

# CONVERGENCE OF PERTURBATION SERIES FOR UNBOUNDED MONOTONE QUASIPERIODIC OPERATORS

ILYA KACHKOVSKIY, LEONID PARNOVSKI, AND ROMAN SHTERENBERG

**ABSTRACT.** We consider a class of unbounded quasiperiodic Schrödinger-type operators on  $\ell^2(\mathbb{Z}^d)$  with monotone potentials (akin to the Maryland model) and show that the Rayleigh–Schrödinger perturbation series for these operators converges in the regime of small kinetic energies, uniformly in the spectrum. As a consequence, we obtain a new proof of Anderson localization in a more general than before class of such operators, with explicit convergent series expansions for eigenvalues and eigenvectors. This result can be restricted to an energy window if the potential is only locally monotone and one-to-one. A modification of this approach also allows the potential to be non-strictly monotone and have a flat segment, under additional restrictions on the frequencies.

## 1. INTRODUCTION

In this paper, we consider a class of quasiperiodic Schrödinger-type operators on  $\ell^2(\mathbb{Z}^d)$

$$(1.1) \quad (H(x_0)\psi)_{\mathbf{n}} = \varepsilon \sum_{\mathbf{m} \in \mathbb{Z}^d} \varphi_{\mathbf{n}-\mathbf{m}} \psi_{\mathbf{m}} + f(x_0 + \mathbf{n} \cdot \omega) \psi_{\mathbf{n}}.$$

The frequency vector  $\omega \in (-1/2, 1/2)^d$  will always satisfy the assumption that  $1, \omega_1, \dots, \omega_d$  are rationally independent. The parameter  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$  is the quasiperiodic phase. We will always assume  $\varphi \in \ell^1(\mathbb{Z}^d)$ , so that the first term in the right hand side (usually referred to as hopping or kinetic term) is a bounded operator on  $\ell^2(\mathbb{Z}^d)$ . Most of the time we will assume that  $\varphi_{\mathbf{n}} \neq 0$  only for finitely many  $\mathbf{n} \in \mathbb{Z}^d$ . The most common example is the discrete nearest-neighbor Laplace operator:

$$(1.2) \quad \varphi = \sum_{\mathbf{n}: |\mathbf{n}|_1=1} e_{\mathbf{n}},$$

where  $\{e_{\mathbf{n}} : \mathbf{n} \in \mathbb{Z}^d\}$  is the standard basis in  $\ell^2(\mathbb{Z}^d)$  and  $|\cdot|_1$  denotes the  $\ell^1$ -norm.

The real-valued function  $f$  is initially defined on  $(-1/2, 1/2)$ ; we assume it to be continuous and satisfy

$$f(-1/2 + 0) = -\infty, \quad f(1/2 - 0) = +\infty.$$

We then extend  $f$  periodically into  $\mathbb{R} \setminus (\mathbb{Z} + 1/2)$ . The assumptions on  $x_0$  and  $\omega$  imply that (1.1) is a well-defined unbounded self-adjoint operator on  $\ell^2(\mathbb{Z}^d)$ . In the present work, the most important assumption will be that  $f$  is non-decreasing on  $(-1/2, 1/2)$ . In most of the paper, we assume that  $f$  is strictly increasing, with the derivative bounded from below, but in the last section  $f$  will be allowed to have flat pieces.

Spectral theory of operators (1.1) with strictly monotone functions  $f$  dates back to the classical Maryland model [5, 6, 4], where  $f(x) = \tan(\pi x)$  and the kinetic term is the discrete Laplacian (1.2). Similarly to many other quasiperiodic models, the spectral types of such

operators depend on the arithmetic properties of the frequencies. We will call the frequency vector  $\omega$  *Diophantine* if, for some  $C_{\text{dio}} > 0$  and  $\tau > d + 1$  we have

$$(1.3) \quad \|\mathbf{n} \cdot \omega\| := \text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z}) \geq C_{\text{dio}} |\mathbf{n}|^{-\tau}, \quad \forall \mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}.$$

