the LDT (2.2) is based exclusively on the transversality condition (1.12) on the potential, and on choosing a sufficiently large disorder λ. The latter is what makes this approach perturbative (and, in the case of the skew-shift model, wasteful, since it does not exploit the weakly-mixing properties of its dynamics). The former implies a Łojasiewicz type inequality, which we prove using a quantitative form of the implicit function theorem.

In the *inductive step* we use the regularity of the potential function $v(\underline{x})$ and the arithmetic properties of the frequency. The regularity of $v(\underline{x})$ allows us to approximate it efficiently by trigonometric polynomials $v_N(\underline{x})$ at each scale N, and to use these approximants in place of $v(\underline{x})$ to get analytic substitutes $\widetilde{M}_N(\underline{x})$ for the transfer matrices $M_N(\underline{x})$. Their corresponding logarithmic averages will be subharmonic in each variable which will allow us to employ the subharmonic functions techniques developed in [4], [5], [6].

The main technical difficulty with this approach, and what restricts it to Gevrey (instead of say, C^α) potential functions, is that the holomorphic extensions of the transfer matrix substitutes $\widetilde{M}_N(\underline{x})$ will have to be restricted to domains of size $\approx N^{-\delta}$ for some $\delta > 0$. In other words, the estimates will not be uniform in N, and this decreasing width of the domain of holomorphicity will have to be over-powered. This will not be possible for a C^α potential function because its trigonometric polynomial approximation is less efficient, so the width of holomorphicity in this case will decrease too fast (exponentially fast).

3. Description of the approximation process

Let $v \in G^s(\mathbb{T}^2)$ be a Gevrey potential function. Then

$$v(\underline{x}) = \sum_{\underline{l} \in \mathbb{Z}^2} \hat{v}(\underline{l}) e^{2\pi i \, \underline{l} \cdot \underline{x}}, \qquad (3.1)$$

where for some constants $M, \rho > 0$, its Fourier coefficients have the decay:

$$|\hat{v}(\underline{l})| \le M e^{-\rho|\underline{l}|^{1/s}}, \quad \text{for all } \underline{l} \in \mathbb{Z}^2. \tag{3.2}$$

We will compare the logarithmic averages of the transfer matrix

$$
\begin{aligned}
L_N(\underline{x}, E) &= \frac{1}{N} \log \|M_N(\underline{x}, E)\| \, d\underline{x} \\
&= \frac{1}{N} \log \Big\| \prod_{j=N}^{1} \Big[\begin{matrix} \lambda v(\mathrm{T}^j \underline{x}) - E & -1 \\ 1 & 0 \end{matrix} \Big] \Big\|,
\end{aligned}
\tag{3.3}
$$

with their means

$$L_N(E) = \int_{\mathbb{T}^2} L_N(\underline{x}, E) \, d\underline{x}. \tag{3.4}$$

To be able to use subharmonic functions techniques, we will have to approximate the potential function $v(\underline{x})$ by trigonometric polynomials $v_N(\underline{x})$ and substitute v by v_N into (3.3). At each scale N we will have a different approximant v_N chosen in such a way that the "transfer matrix substitute" would be close to the original transfer matrix. The approximant v_N will then have to differ from v by a very small error - (super)exponentially small in N. That, in turn, will make the degree $\deg v_N =: \tilde{N}$ of this polynomial very large - based on the rate of decay (1.11) of the Fourier coefficients of v, $\tilde{N}$ should be a power of N, dependent on the Gevrey class s.

The trigonometric polynomial $v_N(\underline{x})$ has an extension $v_N(\underline{z})$, $\underline{z} = (z_1, z_2)$, which is separately holomorphic on the whole complex plane in each variable. We have to restrict $v_N(\underline{z})$ in each variable to a narrow strip (or annulus, if we identify the torus $\mathbb{T}$ with $\mathbb{R}/\mathbb{Z}$) of width ρ_N, where $\rho_N \approx (\deg v_N)^{-1} \approx \tilde{N}^{-1} \approx N^{-\theta}$, for some power $\theta > 0$. This is needed in order to get a uniform in N bound on the extension $v_N(\underline{z})$. Moreover, in the case of the skew-shift, this is also needed because its dynamics expands in the imaginary direction, and in this case, the width of holomorphicity in the second variable will have to be smaller than in the first by a factor of $\approx \frac{1}{N}$.

The fact that the "substitutes" $v_N(\underline{x})$ have different, smaller and smaller widths of holomorphicity creates significant technical problems compared to the case when $v(\underline{x})$ is a real analytic function. It also makes this approach fail when the rate of decay of the Fourier coefficients of the potential function $v(\underline{x})$ is slower.

Therefore, we have to find the optimal "error vs. degree" approximations of $v(\underline{x})$ by trigonometric polynomials $v_N(\underline{x})$. Here are the formal calculations.

For every positive integer N, consider the truncation

$$v_N(\underline{x}) := \sum_{|\underline{l}| \leq \widetilde{N}} \hat{v}(\underline{l}) \, e^{2\pi i \, \underline{l} \cdot \underline{x}}, \tag{3.5}$$

where $\widetilde{N} = \deg v_N$ will be determined later.

Since $v_N(x_1, x_2)$ is in each variable a 1-periodic, real analytic function on $\mathbb{R}$, it can be extended to a separately in each variable 1-periodic holomorphic function on $\mathbb{C}$,

$$v_N(\underline{z}) := \sum_{|\underline{l}| \leq \widetilde{N}} \hat{v}(\underline{l}) e^{2\pi i \, \underline{l} \cdot \underline{z}}. \tag{3.6}$$

To ensure the uniform boundedness in N of $v_N(z_1, z_2)$ we have to restrict $v_N(z_1, z_2)$ to the annulus/strip $[|\Im z| < \rho_{1,N}] \times [|\Im z| < \rho_{1,N}]$, where

$$\rho_{1,N} := \frac{\rho}{2} \widetilde{N}^{-1+1/s}.$$

Indeed, if $z_1 = x_1 + iy_1$, $z_1 = x_1 + iy_1$ and $|y_1|, |y_2| < \rho_{1,N}$, then

$$|v_N(z_1, z_2)| = \Big| \sum_{|\underline{l}| \leq \widetilde{N}} \hat{v}(\underline{l}) e^{2\pi i \, \underline{l} \cdot \underline{z}} \Big|$$

$$\leq \sum_{|\underline{l}| \leq \widetilde{N}} |\hat{v}(\underline{l})| e^{-2\pi \, \underline{l} \cdot \underline{y}}$$

$$\leq M \sum_{|\underline{l}| \leq \widetilde{N}} e^{-\rho |\underline{l}|^{1/s}} e^{|l_1||y_1| + |l_2||y_2|}$$

$$\leq M \sum_{|\underline{l}| \leq \widetilde{N}} e^{-\rho |\underline{l}|^{1/s}} e^{|\underline{l}| \rho_{1,N}}$$

$$\leq M \sum_{|\underline{l}| \leq \widetilde{N}} e^{-\rho |\underline{l}|^{1/s}} \cdot e^{|\underline{l}| \rho/2 |\underline{l}|^{-1+1/s}}$$

$$= M \sum_{|\underline{l}| \leq \widetilde{N}} e^{-\frac{\rho}{2} |\underline{l}|^{1/s}}$$

$$\leq M \sum_{\underline{l} \in \mathbb{Z}^2} e^{-\frac{\rho}{2} |\underline{l}|^{1/s}} =: B < \infty,$$

where B is a constant which depends on v (not on the scale N) and we have used $|y_1|, |y_2| < \rho_{1,N} = \frac{\rho}{2} \widetilde{N}^{-1+1/s} \leq \frac{\rho}{2} |\underline{l}|^{-1+1/s}$ for $|\underline{l}| \leq \widetilde{N}$, since $s > 1$.

We also clearly have $|v(\underline{x}) - v_N(\underline{x})| \lesssim e^{-\rho \widetilde{N}^{1/s}}$ for all $\underline{x} \in \mathbb{T}^2$.

We will need, as mentioned above, super-exponentially small error in how $v_N(\underline{x})$ approximates $v(\underline{x})$, otherwise the error would propagate and the transfer matrix substitutes will not be close to the original transfer matrices. Hence $\widetilde{N}$ should be chosen such that say $e^{-\rho \widetilde{N}^{1/s}} \le e^{-\rho N^2}$. So if $\widetilde{N} := N^{2s}$, then the width of the holomorphic (in each variable) extension $v_N(\underline{z})$ will be $\rho_{1,N} = \frac{\rho}{2} N^{2s(-1+\frac{1}{s})} = \frac{\rho}{2} N^{-2(s-1)} =: \frac{\rho}{2} N^{-\delta}$, where $\delta := 2(s-1) > 0$.

We conclude: for every integer $N \ge 1$, we have a function $v_N(\underline{x})$ on $\mathbb{T}^2$ such that

$$|v(\underline{x}) - v_N(\underline{x})| < e^{-\rho N^2} \tag{3.7}$$

and $v_N(\underline{x})$ has a 1-periodic separately holomorphic extension $v_N(\underline{z})$ to the strip $[|\Im z| < \rho_{1,N}] \times [|\Im z| < \rho_{1,N}]$, where $\rho_{1,N} = \frac{\rho}{2} N^{-\delta}$, for which

$$|v_N(\underline{z})| \le B. \tag{3.8}$$

The positive constants ρ, B, δ above depend only on v (not on the scale N). The constant δ depends on the Gevrey class of v: $\delta := 2(s-1)$ so it is fixed but presumably very large.

We now substitute these approximants $v_N(\underline{x})$ for $v(\underline{x})$ in the definition of the transfer matrix $M_N(\underline{x})$.

Let

$$A(\underline{x}, E) := \begin{bmatrix} \lambda v(\underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix}$$

be the cocycle that defines the transfer matrix $M_N(\underline{x})$.

Consider then

$$\widetilde{A}_N(\underline{x}, E) := \begin{bmatrix} \lambda v_N(\underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix},$$

which leads to the transfer matrix substitutes

$$\widetilde{M}_N(\underline{x}, E) := \prod_{j=N}^{1} \widetilde{A}_N(\mathrm{T}^j \underline{x}, E).$$

To show that the substitutes are close to the original matrices, we use Trotter's formula. This is a wasteful approach, and clearly in part responsible for our inability to apply these methods beyond Gevrey functions. There are other, much more

subtle reasons for why this approach is limited to this class of functions:

$$M_N(\underline{x}) - \widetilde{M}_N(\underline{x})$$

$$= \sum_{j=1}^{N} A(\mathrm{T}^N \underline{x}) \dots A(\mathrm{T}^{j+1} \underline{x}) \, [A(\mathrm{T}^j \underline{x}) - \widetilde{A}_N(\mathrm{T}^j \underline{x})] \, \widetilde{A}_N(\mathrm{T}^{j-1} \underline{x}) \dots$$

$$\widetilde{A}_N(\mathrm{T}\underline{x}) A(\mathrm{T}^j \underline{x}) - \widetilde{A}_N(\mathrm{T}^j \underline{x})$$

$$= \begin{bmatrix} \lambda v(\mathrm{T}^j \underline{x}) - \lambda v_N(\mathrm{T}^j \underline{x}) & 0 \\ 0 & 0 \end{bmatrix},$$

so

$$\|A(\mathrm{T}^j \underline{x}) - \widetilde{A}_N(\mathrm{T}^j \underline{x})\| \le |\lambda| \sup_{\underline{y} \in \mathbb{T}^2} |v(\underline{y}) - v_N(\underline{y})| < |\lambda| \, e^{-\rho N^2}.$$

Since $\sup_{\underline{x} \in \mathbb{T}^2} |v(\underline{x})| \le B$, the spectrum of the operator $H(\underline{x})$ is contained in the interval $[-2 - |\lambda| \, B, 2 + |\lambda| \, B]$. Hence it is enough to consider only the energies E such that $|E| \le 2 + |\lambda| \, B$. We then have

$$\|A(\mathrm{T}^j \underline{x})\| = \left\| \begin{bmatrix} \lambda v(\mathrm{T}^j \underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix} \right\| \le |\lambda| \, B + |E| + 2 \le 2|\lambda| B + 4 \le e^{S(\lambda)}$$

and

$$\|\widetilde{A}_N(\mathrm{T}^j \underline{x})\| \le \left\| \begin{bmatrix} \lambda v_N(\mathrm{T}^j \underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix} \right\| \le |\lambda| \, B + |E| + 2 \le e^{S(\lambda)}$$

Therefore,

$$\|A(\mathrm{T}^j \underline{x})\|, \|\widetilde{A}_N(\mathrm{T}^j \underline{x})\| \le e^{S(\lambda)} \tag{3.9}$$

where $S(\lambda) \approx \log |\lambda|$ is a scaling factor that depends only on the (assumed large) disorder λ and on v (the constants inherent in $\approx$ depend on the number $B = B(v)$ which also determines the range of spectral values E).

We then have

$$\|M_N(\underline{x}, E) - \widetilde{M}_N(\underline{x}, E)\| \le \sum_{j=1}^{N} e^{S(\lambda)} \dots e^{S(\lambda)} |\lambda| \, e^{-\rho N^2} e^{S(\lambda)} \dots e^{S(\lambda)}$$

$$\le e^{N S(\lambda) - \rho N^2}$$

$$\le e^{-\frac{\rho}{2} N^2},$$

provided $N \gtrsim S(\lambda)$.

Hence uniformly in $\underline{x} \in \mathbb{T}^2$ we get

$$\|M_N(\underline{x}, E) - \widetilde{M}_N(\underline{x}, E)\| \le e^{-\frac{\rho}{2} N^2}, \tag{3.10}$$

provided we choose

$$N \gtrsim S(\lambda). \tag{3.11}$$

which means roughly that λ has to be at most exponential in the scale N.

We are now going to turn our attention to the logarithmic averages of the transfer matrices.

Since $\det M_N(\underline{x}) = 1$ and $\det \widetilde{M}_N(\underline{x}) = 1$, we have that $\|M_N(\underline{x})\| \geq 1$ and $\|\widetilde{M}_N(\underline{x})\| \geq 1$. Thus, for all $N \gtrsim S(\lambda)$ and for every $\underline{x} \in \mathbb{T}^2$,

$$|\frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{N} \log \|\widetilde{M}_N(\underline{x})\|| \leq \frac{1}{N} \|M_N(\underline{x}) - \widetilde{M}_N(\underline{x})\| < e^{-\frac{\rho}{2}N^2}$$

.

Recall the notation

$$L_N(\underline{x}, E) = \frac{1}{N} \log \|M_N(\underline{x}, E)\| \, d\underline{x}, \tag{3.12}$$

and define its substitute

$$u_N(\underline{x}, E) := \frac{1}{N} \log \|\widetilde{M}_N(\underline{x})\|. \tag{3.13}$$

Therefore, uniformly in $\underline{x} \in \mathbb{T}^2$ and in the energy E,

$$|L_N(\underline{x}, E) - u_N(\underline{x}, E)| < e^{-\frac{\rho}{2}N^2},$$

and by averaging in $\underline{x}$,

$$|L_N(E) - \langle u_N(E) \rangle| < e^{-\frac{\rho}{2}N^2},$$

where

$$L_N(E) := \int_{\mathbb{T}^2} L_N(\underline{x}, E) \, d\underline{x},$$

and for any function $u(\underline{x})$,

$$\langle u \rangle := \int_{\mathbb{T}^2} u(\underline{x}) \, d\underline{x}.$$

The advantage of the substitutes $u_N(\underline{x})$ is that they extend to pluri-subharmonic functions in a neighborhood of the torus $\mathbb{T}^2$, as explained below.