It is known that for all  $\varepsilon > 0$ , all Diophantine  $\omega$ , and all values of  $x_0$ , the Maryland model has Anderson localization, that is, dense pure point spectrum and exponentially decaying eigenfunctions. In  $d = 1$ , a complete description of the spectral type (in particular, the transitions between pure point and singular continuous spectra depending on whether the frequency is closer to a Diophantine or a Liouville irrational number) was obtained in [7]. See also [9] for an alternative proof in the Diophantine setting. The case of more general  $f$  was studied in [2], and it was shown that the operator has Anderson localization for Diophantine frequencies and  $0 < \varepsilon < \varepsilon_0(\omega, f, \varphi)$ . The function  $f$  was assumed to be strictly monotone and to have a meromorphic continuation into a strip in  $\mathbb{C}$ . The approach of [2] is based on a KAM-type scheme: a diagonalization of  $H(x_0)$  is constructed via an infinite sequence of unitary transformations. In a recent work [11], Anderson localization for  $d = 1$  is shown for all  $\varepsilon > 0$  and all Diophantine frequencies, under the assumption that  $\log |f| \in L^1(-1/2, 1/2)$ . However, the non-perturbative method of [11], based on [8], cannot be extended to  $d > 1$ . We also mention [10], where singular continuous spectrum of operators with meromorphic potentials is studied.

In the present paper, we study Anderson localization for (1.1) in the same perturbative setting as [2], that is, for  $|\varepsilon| < \varepsilon_0(\omega, f, \varphi)$ . Our conditions on  $f$  include all meromorphic functions from [2], but are formulated completely in terms of the first derivative of  $f$ . We believe that the most interesting aspect of our paper is the method: we construct explicit series for eigenvalues and eigenfunctions using the standard perturbation theory, and, to our surprise, in both cases *the Rayleigh-Schrödinger perturbation series converge* as is, without the need of multiple KAM-type steps. As a consequence, we are able to write down complete representations of eigenvalues and eigenvectors in terms of converging power series in  $\varepsilon$ . Compared to the previous work, our method also covers two new cases (the summary of the main results is provided in the next section):

- The case when  $f$  is monotone and one-to-one on an interval  $(a, b) \subset (-1/2, 1/2)$ , with some regularity properties outside of  $(a, b)$  (including  $f^{-1}(f(a, b)) = (a, b)$ ), but does not have to be monotone outside of  $(a, b)$ . In this case, we obtain localization on the energy interval slightly smaller than  $f(a, b)$ , see Theorem 5.19.
- The case when  $f$  has a flat segment, but is Lipschitz monotone outside of that segment. In this case, under some additional assumptions on the length of the segment and the frequencies, we can obtain complete Anderson localization, see Theorem 7.4.

To our best knowledge, this is the first class of examples of convergent perturbation series in the context of Schrödinger operators with dense eigenvalues. A related phenomenon for the classical KAM was observed in [3] in the context of ODEs (which does not appear to be related to monotonicity). Our proof is based on careful observation of cancellations between terms of the perturbation series with the same power of  $\varepsilon$  (see Remarks 4.1, 4.7, 4.8 for the outline of the cancellation mechanism). While the convergence results use the structure (1.1), the actual combinatorial procedure of grouping terms in order to prepare them for cancelling can be formulated in a more abstract context of arbitrary lattice Schrödinger-type operators. We believe that this procedure may be of independent interest. For example, for bounded monotone discontinuous potentials such as  $f(x) = \{x\}$  the perturbation series will converge

on a large (but not full measure) subset of phases  $x_0$  and thus produce a large number of exponentially decaying eigenfunctions, whose contribution to the integrated density of states will approach 1 as  $\varepsilon \rightarrow 0$ . However, the remaining “resonant” phases will require further analysis, possibly by different methods. We intend to explore appropriate modifications of our method in order to study localization and spectral gaps of these operators in subsequent publications. One can also draw some parallels between our procedure and renormalization techniques used in statistical physics, such as the work [13] where Anderson localization in terms of two-particle functions was studied for the quasiperiodic Holstein model.

**Acknowledgments.** The authors are grateful to David Damanik and Leonid Pastur for useful discussions, and to the anonymous referee whose suggestions, hopefully, made the text substantially more readable. The research of LP was partially supported by EPSRC grants EP/J016829/1 and EP/P024793/1. RS was partially supported by NSF grant DMS–1814664. IK was partially supported by NSF grant DMS–1846114.