For the skew-shift transformation $T = S_\omega$ we consider the strip

$$\underline{\mathcal{A}}_{\underline{\rho}_N} := [|\Im z| < \rho_{1,N}] \times [|\Im z| < \rho_{2,N}],$$

where $\rho_{1,N} = \frac{\rho}{4} N^{-\delta}$ and $\rho_{2,N} := \frac{\rho_{1,N}}{2N} = \frac{\rho}{4} N^{-\delta-1}$

We have to reduce the size of the strip in the second variable to account for the fact that the skew-shift expands in the imaginary direction. Our approximation method required a reduction in the size of the holomorphicity strip at each scale, and this additional reduction will be comparatively harmless.

If we extend the map S_ω from $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ to $\mathbb{C}^2$, by

$$S_\omega(z_1, z_2) = (z_1 + z_2, z_2 + \omega)$$

we get as in (1.5) that

$$S_\omega^n(z_1, z_2) = \left(z_1 + n z_2 + \frac{n(n-1)}{2}\omega, z_2 + n\omega \right),$$

Then if $(z_1, z_2) \in \underline{\mathcal{A}}_{\rho_N}$ and if we perform $n \leq N$ iterations,

$$\left| \Im(z_1 + n z_2 + \frac{n(n-1)}{2}\omega) \right| = |\Im(z_1 + n z_2)| = |y_1 + n y_2| < \frac{\rho}{2} N^{-\delta}. \quad (3.14)$$

The matrix function

$$\widetilde{A}_N(\underline{x}) = \begin{bmatrix} \lambda v_N(\underline{x}) - E & -1 \\ 1 & 0 \end{bmatrix}$$

extends to a 1-periodic, separately in each variable holomorphic matrix valued function,

$$\widetilde{A}_N(\underline{z}) := \begin{bmatrix} \lambda v_N(\underline{z}) - E & -1 \\ 1 & 0 \end{bmatrix}.$$

Using (3.8) and the definition of the scaling factor $S(\lambda)$, we have that on the strip $\underline{\mathcal{A}}_{\rho_N}$ the matrix valued function $\widetilde{A}_N(\underline{z})$ is uniformly in N bounded by $e^{S(\lambda)}$. Combining this with (3.14), the transfer matrix substitutes extend on the same strip to separately holomorphic matrix valued functions

$$\widetilde{M}_N(\underline{z}, E) := \prod_{j=N}^{1} \widetilde{A}_N(S_\omega^j \underline{z}, E)$$

such that, for all $\underline{z} \in \underline{\mathcal{A}}_{\rho_N}$ and for all energies E,

$$\|\widetilde{M}_N(\underline{z}, E)\| \leq e^{NS(\lambda)}.$$

Therefore,

$$u_N(\underline{z}) := \frac{1}{N} \log \|\widetilde{M}_N(\underline{z})\|$$

is a pluri-subharmonic function on the strip $\underline{\mathcal{A}}_{\rho_N}$, and for any $\underline{z}$ in this strip,

$$|u_N(\underline{z})| \leq S(\lambda).$$

The same argument applies to the multifrequency shift $\mathrm{T} = \mathrm{T}_{\underline{\omega}}$. The extension of this dynamics to the complex plane

$$\mathrm{T}_{\underline{\omega}}(z_1, z_2) = (z_1 + \omega_1, z_2 + \omega_2)$$

does not expand in the imaginary direction, so there is no need to decrease the width of the strip in the second variable as in the case of the skew-shift. However, for convenience of notations, we will choose the same strip $\underline{\mathcal{A}}_{\rho_N}$ for both transformations.

We can now summarize all of the above into the following.

Lemma 3.1. *For fixed parameters λ, E, for a fixed transformation $\mathrm{T} = \mathrm{S}_{\underline{\omega}}$ or $\mathrm{T} = \mathrm{T}_{\underline{\omega}}$ and for $\delta = 2(s-1)$, at every scale N we have a 1-periodic function*

$$u_N(\underline{x}) := \frac{1}{N} \log \|\widetilde{M}_N(\underline{x})\|$$

which extends to a pluri-subharmonic function $u_N(\underline{z})$ on the strip

$$\underline{\mathcal{A}}_{\rho_N} = [|\Im z| < \rho_{1,N}] \times [|\Im z| < \rho_{2,N}],$$

where $\rho_{1,N} \approx N^{-\delta}$, $\rho_{2,N} \approx N^{-\delta-1}$ so that

$$|u_N(\underline{z})| \leq S(\lambda) \quad \text{for all } \underline{z} \in \underline{\mathcal{A}}_{\rho_N}. \tag{3.15}$$

Note that the bound (3.15) is uniform in N.

Moreover, if $N \gtrsim S(\lambda)$, then the logarithmic averages of the transfer matrices $M_N(\underline{x})$ are well approximated by their substitutes $u_N(\underline{x})$,

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| - u_N(\underline{x}) \right| \lesssim e^{-N^2} \tag{3.16}$$

and

$$|L_N - \langle u_N \rangle| \lesssim e^{-N^2}. \tag{3.17}$$

All the inherent constants in the above (and future) estimates are either universal or depend only on v (and not on the scale N) so they can be ignored. The estimates above are independent of the variable $\underline{x}$, the parameters λ, E and the transformation T.

This s a crucial technical result in our paper, which will allow us to use subharmonic functions techniques as in [4], [6] for the functions u_N, and then transfer the relevant estimates to the rougher functions they substitute.

The logarithmic averages of the transfer matrix have an almost invariance (under the dynamics) property.

Lemma 3.2. *For all $\underline{x} \in \mathbb{T}^2$, for all parameters λ, E and for all transformations* T,

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{N} \log \|M_N(\mathrm{T}\,\underline{x})\| \right| \lesssim \frac{S(\lambda)}{N}. \qquad (3.18)$$

Proof.

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{N} \log \|M_N(\mathrm{T}\,\underline{x})\| \right|$$

$$= \left| \frac{1}{N} \log \frac{\|M_N(\underline{x})\|}{\|M_N(\mathrm{T}\,\underline{x})\|} \right|$$

$$= \left| \frac{1}{N} \log \frac{\|A(\mathrm{T}^N \underline{x}) \cdot \ldots \cdot A(\mathrm{T}^2 \underline{x}) \cdot A(\mathrm{T}\,\underline{x})\|}{\|A(\mathrm{T}^{N+1} \underline{x}) \cdot A(\mathrm{T}^N \underline{x}) \cdot \ldots \cdot A(\mathrm{T}^2 \underline{x})\|} \right|$$

$$\leq \frac{1}{N} \log[\,\|(A(\mathrm{T}^{N+1}\underline{x}))^{-1}\| \cdot \|A(\mathrm{T}\,\underline{x})\|\,]$$

$$\lesssim \frac{S(\lambda)}{N},$$

where the last bound is due to (3.9). The inequality (3.18) then follows. □

4. Averages of shifts of pluri-subharmonic functions

One of the main ingredients in the proof of the LDT (2.2) is an estimate on averages of shifts of pluri-subharmonic functions. These averages are shown to converge in a quantitative way to the mean of the function. The result holds for both the skew-shift and the multi-frequency shift.

For the skew-shift, the result was proven in [6] (see Lemma 2.6 there). We will reproduce here the scaled version of that result, the one that takes into account the size of the domain of subharmonicity and the sup norm of the function. The

reader can verify, by following the details of the proof in [6], that this is indeed
the correct scaled version. For the multi-frequency shift, the result is essentially
contained within the proof of Theorem 5.5 in [4], but for completeness, we will
include here the details of its proof.

Proposition 4.1. *Let $u(\underline{x})$ be a real valued function on $\mathbb{T}^2$, that extends to a pluri-subharmonic function $u(\underline{z})$ on a strip $\underline{A}_\rho = [|\Im z_1| < \rho_1] \times [|\Im z_2| < \rho_2]$. Let $\rho = \min\{\rho_1, \rho_2\}$. Let T be either the skew-shift or the multi-frequency shift on $\mathbb{T}^2$, where the underlying frequency satisfies the DC_κ described in* (1.8) *or* (1.9) *respectively. Assume that*

$$\sup_{\underline{z} \in \underline{A}_\rho} |u(\underline{z})| \le S.$$

Then for some explicit constants $\sigma_0, \tau_0 > 0$, and for $n \ge n(\kappa)$,

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 : \left|\frac{1}{n}\sum_{j=0}^{n-1} u(\mathrm{T}^j \underline{x}) - \langle u \rangle\right| > \frac{S}{\rho} n^{-\tau_0}\right] < e^{-n^{\sigma_0}}. \tag{4.1}$$

Here is how this estimate can be understood. Given the ergodicity of the transformation T for irrational (or rationally independent) frequencies, on the long run, the orbits $\mathrm{T}^j \underline{x}$ of most points $\underline{x}$ will tend to be fairly well distributed throughout the torus $\mathbb{T}^2$ (see the picture below).
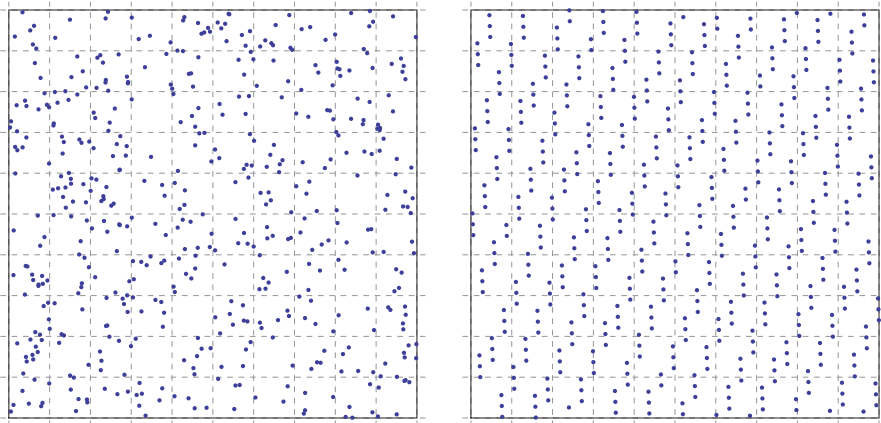
(a) Iterations of the skew-shift   (b) Iterations of the multifrequency shift

Figure 1

The average $\frac{1}{n}\sum_{j=0}^{n-1} u(\mathrm{T}^j \underline{x})$ will then resemble a Riemann sum of the function $u(\underline{x})$ and as such, it will approach the integral $\langle u \rangle$.

Moreover, a quantitative description of the irrationality (or rational independence) of the frequency in the form of a Diophantine condition like (1.8), (1.9), should lead to a quantitative description of the convergence of the average sum to the integral $\langle u \rangle$.

To prove this quantitative convergence result, we consider the Fourier expansion of the function $u(\underline{x})$ and apply it to the average sums. This leads to a convolution of $u(\underline{x})$ with a Fejér-type kernel. It is crucial to have estimates on the Fourier coefficients of the function u, and they are obtained via Riesz' representation theorem for subharmonic functions (see Corollary 4.1. in [4]). Since $u(z_1, z_2)$ is pluri-subharmonic, the scaled version of Corollary 4.1. in [4] implies

$$\sup_{x_2 \in \mathbb{T}} |\hat{u}(l_1, x_2)| \lesssim \frac{S}{\rho_1} \cdot \frac{1}{|l_1|} \quad \text{and} \quad \sup_{x_1 \in \mathbb{T}} |\hat{u}(x_1, l_2)| \lesssim \frac{S}{\rho_2} \cdot \frac{1}{|l_2|}. \tag{4.2}$$

The estimates (4.2) imply (small) upper bounds on the L^2 - norm of the part of the Fourier expansion for which at least one of the indices l_1 and l_2 is large. The difficult part is when both indices l_1 and l_2 are small, in which case we use the Diophantine condition on the frequency to estimate the resulting exponential sums.

In the case of the skew shift dynamics (1.5), the resulting exponential sums are quadratic, and they are estimated using Weyl's method (see [6] for the details of the proof). We will now present the details of the proof for the multi-frequency shift case $\mathrm{T}\underline{x} = \mathrm{T}_{\underline{\omega}}\underline{x} := \underline{x} + \underline{\omega}$.

Proof. Expand $u(\underline{x})$ into a Fourier series

$$u(\underline{x}) = \langle u \rangle + \sum_{\substack{\underline{l} \in \mathbb{Z}^2 \\ \underline{l} \neq (0,0)}} \hat{u}(\underline{l}) \cdot e^{2\pi i\, \underline{l} \cdot \underline{x}}.$$

Then the averages of shifts have the form

$$\frac{1}{n}\sum_{j=0}^{n-1} u(\mathrm{T}^j \underline{x}) = \frac{1}{n}\sum_{j=0}^{n-1} u(\underline{x} + j\underline{\omega})$$

$$= \langle u \rangle + \sum_{\substack{\underline{l} \in \mathbb{Z}^2 \\ \underline{l} \neq (0,0)}} \hat{u}(\underline{l}) \cdot e^{2\pi i\, \underline{l} \cdot \underline{x}} \cdot \left(\frac{1}{n}\sum_{j=0}^{n-1} e^{2\pi i\, j\, \underline{l} \cdot \underline{\omega}} \right)$$

$$= \langle u \rangle + \sum_{\substack{\underline{l} \in \mathbb{Z}^2 \\ \underline{l} \neq (0,0)}} \hat{u}(\underline{l}) \cdot e^{2\pi i\, \underline{l} \cdot \underline{x}} \cdot K_n(\underline{l} \cdot \underline{\omega}),$$

where we denoted by $K_n(t)$ the Fejér kernel

$$K_n(t) = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi i \, jt} = \frac{1}{n} \frac{1 - e^{2\pi i \, nt}}{1 - e^{2\pi i \, t}},$$

which clearly has the bound

$$|K_n(t)| \leq \min\left\{1, \frac{1}{n\|t\|}\right\}. \tag{4.3}$$

We then have

$$\|\frac{1}{n} \sum_{j=0}^{n-1} u(\underline{x} + j\underline{\omega}) - \langle u \rangle \|_{L^2(\mathbb{T}^2)}^2$$

$$= \sum_{\substack{\underline{l} \in \mathbb{Z}^2 \\ \underline{l} \neq (0,0)}} |\hat{u}(\underline{l})|^2 \cdot |K_n(\underline{l} \cdot \underline{\omega})|^2$$

$$= \sum_{1 \leq |\underline{l}| < K} |\hat{u}(\underline{l})|^2 \cdot |K_n(\underline{l} \cdot \underline{\omega})|^2 + \sum_{|\underline{l}| \geq K} |\hat{u}(\underline{l})|^2 \cdot |K_n(\underline{l} \cdot \underline{\omega})|^2.$$

We will estimate the second sum above using the bounds (4.2) on the Fourier coefficients of $u(\underline{x})$ and the first sum using the DC (1.9) on the frequency $\underline{\omega}$. The splitting point K will be chosen to optimize the sum of these estimates.

Clearly (4.2) implies

$$\sum_{l_2 \in \mathbb{Z}} |\hat{u}(l_1, l_2)|^2 = \|\hat{u}(l_1, x_2)\|_{L^2_{x_2}(\mathbb{T})}^2 \lesssim \left(\frac{S}{\rho_1} \frac{1}{|l_1|}\right)^2 \leq \left(\frac{S}{\rho}\right)^2 \frac{1}{|l_1|^2}$$

and

$$\sum_{l_1 \in \mathbb{Z}} |\hat{u}(l_1, l_2)|^2 = \|\hat{u}(x_1, l_2)\|_{L^2_{x_1}(\mathbb{T})}^2 \lesssim \left(\frac{S}{\rho_2} \frac{1}{|l_2|}\right)^2 \leq \left(\frac{S}{\rho}\right)^2 \frac{1}{|l_2|^2}.$$

Then,

$$\sum_{|\underline{l}| \geq K} |\hat{u}(\underline{l})|^2 \cdot |K_n(\underline{l} \cdot \underline{\omega})|^2 \leq \sum_{|\underline{l}| \geq K} |\hat{u}(\underline{l})|^2$$

$$\leq \sum_{\underline{l} \,:\, |l_1| \geq K/2} |\hat{u}(\underline{l})|^2 + \sum_{\underline{l} \,:\, |l_2| \geq K/2} |\hat{u}(\underline{l})|^2$$

$$\lesssim \left(\frac{S}{\rho}\right)^2 \frac{1}{K}.$$

Estimate (4.2) clearly implies

$$|\hat{u}(\underline{l})| \lesssim \frac{S}{\rho} \frac{1}{|\underline{l}|}.$$

Then using the DC (1.9) on $\underline{\omega}$ and (4.3), we obtain

$$\sum_{1 \leq |\underline{l}| < K} |\hat{u}(\underline{l})|^2 \cdot |K_n(\underline{l} \cdot \underline{\omega})|^2 \leq \left(\frac{S}{\rho}\right)^2 \sum_{1 \leq |\underline{l}| < K} \frac{1}{|\underline{l}|^2} \cdot \frac{1}{n^2 \|\underline{l} \cdot \underline{\omega}\|^2}$$

$$\leq \left(\frac{S}{\rho}\right)^2 \sum_{1 \leq |\underline{l}| < K} \frac{1}{|\underline{l}|^2} \cdot \frac{|\underline{l}|^{2A}}{n^2 \kappa^2}$$

$$\lesssim \left(\frac{S}{\rho}\right)^2 \frac{K^{2A}}{n^2 \kappa^2}.$$

We conclude

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} u(\underline{x} + j\underline{\omega}) - \langle u \rangle \right\|_{L^2(\mathbb{T}^2)} \leq \frac{S}{\rho} \left(\frac{1}{K^{1/2}} + \frac{K^A}{n\kappa} \right) \leq \frac{S}{\rho} n^{-a}$$

for some positive constant a that depends on A and for n large enough depending on A and κ.

Using Chebyshev's inequality, the above estimate implies

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{n} \sum_{j=0}^{n-1} u(\underline{x} + j\underline{\omega}) - \langle u \rangle \right| > \frac{S}{\rho} n^{-a/3} \right] < n^{-4a/3}. \qquad (4.4)$$

This is not exactly what we wanted, since the size of the "bad" set above decays only polynomially fast in n, instead of exponentially fast.

To boost this estimate, we will use Lemma 4.12 in J. Bourgain's monograph [4]. This result shows that a weaker a-priori estimate on a subharmonic function implies an upper bound on its BMO norm, which in turn leads, via John-Nirenberg inequality, to a stronger estimate on the function. We reproduce here a "rescaled" version of the estimate in [4], one that takes into account the width ρ of subharmonicity. The reader may verify that this is indeed the correct rescaled version of the statement.

Lemma 4.1. *Assume that* $u = u(\underline{x})\colon \mathbb{T}^2 \to \mathbb{R}$ *has a pluri-subharmonic extension* $u(\underline{z})$ *on* $\underline{\mathcal{A}}_\rho = [|\Im z_1| < \rho_1] \times [|\Im z_2| < \rho_2]$ *such that* $\sup\limits_{\underline{z} \in \underline{\mathcal{A}}_\rho} |u(\underline{z})| \le B$. *Let*

$$\rho = \min\{\rho_1, \rho_2\}.$$

If

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u(\underline{x}) - \langle u \rangle| > \epsilon_0] < \epsilon_1, \tag{4.5}$$

then for an absolute constant $c > 0$,

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u(\underline{x}) - \langle u \rangle| > \epsilon_0{}^{1/4}] < e^{-c\left(\epsilon_0{}^{1/4} + \sqrt{\frac{B}{\rho}}\, \frac{\epsilon_1{}^{1/4}}{\epsilon_0{}^{1/2}}\right)^{-1}}. \tag{4.6}$$

We will apply this result to the average

$$u^\sharp(\underline{x}) := \frac{\rho}{S} \frac{1}{n} \sum_{j=0}^{n-1} u(\underline{x} + j\underline{\omega}).$$

Clearly $u^\sharp(\underline{x})$ is pluri-subharmonic on the same strip $\underline{\mathcal{A}}_\rho$ as $u(\underline{x})$, its upper bound on this strip is $B = \rho$ and its mean is $\langle u^\sharp \rangle = \frac{\rho}{S} \langle u \rangle$.

Then (4.4) implies

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u^\sharp(\underline{x}) - \langle u^\sharp \rangle| > \epsilon_0] < \epsilon_1, \tag{4.7}$$

where $\epsilon_0 := n^{-a/3}$ and $\epsilon_1 := n^{-4a/3}$ so $\epsilon_1 \ll \epsilon_0$.

Applying Lemma 4.1 and performing the obvious calculations, from inequality (4.6), we get

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u^\sharp(\underline{x}) - \langle u^\sharp \rangle| > n^{-a/12}] < e^{-c\, n^{a/12}},$$

which then implies (4.1) for the multi-frequency shift $T_{\underline{\omega}}$. □

5. Łojasiewicz inequality for multivariable smooth functions

To prove the large deviation estimate (2.2) for a large enough initial scale N_0, we will need a quantitative description of the transversality condition (1.12). More precisely, we will show that if a smooth function $v(\underline{x})$ is not flat at any point as defined in (1.12), then the set $[\underline{x}\colon v(\underline{x}) \approx E]$ of points where $v(\underline{x})$ is almost constant has small measure (and bounded complexity).

Such an estimate is called a Łojasiewicz type inequality and it is already available for non-constant analytic functions. For such functions it can be derived using complex analysis methods from [20], namely lower bounds for the modulus of a holomorphic function on a disk (see Lemma 11.4 in [12]).

For non-analytic functions, the proof is more difficult. Using Sard-type arguments, we have obtained a similar result for one-variable functions (see Lemma 5.3 in [16]). For multivariable smooth functions, the argument is more technical and it involves a quantitative form of the implicit function theorem, also used in [7] and [14].

We begin with a simple compactness argument that shows that in the TC (1.12) we can work with finitely many partial derivatives.

Lemma 5.1. *Assume $v(\underline{x})$ is a smooth, 1-periodic function on $\mathbb{R}^2$. Then $v(\underline{x})$ satisfies the transversality condition* (1.12) *if and only if there exist $\underline{m} \in \mathbb{N}^2$, $|\underline{m}| \neq 0$, and $c > 0$ such that*

$$\max_{\substack{\underline{\alpha} \leq \underline{m} \\ |\alpha| \neq 0}} |\partial^{\underline{\alpha}} v(\underline{x})| \geq c \quad \text{for all } \underline{x} \in \mathbb{T}^2. \tag{5.1}$$

The constants m and c in (5.1) *depend only on v.*

Proof. Clearly (5.1) $\Rightarrow$ (1.12). We prove the converse. The TC (1.12) implies that $\underline{x} \in \mathbb{T}^2$ there exists $\underline{m}_x \in \mathbb{N}^2$, $|\underline{m}_x| \neq 0$ such that

$$|\partial^{\underline{m}_x} v(\underline{x})| > c_{\underline{x}} > 0.$$

Then there are radii $r_{\underline{x}} > 0$ such that if $\underline{y}$ is in the disk $D(\underline{x}, r_{\underline{x}})$, then

$$|\partial^{\underline{m}_x} v(\underline{y})| \geq c_{\underline{x}} > 0.$$

The family $\{D(\underline{x}, r_{\underline{x}}) : \underline{x} \in \mathbb{T}^2\}$ covers $\mathbb{T}^2$. Consider a finite subcover

$$\{D(\underline{x}_1, r_{\underline{x}_1}), \dots, D(\underline{x}_k, r_{\underline{x}_k})\}.$$

Let $\underline{m} \in \mathbb{N}^2$ such that $\underline{m} \geq \underline{m}_{x_j}$ for all $1 \leq j \leq k$ and $c := \min_{1 \leq j \leq k} c_{\underline{x}_j}$. Then (5.1) follows. $\qquad\square$

The following is a more precise form of the implicit function theorem (which was also used in [14]).

Lemma 5.2. *Let $f(\underline{x})$ be a C^1 function on a rectangle $\underline{\mathcal{R}} = I \times J \subset [0,1]^2$, let $J = [c, d]$ and $A := \max_{\underline{x} \in \underline{\mathcal{R}}} |\partial_{x_1} f(\underline{x})|$. Assume that*

$$\min_{\underline{x} \in \underline{\mathcal{R}}} |\partial_{x_2} f(\underline{x})| =: \epsilon_0 > 0. \tag{5.2}$$

If $f(a_1, a_2) = 0$ for some point $(a_1, a_2) \in \underline{\mathcal{R}}$, then there is an interval $I_0 = (a_1 - \kappa, a_2 + \kappa) \subset I$ and a C^1 function $\phi_0(x_1)$ on I_0 such that

(i) $f(x_1, \phi_0(x_1)) = 0$ *for all $x_1 \in I_0$,*

(ii) $|\partial_{x_1} \phi_0(x_1)| \leq A \epsilon_0^{-1}$, *and*

(iii) $x_1 \in I_0$ *and* $f(x_1, x_2) = 0 \implies x_2 = \phi_0(x_1)$.

Moreover, the size κ of the domain of ϕ_0 can be taken as large as $\kappa \sim \epsilon_0 A^{-1} \cdot \min\{a_2 - c, d - a_2\}$.

Proof. From (5.2), since $\partial_{x_2} f(\underline{x})$ is either positive on $\underline{\mathcal{R}}$ or negative on $\underline{\mathcal{R}}$ (in which case replace f by $-f$), we may clearly assume that in fact

$$\min_{\underline{x} \in \underline{\mathcal{R}}} \partial_{x_2} f(\underline{x}) =: \epsilon_0 > 0. \tag{5.3}$$

Moreover, note that for any fixed $x_1 \in I$, since $\partial_{x_2} f(x_1, x_2) \neq 0$, the equation $f(x_1, x_2) = 0$ has a unique solution x_2.

Let $x_1 \in I_0$. Then

$$|f(x_1, a_2)| = |f(x_1, a_2) - f(a_1, a_2)| = |\partial_{x_1} f(\xi, a_2) + \cdot |x_1 - a_1| \leq A\kappa.$$

We have two possibilities.

■ $0 \leq f(x_1, a_2) \leq A\kappa$. Then, if $a_2 - t \in J$, we have

$$f(x_1, a_2 - t) - f(x_1, a_2) = \partial_{x_2} f(x_1, \xi) \cdot (-t),$$

and

$$f(x_1, a_2 - t) = f(x_1, a_2) - t \cdot \partial_{x_2} f(x_1, \xi) \leq A\kappa - t\epsilon_0 = 0,$$

provided $t = A\epsilon_0^{-1}\kappa$

For this choice of t, $a_2 - t$ is indeed in J, because of the size κ_0 of the interval I_0: $t = A\epsilon_0^{-1}\kappa \leq A\epsilon_0^{-1}\epsilon_0 A^{-1}(a_2 - c) = a_2 - c$, so $a_2 - t \geq c$.

Therefore,

$$f(x_1, a_2 - t) \leq 0 \leq f(x_1, a_2)$$

so there is a unique $x_2 =: \phi_0(x_1) \in [a_2 - t, a_2]$ such that $f(x_1, \phi_0(x_1)) = 0$.

■ $-A\kappa \leq f(x_1, a_2) \leq 0$. Then, if $a_2 + t \in J$, we have

$$f(x_1, a_2 + t) - f(x_1, a_2) = \partial_{x_2} f(x_1, \xi) \cdot t,$$

and

$$f(x_1, a_2 + t) = f(x_1, a_2) + t \cdot \partial_{x_2} f(x_1, \xi) \geq -A\kappa + t\epsilon_0 = 0,$$

provided $t = A\epsilon_0^{-1}\kappa$

As before, for this choice of t, $a_2 + t$ is in J, because of the size κ of the interval I_0: $t = A\epsilon_0^{-1}\kappa \leq A\epsilon_0^{-1}\epsilon_0 A^{-1}(d - a_2) = d - a_2$, so $a_2 + t \leq d$.

Therefore,

$$f(x_1, a_2) \leq 0 \leq f(x_1, a_2 + t),$$

so there is a unique $x_2 =: \phi_0(x_1) \in [a_2, a_2 + t]$ such that $f(x_1, \phi_0(x_1)) = 0$.

We proved (i) and (iii). The fact that $\phi_0(x_1)$ is C^1 follows from the standard implicit function theorem, while the estimate (ii) follows immediately from (i) using the chain's rule. □

The following is a quantitative and global version of the previous lemma (see also Lemma 8.3 in [14]). It says that under the same conditions as above, the points $(x_1, x_2) \in \underline{\mathcal{R}}$ for which $|f(x_1, x_2)| \leq \epsilon$ are either in a narrow strip at the top or at the bottom of the rectangle $\underline{\mathcal{R}}$, or near the graphs of some functions $\phi_j(x_1)$, in other words $x_2 \approx \phi_j(x_1)$.

Lemma 5.3. *Let $f(\underline{x})$ be a C^1 function on a rectangle $\underline{\mathcal{R}} = I \times J \subset [0, 1]^2$, where $|I| \sim \kappa_0$. Let $J = [c, d]$ and $A := \max\limits_{\underline{x} \in \underline{\mathcal{R}}} |\partial_{x_1} f(\underline{x})|$. Assume that*

$$\min_{\underline{x} \in \underline{\mathcal{R}}} |\partial_{x_2} f(\underline{x})| =: \epsilon_0 > 0 \tag{5.4}$$

Let $\epsilon_1 > 0$ be small enough, i.e. $\epsilon_1 < \frac{\epsilon_0 \kappa_0}{4}$ and $\kappa_1 \sim \epsilon_1 A^{-1}$.

Then there are about $\kappa_0 \kappa_1^{-1}$ sub-intervals I_j of length κ_1 covering I, and on each interval I_j there is a C^1 function $\phi_j(x_1)$ such that

(i) $f(x_1, \phi_j(x_1)) = 0$, *for all $x_1 \in I_j$,*

(ii) $|\partial_{x_1} \phi_j(x_1)| \leq A\epsilon_0^{-1}$, *and*

(iii) $[(x_1, x_2) \in \underline{\mathcal{R}} : |f(x_1, x_2)| < \epsilon_1] \subset \underline{\mathcal{R}}^t \cup \underline{\mathcal{R}}^b \cup (\cup_j \underline{S}_j)$,

where

$$\underline{\mathcal{R}}^t := I \times [d - 2\epsilon_1\epsilon_0^{-1}, d]$$

$$\underline{\mathcal{R}}^b := I \times [c, c + 2\epsilon_1\epsilon_0^{-1}]$$

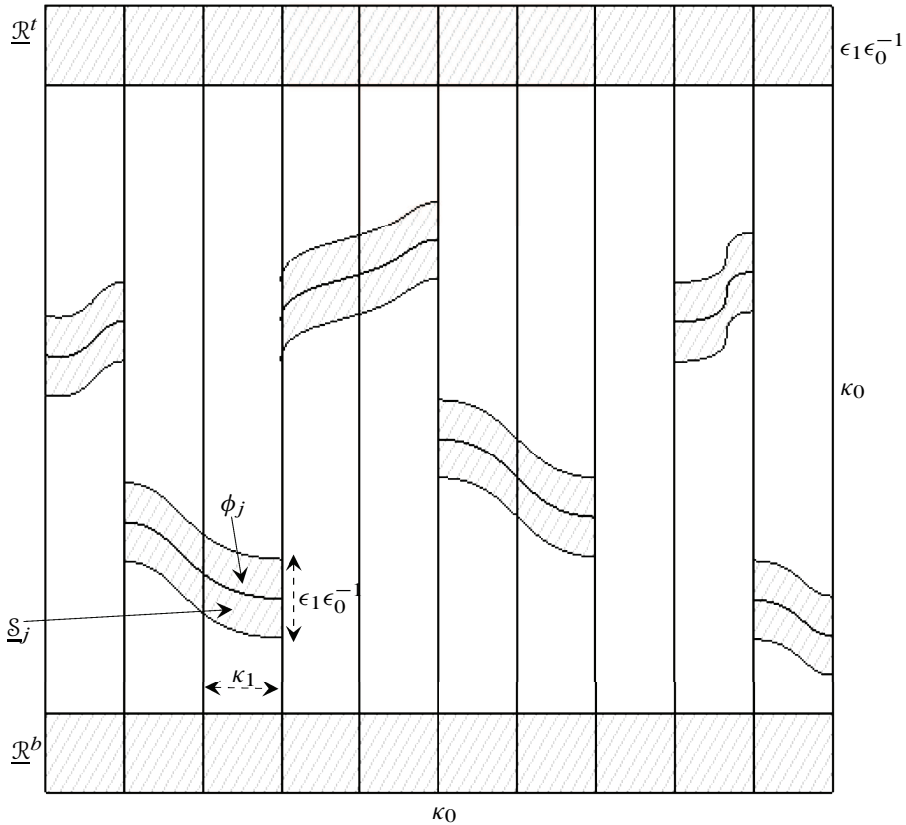$$\underline{\mathcal{S}}_j := [(x_1, x_2) \colon x_1 \in I_j, |x_2 - \phi_j(x_1)| < \epsilon_1\epsilon_0^{-1}]$$

Figure 2

Proof. Divide the interval I, whose length is $\sim \kappa_0$ into $\sim \kappa_0\kappa_1^{-1}$ sub-intervals I_j of length κ_1 each.