## 2. RAYLEIGH–SCHRÖDINGER PERTURBATION SERIES AND MAIN RESULTS

**2.1. The perturbation series.** Let  $V$  be a self-adjoint multiplication operator on  $\ell^2(\mathbb{Z}^d)$ , not necessarily bounded:

$$(2.1) \quad (Vu)_{\mathbf{n}} = V_{\mathbf{n}}u_{\mathbf{n}}, \quad u = \sum_{\mathbf{n} \in \mathbb{Z}^d} u_{\mathbf{n}}e_{\mathbf{n}},$$

where  $\{e_{\mathbf{n}} : \mathbf{n} \in \mathbb{Z}^d\}$  is the standard basis in  $\ell^2(\mathbb{Z}^d)$ . Let also  $\Phi$  be a Töplitz-type operator: for a sequence  $\{\varphi_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{Z}^d}$ , we define

$$(2.2) \quad (\Phi u)_{\mathbf{n}} = \sum_{\mathbf{m} \in \mathbb{Z}^d} \varphi_{\mathbf{n}-\mathbf{m}}u_{\mathbf{m}}.$$

We will assume that  $\Phi$  is self-adjoint:

$$(2.3) \quad \Phi_{\mathbf{m}\mathbf{n}} = \overline{\Phi_{\mathbf{n}\mathbf{m}}}, \quad \text{that is, } \varphi_{-\mathbf{n}} = \overline{\varphi_{\mathbf{n}}}.$$

In Sections 2–5, we also require that  $\varphi_{\mathbf{n}} \neq 0$  only for *finitely many* lattice points  $\mathbf{n} \in \mathbb{Z}^d$ , and  $\varphi_{\mathbf{0}} = 0$ . In Section 6, this will be generalized to an infinite range case in the quasiperiodic setting. Define a family of operators

$$(2.4) \quad H = V + \varepsilon\Phi,$$

parametrized by  $\varepsilon > 0$ . We will not emphasize the dependence on  $\varepsilon$  in the expression for  $H$ , as long as it is clear from the context. Fix some  $\mathbf{n}_0 \in \mathbb{Z}^d$  and assume that

$$(2.5) \quad V_{\mathbf{n}} \neq V_{\mathbf{n}_0} \text{ for } \mathbf{n} \neq \mathbf{n}_0.$$

In other words, the potential attains the value  $V_{\mathbf{n}_0}$  only at the lattice point  $\mathbf{n}_0$ . We will call this the *algebraic non-resonant condition at  $\mathbf{n}_0$* .

If  $\varepsilon = 0$ , then the operator  $H$  has an eigenvalue  $V_{\mathbf{n}_0}$  with an eigenvector  $e_{\mathbf{n}_0}$ . *Rayleigh–Schrödinger series* is a formal perturbation-theoretic expansion, which represents how that eigenvalue changes after the perturbation with  $\varepsilon > 0$ . In order to construct these series, consider the eigenvalue equation

$$(2.6) \quad H\psi = \lambda\psi,$$

where both  $\lambda$  and  $\psi$  are formal series in  $\varepsilon$ :

$$(2.7) \quad \lambda = \lambda_0 + \varepsilon\lambda_1 + \varepsilon^2\lambda_2 + \dots,$$

$$(2.8) \quad \psi = \psi_0 + \varepsilon\psi_1 + \varepsilon^2\psi_2 + \dots$$

with  $\lambda_0 = V_{\mathbf{n}_0}$  being the unperturbed eigenvalue and  $\psi_0 = e_{\mathbf{n}_0}$  the unperturbed eigenfunction. Then, one considers (2.6) as an equality of formal series

$$(2.9) \quad (V + \varepsilon\Phi)(\psi_0 + \varepsilon\psi_1 + \dots) = (\lambda_0 + \varepsilon\lambda_1 + \dots)(\psi_0 + \varepsilon\psi_1 + \dots);$$

in other words, an infinite system of equations obtained by comparing the coefficients at each power of  $\varepsilon$  (note that each equation will have finitely many terms). We will impose the following normalization conditions:

$$(2.10) \quad \psi_j \perp \psi_0, \quad j > 0.$$

In other words, all the correction terms are orthogonal to  $e_{\mathbf{n}_0}$ .

It is well known and easy to see that, under the above assumptions on  $V$  and  $\Phi$ , both  $\lambda_j$  and  $\psi_j$  can be uniquely determined from equating the coefficients in (2.9) at each power of  $\varepsilon$ ; see also Theorem 3.2 below. One can start these expansions at any lattice site  $\mathbf{n}_0$  as long as it satisfies the algebraic non-resonance condition.

**2.2. Summary of the main results.** The results of the present paper rely on certain regularity conditions on the function  $f$ , which can be stated in terms of the first derivative of  $f$  and are discussed in detail in Section 5. Any meromorphic strictly monotone function  $f$ , such as in [2], satisfies these conditions, and one can state the following theorem in a self-contained form as a consequence of our results.