If $|f(x_1, x_2)| \geq \epsilon_1$ for all $(x_1, x_2) \in I_j \times [c + 2\epsilon_1\epsilon_0^{-1}, d - 2\epsilon_1\epsilon_0^{-1}]$, then we are done with the interval I_j.

Otherwise, assume $|f(a_1, a_2)| < \epsilon_1$ for some $a_1 \in I_j$ and $a_2 \in [c + 2\epsilon_1\epsilon_0^{-1}, d - 2\epsilon_1\epsilon_0^{-1}]$.

We may assume $0 \leq f(a_1, a_2) \leq \epsilon_1$, the other case being treated similarly. Then if $a_2 - t \in J$, we have

$$f(a_1, a_2 - t) - f(a_1, a_2) = \partial_{x_2} f(a_1, \xi) \cdot (-t), \quad \text{for some } \xi \in (a_2 - t, a_2),$$

and

$$f(a_1, a_2 - t) = f(a_1, a_2) - t \cdot \partial_{x_2} f(a_1, \xi) \leq \epsilon_1 - \epsilon_0 t = 0,$$

provided $t = \epsilon_1 \epsilon_0^{-1}$. Since $a_2 \geq c + 2\epsilon_1 \epsilon_0^{-1}$, for this t we have $a_2 - t \in J$.

We then have $f(a_1, a_2 - t) \leq 0 \leq f(a_1, a_2)$, so $f(a_1, a_2^*) = 0$ for some $a_2^* \in [a_2 - t, a_2]$.

We can use Lemma 5.2 around the point (a_1, a_2^*). The interval we get has length at least $\epsilon_0 A^{-1} \cdot \min\{a_2 - c, d - a_2\} > \epsilon_0 A^{-1} \cdot 2\epsilon_1 \epsilon_0^{-1} = 2\epsilon_1 A^{-1} > 2\kappa_1$, so it contains I_j, whose length is $\sim \kappa_1$. We have a C^1 function ϕ_j on I_j such that $|\partial_{x_1} \phi_j| \leq A\epsilon_0^{-1}$ and

$$x_1 \in I_j \text{ and } f(x_1, x_2) = 0 \iff x_2 = \phi_j(x_1).$$

Now let $(x_1, x_2) \in \underline{\mathcal{R}}$ such that $|f(x_1, x_2)| < \epsilon_1$. Then either $(x_1, x_2) \in \underline{\mathcal{R}}^t \cup \underline{\mathcal{R}}^b$ or $(x_1, x_2) \in I_j \times [c + 2\epsilon_1 \epsilon_0^{-1}, d - 2\epsilon_1 \epsilon_0^{-1}]$ for some j, in which case

$$\begin{aligned}
\epsilon_1 &> |f(x_1, x_2)| \\
&= |f(x_1, x_2) - f(x_1, \phi_j(x_1))| \\
&= |\partial_{x_2} f(x_1, \xi)| \cdot |x_2 - \phi_j(x_1)| \\
&\geq \epsilon_0 \cdot |x_2 - \phi_j(x_1)|,
\end{aligned}$$

from which we conclude that $|x_2 - \phi_j(x_1)| < \epsilon_1 \epsilon_0^{-1}$.

$\square$

We have shown that the points $\underline{x} = (x_1, x_2) \in \underline{\mathcal{R}}$ for which $|f(\underline{x})| < \epsilon_1$ are within $\sim \epsilon_1$ from the graphs of some functions $\phi_j(x_1)$ that have bounded slopes and are defined on small intervals I_j. This shows that the 'bad' set

$$[\underline{x} \in \underline{\mathcal{R}} \colon |f(\underline{x})| < \epsilon_1]$$

can be covered by small rectangles instead of ϵ_1-neighborhoods of curves, and we have control on the size of these rectangles and on their number. In turn, the 'good' set $[\underline{x} \in \underline{\mathcal{R}} \colon |f(\underline{x})| \geq \epsilon_1]$ can be covered by a comparable number of rectangles, which can be further chopped down into squares, to preserve the symmetry between the two variables. This is the content of the following lemma.

Lemma 5.4. *Given a C^2 function $f(\underline{x})$ on a square $\underline{R}_0 = I_0 \times J_0 \subset [0, 1]^2$, where $|I_0|, |J_0| \sim \kappa_0$. Denote $A := \max\limits_{|\underline{\alpha}| \leq 2} \max\limits_{\underline{x} \in \underline{R}} |\partial^{\underline{\alpha}} f(\underline{x})|$. Assume that*

$$\min_{\underline{x} \in \underline{R}_0} |\partial_{x_2} f(\underline{x})| =: \epsilon_0 > 0 \quad or \quad \min_{\underline{x} \in \underline{R}_0} |\partial_{x_1} f(\underline{x})| =: \epsilon_0 > 0. \tag{5.5}$$

Let $\epsilon_1 > 0$ be small enough, i.e. $\epsilon_1 < \frac{\epsilon_0 \kappa_0}{4}$ and $\kappa_1 \sim \epsilon_1 A^{-1}$. Then there is a set $\underline{B}_1 \subset \underline{R}_0$, with

$$\mathrm{mes}[\underline{B}_1] \lesssim \kappa_0 \, \epsilon_1 \, \epsilon_0^{-1}, \tag{5.6}$$

such that $\underline{R}_0 \setminus \underline{B}_1$ is a union of about $(\kappa_0 \, \kappa_1^{-1})^2$ squares, where each such square has the form $\underline{R}_1 = I_1 \times J_1$, with $|I_1|, |J_1| \sim \kappa_1$.

For each of these squares,

$$\min_{\underline{x} \in \underline{R}_1} |f(\underline{x})| \geq \epsilon_1. \tag{5.7}$$

Proof. We will use Lemma 5.3. I_0 is covered by about $\kappa_0 \kappa_1^{-1}$ subintervals I_j of length κ_1. Consider one such subinterval. There is a C^1 function $\phi_j(x_1)$ on I_j such that $|\partial_{x_1} \phi_j(x_1)| \leq A \, \epsilon_0^{-1}$ and

$$[(x_1, x_2) \in I_j \times J_0 : |f(x_1, x_2)| < \epsilon_1] \subset \underline{R}_j^t \cup \underline{R}_j^b \cup \underline{S}_j,$$

where if $J_0 = [c_0, d_0]$, then

$$\underline{R}_j^t := I_j \times [d_0 - 2\epsilon_1 \epsilon_0^{-1}, d_0],$$

$$\underline{R}_j^b := I_j \times [c_0, c_0 + 2\epsilon_1 \epsilon_0^{-1}],$$

$$\underline{S}_j := [(x_1, x_2) : x_1 \in I_j, |x_2 - \phi_j(x_1)| < \epsilon_1 \epsilon_0^{-1}].$$

Then,

$$\underline{S}_j \subset I_j \times [\min_{x_1 \in I_j} \phi_j(x_1) - \epsilon_1 \epsilon_0^{-1}, \ \max_{x_1 \in I_j} \phi_j(x_1) + \epsilon_1 \epsilon_0^{-1}] =: I_j \times K_j^m =: \underline{R}_j^m.$$

For any $x_1, x_1' \in I_j$,

$$|\phi_j(x_1) - \phi_j(x_1')| \lesssim A\epsilon_0^{-1} \cdot |x_1 - x_1'| \leq A\epsilon_0^{-1}\kappa_1 \sim \epsilon_1\epsilon_0^{-1},$$

which shows that

$$|K_j^m| \lesssim \epsilon_1\epsilon_0^{-1}.$$

We have shown that $[(x_1, x_2) \in I_j \times J_0 : |f(x_1, x_2)| < \epsilon_1]$ is covered by three rectangles: $\underline{R}_j^t, \underline{R}_j^b, \underline{R}_j^m$, each of the form $I_j \times K_j$ where $|I_j| \sim \kappa_1, |K_j| \sim \epsilon_1\epsilon_0^{-1}$.

Summing over $j \lesssim \kappa_0 \kappa_1^{-1}$, we get that the set $[\underline{x} \in \underline{\mathcal{R}}_0 : |f(\underline{x})| < \epsilon_1]$ is contained in the union $\underline{\mathcal{B}}_1$ of about $\kappa_0 \kappa_1^{-1}$ rectangles of size $\kappa_1 \times \epsilon_1 \epsilon_0^{-1}$. Then

$$\mathrm{mes}[\underline{\mathcal{B}}_1] \lesssim \kappa_0 \kappa_1^{-1} \cdot \kappa_1 \cdot \epsilon_1 \epsilon_0^{-1} = \kappa_0 \epsilon_1 \epsilon_0^{-1},$$

which proves (5.6).

The complement of this set, $\underline{\mathcal{R}}_0 \setminus \underline{\mathcal{B}}_1$, consists of about the same number $\kappa_0 \kappa_1^{-1}$ of rectangles - this was the reason for switching from ϵ_1-neighborhoods of curves to rectangles. Each of these rectangles has the form $I_j \times L_j$, where $|I_j| \sim \kappa_1$ and $|L_j| \sim \kappa_0 - O(\epsilon_1 \epsilon_0^{-1}) \sim \kappa_0 \gg \kappa_1$. Divide each of these vertical rectangles into about $\kappa_0 \kappa_1^{-1}$ squares of size $\kappa_1 \times \kappa_1$ each.

We conclude that $\underline{\mathcal{R}}_0 \setminus \underline{\mathcal{B}}_1$ is covered by about $(\kappa_0 \kappa_1^{-1})^2$ squares of the form $\underline{\mathcal{R}}_1 = I_1 \times J_1$, where the size of each square is $|I_1|, |J_1| \sim \kappa_1$. $\qquad \square$

We now have all the ingredients for proving the following Łojasiewicz type inequality.

Theorem 5.1. *Assume that $v(\underline{x})$ is a smooth function on $[0, 1]^2$ satisfying the transversality condition (1.12). Then, for every $\epsilon > 0$,*

$$\sup_{E \in \mathbb{R}} \mathrm{mes}[\underline{x} \in [0, 1]^2 : |v(\underline{x}) - E| < \epsilon] < C \cdot \epsilon^b, \qquad (5.8)$$

where $C, b > 0$ depend only on v.

Proof. Using Lemma 5.1, there exist $\underline{m} = (m_1, m_2) \in \mathbb{N}^2$, $|\underline{m}| \neq 0$, and $c > 0$ such that

$$\max_{\substack{\underline{\alpha} \leq \underline{m} \\ |\alpha| \neq 0}} |\partial^{\underline{\alpha}} v(\underline{x})| \geq c \quad \text{for all } \underline{x} \in \mathbb{T}^2.$$

Let

$$A := \max_{\underline{\alpha} \leq (m_1 + 1, m_2 + 1)} \max_{\underline{x} \in [0, 1]^2} |\partial^{\underline{\alpha}} v(\underline{x})|.$$

We may of course assume that $|E| \leq 2A$, otherwise there is nothing to prove.

All the constants in the estimates that follow will depend only on $|\underline{m}|, c, A$ (so in particular only on v).

Partition $[0, 1]^2$ into about $(\frac{2A}{c})^2$ squares of the form $\underline{\mathcal{R}} = I \times J$ of size $|I|, |J| \sim \frac{c}{2A}$.

Let $\underline{\mathcal{R}}$ be such a square. Then either $|v(\underline{x})| \geq \epsilon$ for all $\underline{x} \in \underline{\mathcal{R}}$, in which case we are done with this square, or for some $\underline{a} = (a_1, a_2) \in \underline{\mathcal{R}}$ we have $|v(\underline{a})| < \epsilon$. But then for one of the partial derivatives $\underline{\alpha} \leq \underline{m}$, $|\alpha| \neq 0$, we have $|\partial^{\underline{\alpha}} v(\underline{a})| \geq c$.

Assume for simplicity that $|\partial^{\underline{m}} v(\underline{a})| \geq c$, which is the worst case scenario.
If $\underline{x} \in \underline{R}$, then $\|\underline{x} - \underline{a}\|_\infty := \max\{|x_1 - a_1|, |x_2 - a_2|\} \leq \frac{c}{2A}$.
Then

$$|\partial^{\underline{m}} v(\underline{x}) - \partial^{\underline{m}} v(\underline{a})| \lesssim \max_{\underline{y} \in \underline{R}} |\nabla \partial^{\underline{m}} v(\underline{y})| \cdot \|\underline{x} - \underline{a}\|_\infty \leq A \cdot \frac{c}{2A} = \frac{c}{2}.$$

It follows that

$$\min_{\underline{x} \in \underline{R}} |\partial^{\underline{m}} v(\underline{x})| \gtrsim \frac{c}{2}.$$

We will use Lemma 5.4 $|\underline{m}| =: m$ times.

Step 1. Let

$$f_1(\underline{x}) := \partial^{(m_1, m_2 - 1)} v(\underline{x}).$$

Then

$$\min_{\underline{x} \in \underline{R}} |\partial_{x_2} f_1(\underline{x})| = \min_{\underline{x} \in \underline{R}} |\partial^{\underline{m}} v(\underline{x})|. \gtrsim c$$

We apply Lemma 5.4 to the function f_1 with the following data:

$$\underline{R}_0 = \underline{R}, \quad \kappa_0 = \frac{c}{2A}, \quad \epsilon_0 \sim c, \quad \epsilon_1 < \frac{\epsilon_0 \kappa_0}{4}, \quad \kappa_1 \sim \epsilon_1 A^{-1},$$

where ϵ_1 will be chosen later.

We get a set $\underline{B}_1^\flat := \underline{B}_1$, $\mathrm{mes}[\underline{B}_1^\flat] \lesssim \kappa_0 \epsilon_1 \epsilon_0^{-1} < \kappa_0^2 A \cdot \epsilon_1 \epsilon_0^{-2}$ such that $\underline{R}_0 \setminus \underline{B}_1^\flat$ is
a union of about $(\kappa_0 \kappa_1^{-1})^2$ squares of the form $\underline{R}_1 = I_1 \times J_1$, of size $|I_1|, |J_1| \sim \kappa_1$.
For each of these squares,

$$\min_{\underline{x} \in \underline{R}_1} |f_1(\underline{x})| \geq \epsilon_1,$$

which means

$$\min_{\underline{x} \in \underline{R}_1} |\partial^{(m_1, m_2 - 1)} v(\underline{x})| \geq \epsilon_1.$$

Step 2. Pick any of the squares $\underline{R}_1 = I_1 \times J_1$ from the previous step and consider
say

$$f_2(\underline{x}) := \partial^{(m_1 - 1, m_2 - 1)} v(\underline{x}).$$

Then

$$\min_{\underline{x} \in \underline{R}_1} |\partial_{x_1} f_2(\underline{x})| = \min_{\underline{x} \in \underline{R}_1} |\partial^{(m_1, m_2 - 1)} v(\underline{x})| \geq \epsilon_1.$$

Apply Lemma 5.4 to the function f_2 with the following data:

$$\underline{\mathcal{R}}_1, \kappa_1, \epsilon_1 \text{ from Step 1}, \quad \epsilon_2 < \tfrac{\epsilon_1 \kappa_1}{4}, \quad \kappa_2 \sim \epsilon_2 A^{-1},$$

where ϵ_2 will be chosen later.