**Theorem 2.1.** *Let  $f: (-1/2, 1/2) \rightarrow \mathbb{R}$  be a strictly increasing function satisfying  $f(-1/2 + 0) = -\infty$ ,  $f(1/2 - 0) = +\infty$  and admitting a meromorphic 1-periodic extension to a neighborhood of the real line in  $\mathbb{C}$ . Suppose that the frequency vector  $\omega$  satisfies the Diophantine condition (1.3)*

$$\|\mathbf{n} \cdot \omega\| = \text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z}) \geq C_{\text{dio}} |\mathbf{n}|^{-\tau}, \quad \forall \mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}.$$

Let  $H(x_0)$  be the operator (1.1) with  $\varphi_{\mathbf{n}} \neq 0$  for finitely many points  $\mathbf{n} \in \mathbb{Z}^d$ . There exists  $\varepsilon_0 = \varepsilon_0(d, f, \varphi, C_{\text{dio}}, \tau) > 0$  such that, for any  $\varepsilon \in [0, \varepsilon_0]$ , any  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ , and any  $\mathbf{n}_0 \in \mathbb{Z}^d$ , the Rayleigh–Schrödinger perturbation series for the eigenvalues and eigenvectors of  $H(x_0)$ , started at  $\mathbf{n}_0$ , converge and provide a complete orthogonal system of eigenvectors for  $H(x_0)$ , parametrized by  $\mathbf{n}_0 \in \mathbb{Z}^d$ .

Note that the strict monotonicity of  $f$  on  $(-1/2, 1/2)$  and rational independence of the components of  $\omega$  (which follows from the Diophantine condition) imply that the potential in  $H(x_0)$  satisfies the algebraic non-resonance condition at all lattice points for all  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ . In other words,  $f(x_0 + \mathbf{m} \cdot \omega) \neq f(x_0 + \mathbf{n} \cdot \omega)$  for  $\mathbf{m} \neq \mathbf{n}$ .

For the remaining results summarized below, it is more convenient to postpone the exact statements until further sections.

- (1) The main result, from which Theorem 2.1 follows, is Theorem 5.11 whose proof is completed in Section 5. The conditions on  $f$ , in both local and global forms, are described in Section 5.1. These conditions, referred to as  $C_{\text{reg}}$ -regularity, can be described entirely in terms of the first derivatives of  $f$ .
- (2) Several generalizations are discussed in Section 6. In particular, the hopping terms  $\varphi_{\mathbf{n}}$  can be allowed to depend on  $x$  in a quasiperiodic way, and infinite range hopping is allowed, with some natural assumptions that higher range hopping terms have larger powers of  $\varepsilon$  in front of them. A somewhat technical case is also considered where  $f$ ,

while still strictly monotone, is allowed to have intervals on which the lower bound on the derivative is, essentially,  $\varepsilon$ -dependent. This case serves as a preparation for the next section. A version of the construction from Section 6 is also used in our subsequent paper [12].

(3) In the last section, we consider a model example of  $f$  which is not necessarily strictly monotone on  $(-1/2, 1/2)$  and is allowed to have a flat segment. As a consequence, the algebraic non-resonance condition is violated for this operator. We describe a scheme that allows to transform such operators to those satisfying the assumptions of a general theorem from Section 6. This allows us to obtain Anderson localization for these operators.

### 3. GRAPHICAL AND SYMBOLIC REPRESENTATIONS OF THE PERTURBATION SERIES

In this section, we will discuss in detail the construction of the perturbation series. For simplicity of notation, we will always assume that the series is started at  $\mathbf{n}_0 = 0$ . The general case can be easily considered by translation (or, in the case of quasiperiodic operators, by changing the phase  $x_0$ ).

**3.1. Representation in terms of paths on a graph.** There are multiple ways of representing terms in the series (2.9). For our purposes, the most convenient way would be to associate them with paths on a certain weighted graph  $\Gamma$ . *The vertices of  $\Gamma$  are defined as follows.* First, take a copy of  $\mathbb{Z}^d$  and call it the sheet of height 0 (“base sheet”). For each vertex  $\mathbf{k}_0$  of  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$ , we add a new separate copy of  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$ , with our vertex  $\mathbf{k}_0$  placed as the origin (so that  $\mathbf{k}_0$  together with a copy of  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$  form a copy of  $\mathbb{Z}^d$ ). Each of these copies of  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$  will be called sheets of height 1. Now, add a separate copy of  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$  for each vertex  $\mathbf{k}_1$  on each sheet of height 1, and call these new copies sheets of height 2. The result of this process, repeated indefinitely, will be the set of vertices of the graph  $\Gamma$ . One can also enumerate the vertices of the graph in a more direct way: each non-zero vertex of  $\Gamma$  on a sheet of height  $s$  can be associated with a sequence  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_s$ , where  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_s \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$ . If  $v$  is a vertex of  $\Gamma$  that is the point  $\mathbf{k}$  from one of the copies of  $\mathbb{Z}^d$  or  $\mathbb{Z}^d \setminus \{\mathbf{0}\}$ , we will call  $\mathbf{k}$  the *coordinate* of  $v$ . In other words, a vertex  $v$  represented by a sequence  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_s$  has coordinate  $\mathbf{k}_s$ . We will also consider the origin (which only exists on the base sheet) to be the only vertex with coordinate  $\mathbf{0}$  and denote it by the same symbol  $\mathbf{0}$  as the origin in  $\mathbb{Z}^d$ .