We get a set $\underline{\mathcal{B}}_2$, $\mathrm{mes}[\underline{\mathcal{B}}_2] \lesssim \kappa_1 \epsilon_2 \epsilon_1^{-1}$ such that $\underline{\mathcal{R}}_1 \setminus \underline{\mathcal{B}}_2$ is a union of about $(\kappa_1 \kappa_2^{-1})^2$ squares of the form $\underline{\mathcal{R}}_2 = I_2 \times J_2$, of size $|I_2|, |J_2| \sim \kappa_2$. For each of these squares,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_2} |f_2(\underline{x})| \geq \epsilon_2,$$

which means

$$\min_{\underline{x} \in \underline{\mathcal{R}}_2} |\, \partial^{(m_1-1, m_2-1)} v(\underline{x})| \geq \epsilon_2.$$

If we do this for each of the $\sim (\kappa_0 \kappa_1^{-1})^2$ squares resulting from Step 1, and if we put together all the 'bad' sets $\underline{\mathcal{B}}_2$ corresponding to each of these squares, we conclude the following.

There is a set $\underline{\mathcal{B}}_2^{\flat} \subset \underline{\mathcal{R}}$ such that

$$\mathrm{mes}[\underline{\mathcal{B}}_2^{\flat}] \lesssim \kappa_1 \epsilon_2 \epsilon_1^{-1} \cdot (\kappa_0 \kappa_1^{-1})^2 = \kappa_0 \epsilon_2 \epsilon_1^{-1} \kappa_1^{-1} \sim \kappa_0^2 A \cdot \epsilon_2 \epsilon_1^{-2}.$$

Hence the total measure of the 'bad' set in Step 2 is

$$\mathrm{mes}[\underline{\mathcal{B}}_2^{\flat}] \lesssim \kappa_0^2 A \cdot \epsilon_2 \epsilon_1^{-2}.$$

Moreover, $\underline{\mathcal{R}} \setminus (\underline{\mathcal{B}}_1^{\flat} \cup \underline{\mathcal{B}}_2^{\flat})$ is covered by squares of the form $\underline{\mathcal{R}}_2 = I_2 \times J_2$, of size $|I_2|, |J_2| \sim \kappa_2$.

The total number of such squares is about

$$(\kappa_1 \kappa_2^{-1})^2 \cdot (\kappa_0 \kappa_1^{-1})^2 = (\kappa_0 \kappa_2^{-1})^2.$$

On each of these squares,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_2} |\, \partial^{(m_1-1, m_2-1)} v(\underline{x})| \geq \epsilon_2$$

It is clear how this procedure continues. Perform it for $m-1$ steps. We will get sets $\underline{\mathcal{B}}_1^{\flat}, \ldots, \underline{\mathcal{B}}_{m-1}^{\flat}$ such that $\underline{\mathcal{R}} \setminus (\underline{\mathcal{B}}_1^{\flat} \cup \ldots \cup \underline{\mathcal{B}}_{m-1}^{\flat})$ consists of about $(\kappa_0 \kappa_{m-1}^{-1})^2$ squares of the form $\underline{\mathcal{R}}_{m-1} = I_{m-1} \times J_{m-1}$, of size $|I_{m-1}|, |J_{m-1}| \sim \kappa_{m-1}$. On each of these squares,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_{m-1}} |\, \partial_{x_2} v(\underline{x})| \geq \epsilon_{m-1} \quad \text{or} \quad \min_{\underline{x} \in \underline{\mathcal{R}}_{m-1}} |\, \partial_{x_1} v(\underline{x})| \geq \epsilon_{m-1}.$$

Step m.   Assume the former inequality above and apply Lemma 5.4 one more time. Let

$$f_m(\underline{x}) := v(\underline{x}) - E,$$

for some fixed energy E with $|E| \le 2A$ (the estimates will not depend on E). Then for each of the squares $\underline{\mathcal{R}}_{m-1}$ from the previous step,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_{m-1}} |\partial_{x_2} f_m(\underline{x})| = \min_{\underline{x} \in \underline{\mathcal{R}}_{m-1}} |\partial_{x_2} v(\underline{x})| \ge \epsilon_{m-1}.$$

Apply Lemma 5.4 to the function f_m with the following data:

$\underline{\mathcal{R}}_{m-1}, \kappa_{m-1}, \epsilon_{m-1}$ from the previous step, $\epsilon_m < \frac{\epsilon_{m-1}\kappa_{m-1}}{4}$, and $\kappa_m \sim \epsilon_m A^{-1}$,

where ϵ_m will be chosen later.

We get a set $\underline{\mathcal{B}}_m$, $\mathrm{mes}[\underline{\mathcal{B}}_m] \lesssim \kappa_{m-1}\epsilon_m\epsilon_{m-1}^{-1}$ such that $\underline{\mathcal{R}}_{m-1} \setminus \underline{\mathcal{B}}_m$ is a union of about $(\kappa_{m-1}\kappa_m^{-1})^2$ squares of the form $\underline{\mathcal{R}}_m = I_m \times J_m$, of size $|I_m|, |J_m| \sim \kappa_m$. For each of these squares,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_m} |f_m(\underline{x})| \ge \epsilon_m,$$

which means

$$\min_{\underline{x} \in \underline{\mathcal{R}}_m} |v(\underline{x}) - E| \ge \epsilon_m.$$

If we do this for each of the $\sim (\kappa_0 \kappa_{m-1}^{-1})^2$ squares resulting from the previous step, and if we put together all the corresponding 'bad' sets, we conclude.

There is a set $\underline{\mathcal{B}}_m^\flat \subset \underline{\mathcal{R}}$ such that

$$\mathrm{mes}[\underline{\mathcal{B}}_m^\flat] \lesssim \kappa_{m-1}\epsilon_m\epsilon_{m-1}^{-1} \cdot (\kappa_0\kappa_{m-1}^{-1})^2 = \kappa_0\epsilon_m\epsilon_{m-1}^{-1}\kappa_{m-1}^{-1} \sim \kappa_0^2 A \cdot \epsilon_m\epsilon_{m-1}^{-2}.$$

Hence the total measure of the 'bad' set in Step m is

$$\mathrm{mes}[\underline{\mathcal{B}}_m^\flat] \lesssim \kappa_0^2 A \cdot \epsilon_m\epsilon_{m-1}^{-2}.$$

Moreover, $\underline{\mathcal{R}} \setminus (\underline{\mathcal{B}}_1^\flat \cup \underline{\mathcal{B}}_2^\flat \ldots \cup \underline{\mathcal{B}}_m^\flat)$ is covered by squares of the form $\underline{\mathcal{R}}_m = I_m \times J_m$, of size $|I_m|, |J_m| \sim \kappa_m$.

The total number of such squares is about

$$(\kappa_{m-1}\kappa_m^{-1})^2 \cdot (\kappa_0\kappa_{m-1}^{-1})^2 = (\kappa_0\kappa_m^{-1})^2.$$

On each of these squares,

$$\min_{\underline{x} \in \underline{\mathcal{R}}_m} |v(\underline{x}) - E| \ge \epsilon_m.$$

Therefore, the total measure of the bad set from all steps is

$$\text{mes}[\underline{\mathcal{B}}_1^\flat \cup \underline{\mathcal{B}}_2^\flat \ldots \cup \underline{\mathcal{B}}_m^\flat] \lesssim \kappa_0^2 A \cdot [\epsilon_1 \epsilon_0^{-2} + \epsilon_2 \epsilon_1^{-2} + \ldots \epsilon_m \epsilon_{m-1}^{-2}.] \qquad (5.9)$$

We choose

$$\epsilon_j := \epsilon^{1/3^{m-j}} \quad \text{for } 1 \le j \le m$$

If $\epsilon < \epsilon^*(c, m)$, then $\epsilon_0 \sim c > \epsilon^{1/3^m}, \epsilon_0^{-2} \sim c^{-2} < \epsilon^{-2}$, so there is no harm in also putting (for simplicity) $\epsilon_0 = \epsilon^{1/3^m}$.

It is a simple calculation to see that for any $\epsilon < \epsilon^*(m, A)$, we have $\epsilon_{j+1} < \frac{\epsilon_j \kappa_j}{4}$ for all $j = \overline{0 \ldots m-1}$, which allows our inductive process to work.

Note that $\epsilon_j^3 = \epsilon_{j+1}$ so $\epsilon_{j+1} \epsilon_j^{-2} = \epsilon_j$. This implies

$$\epsilon_1 \epsilon_0^{-2} + \epsilon_2 \epsilon_1^{-2} + \ldots \epsilon_m \epsilon_{m-1}^{-2} = \epsilon_0 + \epsilon_1 + \ldots + \epsilon_{m-1} \le m\epsilon_0 = m \cdot \epsilon^{1/3^m}.$$

From (5.9) it follows that the total measure of the bad set inside the square $\underline{\mathcal{R}}$ is

$$\text{mes}[\underline{\mathcal{B}}_1^\flat \cup \underline{\mathcal{B}}_2^\flat \ldots \cup \underline{\mathcal{B}}_m^\flat] \lesssim \kappa_0^2 A m \cdot \epsilon^{1/3^m}.$$

There are about $(\frac{2A}{c})^2 = \kappa_0^{-2}$ such squares.

Therefore, outside a bad set $\underline{\mathcal{B}}$, $\text{mes}[\underline{\mathcal{B}}] < A m \cdot e^{1/3^m}$, we have $|v(\underline{x}) - E| \ge \epsilon$, which proves (5.1) with $C \sim A m$ and $b = \frac{1}{3^m}$.

$\square$

Remark 5.1. The exponent b in (5.8) is related to the Łojasiewicz exponent of the function v (see [17] and [10]). Determining the optimal exponent in such an inequality is an interesting problem in itself, and has been studied extensively for polynomials and analytic functions. It is clear that for a polynomial, the Łojasiewicz exponent should be related to its degree d, and it is in fact shown to be $O(\frac{1}{d^2})$ with explicit underlying constants (see [10] and [17]). The proof of the Łojasiewicz inequality for analytic functions in [12] (see Lemma 11.4 there) does not provide an explicit value for the exponent, but Theorem 4 in [21] provides a scheme for computing it via the Newton distance of v.

In our proof for smooth, transversal functions, we obtain the exponent $\frac{1}{3^m}$, where m is the maximum number of partial derivatives needed for transversality. If v were a polynomial of degree d, then m would be d, which shows that our estimate is very wasteful (we have obtained a better estimate, $O(\frac{1}{m})$, for one-variable functions, see Lemma 5.3 in [16]). This, however, seems to be the only such estimate available now for non-analytic functions of two variables.

A similar argument can be made for functions of more than two variables, so (5.8) will hold for such functions as well.

6. Large deviation theorem, the proof of main results

Using induction on the scale N, we will prove the large deviation estimate (2.2) for the logarithmic average of transfer matrices,

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 : \left|\frac{1}{N}\log\|M_N(\underline{x}, E)\| - L_N(E)\right| > N^{-\tau}\right] < e^{-N^\sigma},$$

as well as a lower bound on the mean of the quantities

$$L_N(E) \geq \gamma_N \log|\lambda|.$$

The base step of the induction uses the quantitative description (5.8) of the transversality condition (1.12) on the potential function, and the large size of the coupling constant. The inductive step uses only the regularity of the potential function via Lemma 3.1, which provides a good approximation of these logarithmic averages by pluri-subharmonic functions.

Lemma 6.1 (base step of the induction). *Assume that $v(\underline{x})$ is smooth and satisfies the transversality condition* (1.12). *Then given any constant $C > 0$, there are positive constants λ_1 and B which depend on v and C, such that for any scale N_0, for any λ subject to $|\lambda| \geq \max\{\lambda_1, N_0^B\}$ and, for any $E \in \mathbb{R}$,*

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 : \left|\frac{1}{N_0}\log\|M_{N_0}(\underline{x}, \lambda, E)\| - L_{N_0}(\lambda, E)\right| > \frac{1}{20}S(\lambda)\right] < N_0^{-C}. \quad (6.1)$$

Furthermore, for these λ, N_0 and for all E,

$$L_{N_0}(\lambda, E) \geq \frac{1}{2}S(\lambda), \quad (6.2)$$

and

$$L_{N_0}(\lambda, E) - L_{2N_0}(\lambda, E) \leq \frac{1}{80}S(\lambda). \quad (6.3)$$

Proof. The proof of this result is similar to the analytic potential function case. That is because the only fact about analyticity needed here is the Łojasiewicz inequality (5.8), which holds for any non-constant analytic functions, and which we have established in section 5 for smooth functions satisfying the transversality condition (1.12). We will then omit the proof, but the reader is referred to the proof of Lemma 2.10 in [6] for details. □

We will now explain the idea of the proof of the inductive step.

If at scale N_0 we apply the almost invariance property (3.18) n times and then average, we get

$$\Big| L_{N_0}(\underline{x}) - \frac{1}{n} \sum_{j=0}^{n-1} L_{N_0}(\mathrm{T}^j \underline{x}) \Big| \lesssim \frac{n S(\lambda)}{N_0}, \qquad (6.4)$$

so using the approximation (3.16), we also get

$$\Big| u_{N_0}(\underline{x}) - \frac{1}{n} \sum_{j=0}^{n-1} u_{N_0}(\mathrm{T}^j \underline{x}) \Big| \lesssim \frac{n S(\lambda)}{N_0}. \qquad (6.5)$$

To have a decay above, we need to take a smaller number of shifts $n \ll N_0$.

Apply the estimate (4.1) on averages of shifts of pluri-subharmonic functions to $u_{N_0}(x)$ and get

$$\mathrm{mes}\Big[\underline{x} \in \mathbb{T}^2 \colon \Big| \frac{1}{n} \sum_{j=0}^{n-1} u_{N_0}(\mathrm{T}^j \underline{x}) - \langle u_{N_0} \rangle \Big| > \frac{S}{\rho_{N_0}} n^{-\tau_0} \Big] < e^{-n^{\sigma_0}}. \qquad (6.6)$$

We may combine (6.5) and (6.6) to directly obtain a large deviation estimate for $u_{N_0}(x)$ and then, via the approximations (3.16) and (3.17) to obtain the LDT for $L_{N_0}(x)$, *only* when the deviation $\frac{S}{\rho_{N_0}} n^{-\tau_0} \ll 1$. In other words, this approach works only when the scaling factor $\frac{S}{\rho_{N_0}}$ is not too large to cancel the decay $n^{-\tau_0}$. This is the case of the single or multi-frequency shift model with *analytic* potential (see [4], [5]) where $\frac{S}{\rho_{N_0}} = \frac{S}{\rho}$ is just a constant depending on the potential function v. This approach also works for the *single*-frequency model with potential function in a Gevrey class of order $s < 2$, since in this case sharper estimates than (6.6) are available for averages of shifts of single-variable subharmonic functions (see [16]). This approach fails for the skew-shift model (whether the potential function is analytic or Gevrey) and also for the multi-frequency model with Gevrey potential function, because the size ρ_{N_0} of the subharmonic extension depends on the scale N_0.

In order to beat the scaling factor $\frac{S}{\rho_{N_0}}$ when applying the estimate (6.6) to a transfer matrix substitute $u_{N_0}(\underline{x})$ at scale N_0, we need to consider a large number of shifts $n \gg N_0$. The averages of shifts thus obtained will be close to the mean $\langle u_{N_0} \rangle$. Moreover, we will get

$$L_{N_0} \overset{(1)}{\approx} \langle u_{N_0} \rangle \overset{(2)}{\approx} \frac{1}{n} \sum_{j=0}^{n-1} u_{N_0}(\mathrm{T}^j \underline{x})$$

$$\overset{(3)}{\approx} \frac{1}{n} \sum_{j=0}^{n-1} \frac{1}{N_0} \log \| M_{N_0}(\mathrm{T}^j \underline{x}) \| \overset{(4)}{\approx} \frac{1}{n N_0} \log \| M_{n N_0}(\underline{x}) \|.$$

The first approximation above is just (3.17). The second is exactly (6.6). The third is due to (3.16). The last approximation above essentially says that

$$\prod_{j=0}^{n-1} \|M_{N_0}(\mathrm{T}^j \underline{x})\| \approx \| \prod_{j=0}^{n-1} M_{N_0}(\mathrm{T}^j \underline{x})\| \approx \|M_{nN_0}(\underline{x})\|,$$

or in other words, that the product of the norms of certain transfer matrices is approximately equal to the norm of the product of these matrices, the latter giving us the transfer matrix at the larger scale nN_0.

If these heuristics were true, then for $n \gg N_0$ we would get

$$L_{N_0} \approx \frac{1}{nN_0} \log \|M_{nN_0}(\underline{x})\|,$$

which would establish the large deviation estimate for transfer matrices at a larger scale nN_0.

The avalanche principle, which is a deterministic result, describes how estimates on the norms of individual (and of products of two consecutive) $\mathrm{SL}_2(\mathbb{R})$ matrices can lead to estimates on the norm of the product of all matrices (see [12] and [4]), thus providing the basis for establishing the above heuristics. It requires a uniform lower bound on the norms of individual matrices in the product, as well as knowing that the norm of the product of any two consecutive matrices is comparable to the product of their norms.

The following lemma provides the inductive step in proving the LDT for an increasing sequence of scales N. It also provides the inductive step in proving the positivity and continuity of the Lyapunov exponent. The proof of this lemma is based on the heuristics described above, and combines the averages of shifts estimate (4.1), the almost invariance property (3.18) and the avalanche principle (see Proposition 2.2 in [12]).

Before stating the lemma let us describe the various parameters and constants that will appear.

List of constants and parameters

- $s > 1$ is the order of the Gevrey class.

- $\delta = 2(s-1)$ refers to the size ($\approx N^{-\delta}$) of the holomorphic extensions of the transfer matrix substitutes.

- $D := 2\delta + 8$, $A := \max\{\frac{2(\delta+1)}{\tau_0}, 2\}$ are some well chosen powers of the scale N, τ_0 is the exponent from (4.1).

- $\gamma > \frac{1}{4}$ is a fixed number.

Note that all these constants are either universal or depend on the order s of the Gevrey class.

- λ, E are fixed parameters such that $|E| \leq |\lambda|B + 2$, and $B := \sup_{\underline{x} \in \mathbb{T}} |v(\underline{x})|$.

- The transformation $T = S_\omega$ where $\omega \in DC_\kappa$ or $T = T_{\underline{\omega}}$ where $\underline{\omega} \in DC_\kappa$ for some $\kappa > 0$.

- $N_{00} = N_{00}(s, \kappa, B)$ is a sufficiently large integer, such that the asymptotic behavior of various powers and exponentials applies to N_{00} and such that (4.1) holds for N_{00} shifts.

Lemma 6.2 (the inductive step). *Consider two scales N_0 and N such that $N_0 \geq N_{00}$, (3.16) holds at scale N_0, that is*

$$N_0 \geq S(\lambda) \iff |\lambda| \leq e^{N_0} \tag{6.7}$$

and

$$N_0^A \leq N \leq e^{N_0}. \tag{6.8}$$

Assume that a weak LDT holds at scales N_0 and $2N_0$:

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{N_0} \log \|M_{N_0}(\underline{x}, \lambda, E)\| - L_{N_0}(\lambda, E)\right| > \frac{\gamma}{10} S(\lambda)\right] < N^{-D} \tag{6.9}$$

and

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{2N_0} \log \|M_{2N_0}(\underline{x}, \lambda, E)\| - L_{2N_0}(\lambda, E)\right| > \frac{\gamma}{10} S(\lambda)\right] < N^{-D}, \tag{6.10}$$

and that the means L_{N_0}, L_{2N_0} have a lower bound and are close to each other:

$$L_{N_0}(\lambda, E), \; L_{2N_0}(\lambda, E) \geq \gamma S(\lambda) \tag{6.11}$$

and

$$L_{N_0}(\lambda, E) - L_{2N_0}(\lambda, E) \leq \frac{\gamma}{40} S(\lambda). \tag{6.12}$$

Then similar (but stronger) estimates hold at the larger scale N:

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{N} \log \|M_N(\underline{x}, \lambda, E)\| - L_N(\lambda, E)\right| > S(\lambda)N^{-\tau}\right] < e^{-N^\sigma}, \tag{6.13}$$

$$L_N(\lambda, E) \geq \gamma S(\lambda) - 2[L_{N_0}(\lambda, E) - L_{2N_0}(\lambda, E)] - C_0 S(\lambda) N_0 N^{-1}, \tag{6.14}$$

and

$$L_N(\lambda, E) - L_{2N}(\lambda, E) \leq C_0 S(\lambda) N_0 N^{-1}, \tag{6.15}$$

for some positive absolute constants C_0, τ, σ.

Proof. The parameters λ, E and the transformation $T = S_\omega$ or $T = T_{\underline{\omega}}$ are fixed, so they can be suppressed from notations. For instance $M_N(\underline{x}) = M_N(\underline{x}, \lambda, E)$, $S(\lambda) = S$ etc.

- We can assume without loss of generality that N is a multiple of N_0, that is, that $N = n \cdot N_0$. Indeed, if $N = n \cdot N_0 + r, 0 \leq r < N_0$, then

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{n \cdot N_0} \log \|M_{n \cdot N_0}(\underline{x})\| \right| \leq 2S N_0 N^{-1}. \tag{6.16}$$

Therefore, if we prove (6.14), (6.15), and (6.13) at scale $n \cdot N_0$, then they hold at scale N too.

To prove (6.16), first note that $M_N(\underline{x}) = B(\underline{x}) \cdot M_{n \cdot N_0}(\underline{x})$, where

$$B(\underline{x}) := \prod_{j=N}^{n \cdot N_0+1} A(T^j \underline{x}) = \prod_{j=n \cdot N_0+r}^{n \cdot N_0+1} A(T^j \underline{x}),$$

so

$$\|B(\underline{x})\| \leq e^{r \cdot S} \leq e^{N_0 \cdot S} \quad \text{and} \quad \|B(\underline{x})^{-1}\| \leq e^{r \cdot S} \leq e^{N_0 \cdot S}.$$

Since $\|M_{n \cdot N_0}(\underline{x})\| \geq 1$ and $\|M_N(\underline{x})\| \geq 1$, it follows that

$$\frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{n \cdot N_0} \log \|M_{n \cdot N_0}(\underline{x})\|$$

$$= \frac{1}{n \cdot N_0} \log \frac{\|M_N(\underline{x})\|^{\frac{n \cdot N_0}{N}}}{\|M_{n \cdot N_0}(\underline{x})\|}$$

$$\leq \frac{1}{n \cdot N_0} \log \frac{\|B(\underline{x})\|^{\frac{n \cdot N_0}{N}} \cdot \|M_{n \cdot N_0}(\underline{x})\|^{\frac{n \cdot N_0}{N}}}{\|M_{n \cdot N_0}(\underline{x})\|}$$

$$\leq \frac{1}{n \cdot N_0} \log (e^{N_0 S})^{\frac{n \cdot N_0}{N}}$$

$$= S N_0 N^{-1}.$$

Similarly

$$\frac{1}{n \cdot N_0} \log \|M_{n \cdot N_0}(\underline{x})\| - \frac{1}{N} \log \|M_N(\underline{x})\|$$

$$= \frac{1}{n \cdot N_0} \log \frac{\|M_{n \cdot N_0}(\underline{x})\|}{\|M_N(\underline{x})\|^{\frac{n \cdot N_0}{N}}}$$

$$= \frac{1}{n \cdot N_0} \log \left[\left(\frac{\|M_{n \cdot N_0}(\underline{x})\|}{\|M_N(\underline{x})\|} \right)^{\frac{n \cdot N_0}{N}} \cdot \|M_{n \cdot N_0}(\underline{x})\|^{\frac{r}{N}} \right]$$

$$\leq \frac{1}{n \cdot N_0} \log [\|(B(\underline{x}))^{-1}\|^{\frac{n \cdot N_0}{N}} \cdot \|M_{n \cdot N_0}(\underline{x})\|^{\frac{r}{N}}]$$

$$\leq \frac{1}{n \cdot N_0} \log [(e^{N_0 S})^{\frac{n \cdot N_0}{N}} \cdot (e^{n N_0 S})^{\frac{N_0}{N}}]$$

$$= 2 S N_0 N^{-1}$$

and inequality (6.16) now follows.

- We are going to show that (6.8)–(6.12) allow us to apply the avalanche principle to the "blocks" $M_{N_0}(\mathrm{T}^{(j-1)N_0} \underline{x})$, for $j = \overline{1, n}$. Each of these blocks is a product of N_0 matrices, and they multiply up to $M_N(\underline{x})$.

Denote the set in (6.9) by B_{N_0} and similarly the set in (6.10) by B_{2N_0}. If $\underline{x} \notin B_{N_0}$ then using (6.9), (6.11) and (6.8) we get

$$\|M_{N_0}(\underline{x})\| > e^{-\frac{\gamma}{10} S N_0 + L_{N_0} \cdot N_0} \geq e^{\frac{9\gamma}{10} S N_0} =: \mu > e^{N_0} \geq N > n,$$

so

$$\|M_{N_0}(\underline{x})\| \geq \mu \geq n, \quad \text{if } \underline{x} \notin B_{N_0}. \tag{6.17}$$

For $1 \leq j \leq n = \frac{N}{N_0}$ consider

$$A_j = A_j(\underline{x}) := M_{N_0}(\mathrm{T}^{(j-1)N_0} \underline{x}).$$

Then (6.17) implies

$$\min_{1 \leq j \leq n} \|A_j(\underline{x})\| \geq \mu, \quad \text{for all } x \notin \bigcup_{j=0}^{n} \mathrm{T}^{-jN_0} B_{N_0}. \tag{6.18}$$

Since $A_{j+1}(\underline{x}) \cdot A_j(\underline{x}) = M_{2N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})$, using equations (6.9), (6.10), and (6.12), for $\underline{x} \notin \bigcup_{j=0}^{n}(\mathrm{T}^{-jN_0}B_{N_0}) \cup \bigcup_{j=0}^{n}(\mathrm{T}^{-jN_0}B_{2N_0})$ (which is a set of measure $< 2N^{-D} \cdot N = 2N^{-D+1}$), we have

$$\log \|A_{j+1}(\underline{x})\| + \log \|A_j(\underline{x})\| - \log \|A_{j+1}(\underline{x}) \cdot A_j(\underline{x})\|$$

$$= \log \|M_{N_0}(\mathrm{T}^{jN_0}\underline{x})\| + \log \|M_{N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})\|$$
$$- \log \|M_{2N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})\|$$

$$\leq N_0\Big(L_{N_0} + \frac{S\gamma}{10}\Big) + N_0\Big(L_{N_0} + \frac{S\gamma}{10}\Big) + 2N_0\Big(\frac{S\gamma}{10} - L_{2N_0}\Big)$$

$$= 2N_0(L_{N_0} - L_{2N_0}) + \frac{4S\gamma}{10}N_0 \leq \frac{9S\gamma}{20}N_0 = \frac{1}{2}\log\mu.$$

Therefore,

$$\log \|A_{j+1}(\underline{x})\| + \log \|A_j(\underline{x})\| - \log \|A_{j+1}(\underline{x}) \cdot A_j(\underline{x})\| \leq \frac{1}{2}\log\mu, \quad (6.19)$$

for $\underline{x}$ outside a set of measure $< 2N^{-D+1}$.

Estimates (6.18), (6.19) are exactly the assumptions in the avalanche principle (Proposition 2.2 in [12]). We then conclude

$$|\log \|A_n(\underline{x}) \cdot \ldots \cdot A_1(\underline{x})\| + \sum_{j=2}^{n-1}\log \|A_j(\underline{x})\| - \sum_{j=1}^{n-1}\log \|A_{j+1}(\underline{x}) \cdot A_j(\underline{x})\|| \lesssim \frac{n}{\mu}$$
$$(6.20)$$

for $\underline{x}$ outside a set of measure $< 2N^{-D+1}$.

Hence, since $N = n \cdot N_0$ and $A_n(\underline{x}) \cdot \ldots \cdot A_1(\underline{x}) = M_N(\underline{x})$,

$$|\log \|M_N(\underline{x})\| + \sum_{j=2}^{n-1}\log |M_{N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})|$$

$$- \sum_{j=1}^{n-1}\log \|M_{2N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})\|| \lesssim \frac{n}{\mu}.$$

Therefore

$$\Big|\frac{1}{N}\log \|M_N(\underline{x})\| + \frac{1}{n}\sum_{j=2}^{n-1}\frac{1}{N_0}\log \|M_{N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})\|$$

$$(6.21)$$

$$- \frac{2}{n}\sum_{j=1}^{n-1}\frac{1}{2N_0}\log \|M_{2N_0}(\mathrm{T}^{(j-1)N_0}\underline{x})\|\Big| \lesssim \frac{1}{\mu}.$$

■ We will go from averages of n blocks in (6.21), to averages of N shifts. In (6.21) replace $\underline{x}$ by $\underline{x}, \mathrm{T}\underline{x}, \dots \mathrm{T}^{N_0-1}\underline{x}$ and then average (i.e. add up all these N_0 inequalities and divide by N_0) to get

$$
\Big| \frac{1}{N_0} \sum_{j=0}^{N_0-1} \frac{1}{N} \log \|M_N(\mathrm{T}^j\underline{x})\| + \frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{N_0} \log \|M_{N_0}(\mathrm{T}^j\underline{x})\|
$$
$$
- \frac{2}{N} \sum_{j=0}^{N-1} \frac{1}{2N_0} \log \|M_{2N_0}(\mathrm{T}^j\underline{x})\| \Big| \lesssim \frac{1}{\mu}. \tag{6.22}
$$

The almost invariance property - Lemma (3.18) implies

$$
\Big| \frac{1}{N} \log \|M_N(\underline{x})\| - \frac{1}{N_0} \sum_{j=0}^{N_0-1} \frac{1}{N} \log \|M_N(\mathrm{T}^j\underline{x})\| \Big| \lesssim \frac{SN_0}{N}. \tag{6.23}
$$

From (6.22) and (6.23) we get

$$
\Big| \frac{1}{N} \log \|M_N(\underline{x})\| + \frac{1}{N} \sum_{j=0}^{N-1} \frac{1}{N_0} \log \|M_{N_0}(\mathrm{T}^j\underline{x})\|
$$
$$
- \frac{2}{N} \sum_{j=0}^{N-1} \frac{1}{2N_0} \log \|M_{2N_0}(\mathrm{T}^j\underline{x})\| \Big|,
$$
$$
\lesssim \frac{SN_0}{N} + \frac{1}{\mu} \lesssim SN_0 N^{-1}
$$

for

$$
\underline{x} \notin B_1 := \bigcup_{j=0}^{N} (T^{-j} B_{N_0}) \cup \bigcup_{j=0}^{n} (T^{-j} B_{2N_0}),
$$

where $\mathrm{mes}[B_1] < 2N^{-D+1}$.

Integrating the left hand side of (6.24) in $\underline{x}$, we get

$$
|L_N + L_{N_0} - 2L_{2N_0}| < CSN_0 N^{-1} + 4S \cdot 2N^{-D+1} < C_0 SN_0 N^{-1}, \tag{6.24a}
$$
$$
L_N + L_{N_0} - 2L_{2N_0} > -C_0 SN_0 N^{-1}, \tag{6.24b}
$$

and

$$
L_N > L_{N_0} - 2(L_{N_0} - L_{2N_0}) - C_0 SN_0 N^{-1}
$$
$$
> \gamma S - 2(L_{N_0} - L_{2N_0}) - C_0 SN_0 N^{-1}, \tag{6.24c}
$$

which proves (6.14).

Clearly all the arguments above work for N replaced by $2N$, so we get the analogue of (6.24)

$$|L_{2N} + L_{N_0} - 2L_{2N_0}| < C_0 S N_0 N^{-1}. \tag{6.25}$$

From (6.24) and (6.25) we obtain

$$L_N - L_{2N} \leq C_0 S N_0 N^{-1},$$

which is exactly (6.15).

■ To prove the LDT (6.13) at scale N, we are going to apply the estimate (4.1) on averages of shifts of pluri-subharmonic functions to the transfer matrix substitutes u_{N_0} and u_{2N_0}. Their widths of subharmonicity in each variable are $\rho_{N_0}, \rho_{2N_0} \approx N_0^{-\delta-1}$ and they are uniformly bounded by S.

Using (3.16) which holds at scales N_0 and $2N_0$ due to (6.7), we can 'substitute' in (6.24)

$$\frac{1}{N_0} \log \|M_{N_0}(\mathrm{T}^j(\underline{x})\| \quad \text{by } u_{N_0}(\mathrm{T}^j \underline{x})$$

and

$$\frac{1}{2N_0} \log \|M_{2N_0}(\mathrm{T}^j(\underline{x})\| \quad \text{by } u_{2N_0}(\mathrm{T}^j \underline{x}),$$

and get, for $\underline{x} \notin B_1$,

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| + \frac{1}{N} \sum_{j=0}^{N-1} u_{N_0}(\mathrm{T}^j \underline{x}) - \frac{2}{N} \sum_{j=0}^{N-1} u_{2N_0}(\mathrm{T}^j \underline{x}) \right| \lesssim S N_0 N^{-1}. \tag{6.26}$$

Applying (4.1) to u_{N_0} and u_{2N_0} we get

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{N} \sum_{j=0}^{N-1} u_{N_0}(\mathrm{T}^j \underline{x}) - \langle u_{N_0} \rangle \right| > S \cdot N_0^{\delta+1} \cdot N^{-\tau_0} \right] < e^{-N^{\sigma_0}} \tag{6.27}$$

and

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{N} \sum_{j=0}^{N-1} u_{2N_0}(\mathrm{T}^j \underline{x}) - \langle u_{2N_0} \rangle \right| > S \cdot N_0^{\delta+1} \cdot N^{-\tau_0} \right] < e^{-N^{\sigma_0}}. \tag{6.28}$$

Denote the union of the two sets in (6.27) and (6.28) by B_2.

Since N satisfies (6.8),

$$S \cdot N_0^{\delta+1} \cdot N^{-\tau_0} < S \cdot (N^{1/A})^{\delta+1} \cdot N^{-\tau_0} < S \cdot N^{-\tau_1}, \quad \text{where } \tau_1 < \frac{\tau_0}{2},$$

so, by (6.26), (6.27), and (6.28), we get

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| + \langle u_{N_0} \rangle - 2\langle u_{2N_0} \rangle \right| \lesssim S N_0 N^{-1} + S \cdot N^{-\tau_1} \\ \lesssim S \cdot N^{-\tau_1}, \tag{6.29}$$

for $\underline{x} \notin B := B_1 \cup B_2$, where

$$\text{mes}[B] < 2N^{-D+1} + 2e^{-N^\sigma} < 3N^{-D+1} < N^{-D+2} p.$$

By (3.17) at scales N_0, $2N_0$ and taking into account (6.8), estimate (6.29) becomes

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| + L_{N_0} - 2L_{2N_0} \right| < 2S \cdot N^{-\tau_1} + 2e^{-N_0^2} \\ < 3S N^{-\tau_1}, \tag{6.30}$$

provided $x \notin B$.

Combine (6.30) with (6.24) to get

$$\left| \frac{1}{N} \log \|M_N(\underline{x})\| - L_N \right| < C_0 S N_0 N^{-1} + 3S \cdot N^{-\tau_1} \\ < S \cdot N^{-\tau_2}, \tag{6.31}$$

for all $\underline{x} \notin B$, where $\text{mes}[B] < N^{-D+2}$ and $\tau_2 < \tau_1$.

However, (6.31) is not exactly what we need in order to prove the estimate (6.13). We have to prove an estimate like (6.31) for $\underline{x}$ outside an exponentially small set, and we only have it outside a polynomially small set. To boost this estimate, we employ again Lemma 4.1.

From (6.31), using again (3.16), (3.17) at scale N, we get

$$\text{mes}[\underline{x} \in \mathbb{T}^2 : |u_N(\underline{x}) - \langle u_N \rangle| > S \cdot N^{-\tau_2}] < N^{-D+2}. \tag{6.32}$$

We apply Lemma 4.1 to

$$u(\underline{x}) := \frac{1}{S} u_N(\underline{x}),$$

which is a pluri-subharmonic function on the strip $\underline{\mathcal{A}}_{\rho_N}$, with upper bound $B = 1$ on this strip.

Estimate (6.32) implies

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u(\underline{x}) - \langle u \rangle| > N^{-\tau_2}] < N^{-D+2}. \qquad (6.33)$$

Then, for $\epsilon_0 := N^{-\tau_2}, \epsilon_1 := N^{-D+2}, B = 1, \rho = \rho_N \approx N^{-\delta-1}$,

$$\epsilon_0{}^{1/4} + \sqrt{\frac{B}{\rho} \frac{\epsilon_1{}^{1/4}}{\epsilon_0{}^{1/2}}} = N^{-\tau_2/4} + N^{\frac{\delta+1}{2}} N^{-\frac{D+2}{4}} N^{\tau_2/2}$$

$$= N^{-\tau_2/4} + N^{-1} N^{\tau_2/2}$$

$$< N^{-\sigma_1},$$

for some positive constant σ_1.

The conclusion (4.6) of Lemma 4.1 then boosts (6.32) from a small deviation outside a polynomially small set, to one outside an exponentially small set, amid a small power loss in the deviation

$$\mathrm{mes}[\underline{x} \in \mathbb{T}^2 \colon |u_N(\underline{x}) - \langle u_N \rangle| > S \, N^{-\tau_2/4}] < e^{-cN^{\sigma_1}} < e^{-N^{\sigma}}, \quad (6.34)$$

which proves estimate (6.13).   □

Remark 6.1. The scaling factor $\sqrt{\frac{B}{\rho}}$ in estimate (4.6) of Lemma 4.1 is what prevents this approach via polynomial approximation to extend to more general Carleman classes of potential functions. This is because when the estimates on the Fourier coefficients of the potential function are weaker than estimate (1.11) for Gevrey functions, the size $\rho = \rho_N$ of the holomorphic extension of the Nth transfer matrix substitute will cancel any decay in the expression $\sqrt{\frac{B}{\rho}} \frac{\epsilon_1{}^{1/4}}{\epsilon_0{}^{1/2}}$.

We will combine the base step (Lemma 6.1) with the inductive step (see Lemma 6.2) to prove the large deviation estimate for transfer matrices and the positivity of the Lyapunov exponent. The proof of the LDT will also provide us with the major ingredient for deriving the continuity of the Lyapunov exponent.

Theorem 6.1. *Consider the Schrödinger operator* (1.1) *on* $l^2(\mathbb{Z})$,

$$[H(\underline{x}) \, \psi]_n := -\psi_{n+1} - \psi_{n-1} + \lambda \, v(\mathrm{T}^n \underline{x}) \, \psi_n,$$

where the transformation T *is either the skew-shift* (1.4) *or the multi-frequency shift* (1.6). *Assume that for some* $\kappa > 0$ *the underlying frequency satisfies the Diophantine condition* DC_κ *described in* (1.8) *or* (1.9) *respectively.*

Assume moreover that the potential function $v(\underline{x})$ *belongs to a Gevrey class* $G^s(\mathbb{T}^2)$ *and that it is transversal as in* (1.12).

Then there exists $\lambda_0 = \lambda_0(v, \kappa)$ so that for every fixed λ with $|\lambda| \geq \lambda_0$ and for every energy E,

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{N}\log\|M_N(\underline{x}, \lambda, E)\| - L_N(\lambda, E)\right| > N^{-\tau}\right] < e^{-N^\sigma} \quad (6.35)$$

for some absolute constants $\tau, \sigma > 0$, and for all $N \geq N_0(\lambda, \kappa, v, s)$.

Furthermore, for every such transformation T *and coupling constant λ and for all energies $E \in \mathbb{R}$,*

$$L(\lambda, E) \geq \frac{1}{4}\log|\lambda| > 0. \quad (6.36)$$

Proof. We refer to the list of constants preceding Lemma 6.2.

We use the initial step - Lemma 6.1 at a sufficiently large initial scale $N_0 \geq N_{00} = N_{00}(v)$. We will explain how the scale N_0 is chosen later. We get constants $\lambda_1, B > 0$ such that for every λ with $|\lambda| \geq \max\{\lambda_1, (2N_0)^B\}$ (we want Lemma 6.1 to apply at both scales N_0 and $2N_0$),

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{N_0}\log\|M_{N_0}(\underline{x})\| - L_{N_0}\right| > \frac{1}{20}S\right] < N_0^{-A^2 \cdot D} \\ \leq N^{-D}, \quad (6.37)$$

$$\text{mes}\left[\underline{x} \in \mathbb{T}^2 \colon \left|\frac{1}{2N_0}\log\|M_{2N_0}(\underline{x})\| - L_{2N_0}\right| > \frac{1}{20}S\right] < (2N_0)^{-A^2 \cdot D} \\ \lesssim N^{-D}, \quad (6.38)$$

$$L_{N_0}, \, L_{2N_0} \geq \frac{1}{2}S, \quad (6.39)$$

and

$$L_{N_0} - L_{2N_0} \leq \frac{1}{80}S. \quad (6.40)$$

Of course (6.37) and (6.38) hold provided N satisfies

$$N \leq N_0^{A^2}. \quad (6.41)$$

Estimates (6.37)–(6.40) above are exactly (6.9)–(6.12) (at scale N_0, with $\gamma = \gamma_0 = \frac{1}{2}$) in Lemma 6.2. of the inductive step of LDT.

However, in order to apply this inductive step lemma and obtain similar estimates at the larger scale N, the initial scale N_0 and the disorder λ have to satisfy the condition (6.7). Together with the conditions on λ and N_0 from the initial step (Lemma 6.1), N_0 and λ have to satisfy

$$(2N_0)^B \leq |\lambda| \leq e^{N_0}, \tag{6.42}$$

$$N_0 \geq N_{00}, \tag{6.43}$$

and

$$|\lambda| \geq \lambda_1. \tag{6.44}$$

We want to prove the LDT for every disorder λ large enough, $|\lambda| \geq \lambda_0$ and not just for λ in a bounded interval as in (6.42). To do that, we will have to first choose λ large enough, and then to pick $N_0 = N_0(\lambda) \geq N_{00}$ appropriately. Here is how we can accomplish that.

Condition (6.42) is equivalent to

$$\log|\lambda| \leq N_0 \leq \frac{1}{2}|\lambda|^{1/B}. \tag{6.45}$$

We can find λ_0 large enough, $\lambda_0 = \lambda_0(v, \kappa)$, $\lambda_0 \geq \lambda_1$, so that if $|\lambda| \geq \lambda_0$, then

$$\log|\lambda| \geq N_{00} \quad \text{and} \quad \log|\lambda| \ll \frac{1}{2}|\lambda|^{1/B}. \tag{6.46}$$

Then for every such λ we can pick $N_0 = N_0(\lambda)$ so that (6.45) holds. Combining this with (6.46), we get that (6.42), (6.43), and (6.44) hold.

All the assumptions on the small scale N_0 in the inductive step - Lemma 6.2 hold now, so if we choose the large scale N such that

$$N_0^A \leq N \leq N_0^{A^2} (< e^{N_0}), \tag{6.47}$$

then (6.41) and (6.8) hold, so we can apply Lemma 6.2 to get

$$\mathrm{mes}\left[\underline{x} \in \mathbb{T}^2 : \left|\frac{1}{N}\log\|M_N(\underline{x})\| - L_N\right| > SN^{-\tau}\right] < e^{-N^\sigma}, \tag{6.48}$$

$$L_N \geq \gamma_0 S - 2(L_{N_0} - L_{2N_0}) - C_0 SN_0 N^{-1}, \tag{6.49}$$

and

$$L_N - L_{2N} \leq C_0 SN_0 N^{-1}, \tag{6.50}$$

for some positive absolute constants C_0, τ, σ.

Estimate (6.48) proves the LDT (6.35) at scale N in the range $[N_0^A, N_0^{A^2}]$. If N_1 is in this range, say $N_1 = N_0^A$, then (6.50) and (6.49) imply

$$L_{N_1} - L_{2N_1} \leq C_0 S N_0 N^{-1},$$

and

$$L_{N_1} \geq \gamma_0 S - 3 C_0 S N_0 N^{-1} = \gamma_0 S - 3 C_0 N_0^{-A+1} S =: \gamma_1 \cdot S,$$

where

$$\gamma_1 := \gamma_0 - 3 C_0 N_0^{-A+1} = \frac{1}{2} - 3 C_0 N_0^{-A+1} > \frac{1}{4},$$

provided we chose N_{00} (and so N_0) large enough depending on A, C_0.

Therefore,

$$L_{N_1} \geq \gamma_1 S \tag{6.51}$$

and

$$L_{N_1} - L_{2N_1} \leq C_0 S N_0 N^{-1} = C_0 S N_0^{-A+1} < \frac{1}{160} \cdot S < \frac{\gamma_1}{40} \cdot S,$$

so

$$L_{N_1} - L_{2N_1} < \frac{\gamma_1}{40} \cdot S. \tag{6.52}$$

Since $2N_1 = 2N_0^A$ is in the range $[N_0^A, N_0^{A^2}]$, (6.51) holds at scale $2N_1$ too, so

$$L_{N_1}, L_{2N_1} \geq \gamma_1 S. \tag{6.53}$$

Choosing the next large scale N_2 so that $N_1^A \leq N_2 \leq N_1^{A^2} (< e^{N_1})$, we have $e^{-N_1^\sigma} < N_1^{-A^2 \cdot D} \leq N_2^{-D}$, so (6.48) implies

$$\mathrm{mes} \left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{N_1} \log \|M_{N_1}(\underline{x})\| - L_{N_1} \right| > \frac{1}{20} S \right] < e^{-N_1^\sigma} < N_2^{-D}, \tag{6.54}$$

and

$$\mathrm{mes} \left[\underline{x} \in \mathbb{T}^2 : \left| \frac{1}{2N_1} \log \|M_{2N_1}(\underline{x})\| - L_{2N_1} \right| > \frac{1}{20} S \right] \lesssim N_2^{-D}. \tag{6.55}$$

Estimates (6.54), (6.55), (6.53), and (6.52) are the assumptions in the inductive step – Lemma 6.2 with small scale N_1 and large scale N_2, where $N_2 \in [N_1^A, N_1^{A^2}] = [N_0^{A^2}, N_0^{A^3}]$. Applying Lemma 6.2, we get the LDT (6.35) for N_2 in this range. Moreover,

$$L_{N_2} - L_{2N_2} \leq C_0 S N_1 N_2^{-1}$$

and

$$L_{N_2} \geq \gamma_1 S - 2(L_{N_1} - L_{2N_1}) - C_0 S N_1 N_2^{-1} \geq (\gamma_1 - 3C_0 N_1^{-A+1}) \cdot S =: \gamma_2 \cdot S,$$

where

$$\gamma_2 := \gamma_1 - 3C_0 N_1^{-A+1} \geq \frac{1}{2} - 3C_0 N_0^{-A+1} - 3C_0 N_0^{A\cdot(-A+1)} > \frac{1}{4},$$

again, provided N_{00} (thus N_0) was chosen large enough depending on A, C_0.

Hence we have $L_{N_2} \geq \gamma_2 \cdot S$ and $L_{N_2} - L_{2N_2} \leq \frac{\gamma_2}{40} \cdot S$.

Continuing this inductively, we obtain (6.35) at every scale $N \geq N_0^A$.

Also, at each step k in the induction process, if $N \in [N_k^A, N_k^{A^2}]$, then $L_N \geq \gamma_k \cdot S > \frac{1}{4} \cdot S$ so

$$L = \inf_N L_N \geq \frac{1}{4} \cdot S,$$

and (6.36) is proven.   □

We now prove that the Lyapunov exponent is continuous as a function of the energy.

Theorem 6.2. *Under the same conditions as in Theorem* 6.1 *above, and for any* $|\lambda| \geq \lambda_0(v, \kappa)$, *the Lyapunov exponent $L(E)$ is a continuous function of the energy E with modulus of continuity on each compact interval $\mathcal{E}$ at least,*

$$w(t) = C \left(\log \frac{1}{t} \right)^{-\beta}, \tag{6.56}$$

where $C = C(\mathcal{E}, \lambda, v, \kappa, s)$ and $\beta \in (0, 1)$ is a universal constant that can be chosen, at the expense of C, to be arbitrarily close to 1.

Proof. We will fix λ, T and omit them from notations. We also fix the compact interval $\mathcal{E}$.

It is easy to show (see below) that for every scale N, the functions $L_N(E)$ are (Lipschitz) continuous. To prove that their limits $L(E)$ are also continuous with a certain modulus of continuity, we need a quantitative description of the convergence $L_N(E) \to L(E)$ as $N \to \infty$. The better this rate of convergence, the sharper the modulus of continuity of $L(E)$.

It follows from the proof of Theorem 6.1 above (see (6.50) and the inductive process thereafter) that for every scales N_0 and N such that $N_0 \geq N_{00}(\lambda, v, \kappa)$ and $N_0^A \leq N \leq N_0^{A^2}$,

$$L_N(E) - L_{2N}(E) \lesssim N_0 N^{-1} \leq N^{1/A} N^{-1} =: N^{-\beta},$$

so

$$L_N(E) - L_{2N}(E) \lesssim N^{-\beta} \quad \text{for all } N \geq N_{00}. \tag{6.57}$$

Summing up over dyadic N's we conclude

$$L_N(E) - L(E) \lesssim N^{-\beta}, \quad \text{for all } N \geq N_{00}, \tag{6.58}$$

which is the quantitative convergence we were seeking.

To show that

$$L_N(E) = \frac{1}{N} \int_{\mathbb{T}^2} \log \|M_N(\underline{x}, E)\| \, d\underline{x}$$

are continuous, we use Trotter's formula for the transfer matrix $M_N(\underline{x}, E)$,

$$M_N(\underline{x}, E) - M_N(\underline{x}, E')$$

$$= \sum_{j=1}^N A(\mathrm{T}^N \underline{x}, E) \dots [A(\mathrm{T}^j \underline{x}, E) - A(\mathrm{T}^j \underline{x}, E')] \dots A(\mathrm{T} \underline{x}, E').$$

But

$$A(\mathrm{T}^j \underline{x}, E) - A(\mathrm{T}^j \underline{x}, E') = \begin{bmatrix} E' - E & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$\|A(\mathrm{T}^j \underline{x}, E)\| \leq e^S, \quad \text{for all } E \in \mathcal{E},$$

so

$$\|M_N(\underline{x}, E) - M_N(\underline{x}, E')\| \leq e^{SN} |E - E'|.$$

Therefore, since $\|M_N(\underline{x}, E)\| \geq 1$ and $\|M_N(\underline{x}, E')\| \geq 1$,

$$| \log \|M_N(\underline{x}, E)\| - \log \|M_N(\underline{x}, E')\| |$$

$$\leq \|M_N(x, E) - M_N(x, E')\|$$

$$\leq e^{SN} |E - E'|.$$

Integrating in $\underline{x}$ we obtain

$$|L_N(E) - L_N(E')| \leq e^{SN} |E - E'|, \tag{6.59}$$

which shows Lipschitz continuity for the maps $L_N(E)$.

Combining (6.58) and (6.59) we obtain

$$|L(E) - L(E')| \lesssim N^{-\beta} + e^{SN} |E - E'|, \quad \text{for all } N \geq N_{00}(\lambda, v, \kappa). \tag{6.60}$$

For every such N let

$$|E - E'| \sim e^{-SN} N^{-\beta},$$

so

$$|L(E) - L(E')| \lesssim N^{-\beta}.$$

Since

$$\log \frac{1}{|E - E'|} \sim SN + \beta \log N \lesssim SN,$$

we have

$$N^{-\beta} \sim \left(\frac{1}{S}\right)^{-\beta} \left(\log \frac{1}{|E - E'|}\right)^{-\beta} = C \left(\log \frac{1}{|E - E'|}\right)^{-\beta},$$

where $C = C(\lambda, v, \kappa)$.

We conclude, using the compactness of $\mathcal{E}$, that for some constant $C = C(\mathcal{E}, \lambda, v, \kappa)$, and for a constant β that can be chosen arbitrarily close to 1 by starting off with a large enough constant A,

$$|L(E) - L(E')| < C \left(\log \frac{1}{|E - E'|}\right)^{-\beta}.$$

<div style="text-align: right">□</div>

Remark 6.2. The rate of convergence (6.58) can be improved to

$$|L(E) + L_N(E) - 2L_{2N}(E)| \lesssim e^{-cN^\eta} \quad \text{for all } N \geq N_{00}, \tag{6.61}$$

which follows from the proof of the inductive step, Lemma 6.2 (see (6.24)) and uses the avalanche principle. This faster rate of convergence leads to the sharper modulus of continuity (1.14) (see [16], [6] for details).

We will now explain how Anderson localization is derived from the large deviation Theorem 6.1.

Given the Schrödinger operator

$$[H(\underline{x}) \psi]_n := -\psi_{n+1} - \psi_{n-1} + \lambda\, v(\mathrm{T}^n \underline{x})\, \psi_n, \tag{6.62}$$

for every scale N we denote

$$H_N(\underline{x}) := R_{[1,N]} H(\underline{x}) R_{[1,N]},$$

where $R_{[1,N]}$ is the coordinate restriction to $[1, N] \subset \mathbb{Z}$ with Dirichlet boundary conditions.

Then the associated Green's functions are defined as

$$G_N(\underline{x}, E) := [H_N(\underline{x}) - E]^{-1},$$

if the $N \times N$ matrix $H_N(\underline{x}) - E$ is invertible.

The large deviation estimate (6.35) implies, via Cramer's rule, 'good bounds' on the Green's functions $G_N(\underline{x}, E)$ associated with (6.62).

Indeed, for $1 \le n_1 \le n_2 \le N$, we have:

$$G_N(\underline{x}, E)(n_1, n_2)$$
$$= [H_N(\underline{x}) - E]^{-1}(n_1, n_2)$$
$$= \frac{\det[H_{n_1-1}(\underline{x}) - E] \cdot \det[H_{N-n_2}(\mathrm{T}^{n_2}\underline{x}) - E]}{\det[H_N(\underline{x}) - E]}.$$

There is the following relation between transfer matrices and determinants:

$$M_N(\underline{x}, E) = \begin{bmatrix} \det[H_N(\underline{x}) - E] & -\det[H_{N-1}(\mathrm{T}\underline{x}) - E] \\ \det[H_{N-1}(\underline{x}) - E] & -\det[H_{N-2}(\mathrm{T}\underline{x}) - E] \end{bmatrix}. \qquad (6.63)$$

Therefore, we get the estimate on the Green's functions

$$|G_N(\underline{x}, E)(n_1, n_2)| \le \frac{\|M_{n_1}(\underline{x}, E)\| \cdot \|M_{N-n_2}(\mathrm{T}^{n_2}\underline{x}, E)\|}{|\det(H_N(\underline{x}) - E)|}.$$

Combining this with the LDT (6.35), we obtain the following bounds on the Green's functions $G_\Lambda(E, \underline{x})$ associated with the operator (6.62).

For every N large enough and for every energy E, there is a set $\Omega_N(E) \subset \mathbb{T}^2$ with $\mathrm{mes}[\Omega_N(E)] < e^{-N^\sigma}$ so that for any $\underline{x} \notin \Omega_N(E)$, one of the intervals

$$\Lambda = \Lambda(\underline{x}) = [1, N], [1, N-1], [2, N], [2, N-1]$$

will satisfy

$$|G_\Lambda(E, \underline{x})(n_1, n_2)| < e^{-c|n_1-n_2| + N^{1-}}. \qquad (6.64)$$

Since $v(\underline{x}) = \sum_{\underline{l} \in \mathbb{Z}^2} \hat{v}(\underline{l}) e^{2\pi i \underline{l} \cdot \underline{x}}$ and $|\hat{v}(\underline{l})| \le M e^{-\rho|\underline{l}|^{1/s}}$ for all $\underline{l} \in \mathbb{Z}^2$, substituting in (6.64) $v(\underline{x})$ by $v_1(\underline{x}) := \sum_{|\underline{l}| \le CN^s} \hat{v}(\underline{l}) e^{2\pi i \underline{l} \cdot \underline{x}}$ we can assume that the 'bad set' $\Omega_N(E)$ above not only has exponentially small measure, but it also has bounded algebraic complexity - it is semi-algebraic of degree $\le N^{d(s)}$.

These sets depend on the energy E. The rest of the proof of localization for (6.62) involves the elimination of the energy, which uses semi-algebraic set theory, and follows exactly the same pattern as the proof of the corresponding result for the analytic case (see [6], [5] or Chapter 15 in [4]).

Our statement for the skew-shift model is weaker than the one for the multi-frequency shift, since they both mirror the corresponding results in the analytic case.

Remark 6.3. We do not know if the transversality condition (1.12) is indeed necessary, either for the models considered here or for the single-variable shift considered in [16]. In particular, we do not know if the Lyapunov exponent is still positive throughout the spectrum for potential functions that have flat parts but are very smooth otherwise. This is a difficult and interesting problem.

Finally, a more challenging problem regarding Gevrey potential functions is proving localization for a long range model, one where the Laplacian is replaced by a Toeplitz matrix. In the case of the skew-shift dynamics, this could lead to applications to more general quantum kicked rotator equations.

References

[1] Kristian Bjerklöv, Positive Lyapunov exponent and minimality for a class of one-dimensional quasi-periodic Schrödinger equations. *Ergodic Theory Dynam. Systems* **25** (2005), 1015–1045. MR 2158395 Zbl 1094.47037

[2] J. Bourgain, Positive Lyapounov exponents for most energies. In V. D. Milman and G. Schechtman, *Geometric aspects of functional analysis.* Proceedings of the Israel seminar (GAFA) 1996–2000. Lecture Notes in Mathematics, 1745. Springer, Berlin etc., 2000, 37–66. MR 1796712 MR 1797971 (collection) Zbl 1160.81381 Zbl 0949.00025 (collection)

[3] J. Bourgain, On the spectrum of lattice Schrödinger operators with deterministic potential. II. Dedicated to the memory of T. Wolff. *J. Anal. Math.* **88** (2002), 221–254. MR 1984594 Zbl 1058.47002

[4] J. Bourgain, *Green's function estimates for lattice Schrödinger operators and applications.* Annals of Mathematics Studies, 158. Princeton University Press, Princeton, N.J., 2005. MR 2100420 Zbl 1137.35001

[5] J. Bourgain and M. Goldstein, On nonperturbative localization with quasi-periodic potential. *Ann. of Math.* (2) **152** (2000), 835–879. MR 1815703 Zbl 1053.39035

[6] J. Bourgain, M. Goldstein, and W. Schlag, Anderson localization for Schrödinger operators on $\mathbb{Z}$ with potentials given by the skew-shift. *Comm. Math. Phys.* **220** (2001), 583–621. MR 1843776 Zbl 0994.82044

[7] J. Chan, Method of variations of potential of quasi-periodic Schrödinger equations. *Geom. Funct. Anal.* **17** (2008), 1416–1478. MR 2377494 Zbl 1165.82014

[8] J. Chan, M. Goldstein, and W. Schlag, On non-perturbative anderson localization for C^α potentials generated by shifts and skew-shifts. Preprint 2006. arXiv:math/0607302 [math.DS]

[9] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon, *Schrödinger operators with application to quantum mechanics and global geometry.* Springer Study edition. Texts and Monographs in Physics. Springer, Berlin etc., 1987. MR 0883643 Zbl 0619.47005

[10] D. D'Acunto and K. Kurdyka, Explicit bounds for the Łojasiewicz exponent in the gradient inequality for polynomials. *Ann. Polon. Math.* **87** (2005), 51–61. MR 2208535 Zbl 1093.32011

[11] L. H. Eliasson, Discrete one-dimensional quasi-periodic Schrödinger operators with pure point spectrum. *Acta Math.* **179** (1997), 153–196. MR 1607554 Zbl 0908.34072

[12] M. Goldstein and W. Schlag, Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions. *Ann. of Math.* (2) **154** (2001), 155–203. MR 1847592 Zbl 0990.39014

[13] M. Goldstein and W. Schlag, Fine properties of the integrated density of states and a quantitative separation property of the Dirichlet eigenvalues. *Geom. Funct. Anal.* **18** (2008), 755–869. MR 2438997 Zbl 1171.82011

[14] M. Goldstein and W. Schlag, On resonances and the formation of gaps in the spectrum of quasi-periodic Schrödinger equations. *Ann. of Math.* (2) **173** (2011), 337–475. MR 2753606 Zbl 1268.82013

[15] Y. Katznelson, *An introduction to harmonic analysis.* John Wiley & Sons Inc., New York etc., 1968. MR 0248482 Zbl 0169.17902

[16] S. Klein, Anderson localization for the discrete one-dimensional quasi-periodic Schödinger operator with potential defined by a Gevrey-class function. *J. Funct. Anal.* **218** (2005), 255–292. MR 2108112 Zbl 1068.81023

[17] J. Kollár, An effective Łojasiewicz inequality for real polynomials. *Period. Math. Hungar.* **38** (1999), 213–221. MR 1756239 Zbl 0973.26012

[18] H. Krüger, Multiscale analysis for ergodic Schrödinger operators and positivity of Lyapunov exponents. *J. Anal. Math.* **115** (2011), 343–387. MR MR2855042

[19] H. Krüger, The spectrum of skew-shift Schrödinger operators contains intervals. *J. Funct. Anal.* **262** (2012), 773–810. MR 2863848 Zbl 1252.35211

[20] B. Ya. Levin, *Lectures on entire functions.* Translations of Mathematical Monographs, 150. In collaboration with and with a preface by Yu. Lyubarskii, M. Sodin, and V. Tkachenko, Translated from the Russian manuscript by Tkachenko. American Mathematical Society, Providence, R.I., 1996. MR 1400006 Zbl 0856.30001

[21] D. H. Phong, E. M. Stein, and J. A. Sturm, On the growth and stability of real-analytic functions. *Amer. J. Math.* **121** (1999), 519–554. MR 1738409 Zbl 1015.26031

Silvius Klein, Centro de Matemática e Aplicações Fundamentais,
Faculdade de Ciências da Universidade de Lisboa, Avenida Professor Gama Pinto, 2,
1649–003 Lisboa, Portugal

Department of Mathematical Sciences,
Norwegian University of Science and Technology (NTNU), Trondheim, N-7491,
Norway

Simion Stoilow Institute of Mathematics of the Romanian Academy (IMAR),
21 Calea Grivitei Street, 010702 Bucharest, Romania

e-mail: silviusaklein@gmail.com