We will say that the vertex associated with a sequence  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_{s-1}, \mathbf{k}_s$  is *directly above* the vertex associated to  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_{s-1}$ . We will also say that the vertex associated with  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_s, \dots, \mathbf{k}_t$  is above the vertex associated with  $\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_s$ . In the remaining text, we will avoid using the sequence notation for vertices, but we will often use the terms “above” or “directly above”. We will also use the words “the sheet of  $\Gamma$  directly above the vertex  $v$ ” in the obvious interpretation. Sometimes it is also convenient to use words “below” or “directly below”, whose meaning is opposite to “above” and “directly above”. Note that, by construction, none of the vertices of  $\Gamma$  are above or below  $\mathbf{0}$ .

*The edges of the graph  $\Gamma$  are defined as follows.* In all cases,  $\mathbf{n}$  and  $\mathbf{n}'$  are coordinates of two vertices  $v, v'$ . The graph  $\Gamma$  is oriented; however, with each edge it will also contain the edge in the opposite direction (which may have a different weight — see below). If  $v$  and  $v'$  are on the same sheet, there is an edge between  $v$  and  $v'$  if and only if  $\Phi_{\mathbf{n}'\mathbf{n}} \neq 0$ . That includes the origin on the base sheet (recall that other sheets do not have an origin on them). In addition, if  $v'$  is located on the sheet directly above  $v$ , then there is an edge from  $v$  to  $v'$ .

if and only if  $\Phi_{\mathbf{n}'\mathbf{0}} \neq 0$ . In a way,  $v$  plays the role of the origin in the sheet that is directly above  $v$ .

*To each edge of the graph  $\Gamma$ , we will associate a weight:*

- If  $v$  and  $v'$  are on the same sheet, then the edge from  $v$  to  $v'$  will have weight  $\frac{\Phi_{\mathbf{n}'\mathbf{n}}}{V_0 - V_{\mathbf{n}'}}$  if  $\mathbf{n}' \neq \mathbf{0}$ , and weight  $\Phi_{\mathbf{n}\mathbf{0}}$  if  $\mathbf{n}' = \mathbf{0}$ ; the latter can happen only on the base sheet.
- If  $v'$  is directly above  $v$ , then the “ascending” edge from  $v$  to  $v'$  has weight  $\frac{\Phi_{\mathbf{n}'\mathbf{0}}}{V_0 - V_{\mathbf{n}'}}$ , and the “descending” edge from  $v'$  to  $v$  has weight  $-\frac{\Phi_{\mathbf{0}\mathbf{n}'}}{V_0 - V_{\mathbf{n}}}$ .

Just like on any oriented graph, one can define a path on  $\Gamma$  as a sequence of (not necessarily distinct) edges  $(e_1, \dots, e_p)$ , where the edge  $e_j$  ends at the starting vertex of  $e_{j+1}$ . We will say that this path starts from the starting vertex of  $e_1$  and ends at the ending vertex of  $e_p$ . We will say that a path visits a vertex  $v$  if one of the edges  $e_1, \dots, e_p$  starts or ends with  $v$ . A path may visit the same vertex multiple times.

We will consider two types of paths on  $\Gamma$ . An *eigenvalue path* is a path that starts and ends at the origin and never visits the origin in between. An *eigenvector path* starts at the origin, ends at some non-zero vertex with coordinate  $\mathbf{n}$  on the base sheet, and does not visit the origin in between. It will also be convenient to consider eigenvalue paths that do not start from the origin. By definition, a *non-base (eigenvalue) path* is a path on  $\Gamma$  that starts and ends from the same vertex  $v$  and only visits vertices above  $v$  in between. An eigenvalue path would be a non-base path that starts from the origin. Since the graph  $\Gamma$  above any point looks identical, one can translate any eigenvalue path into any vertex of  $\Gamma$  and obtain a non-base path. We will not consider non-base eigenvector paths.

Since we will mostly consider eigenvalue/eigenvector paths, *any claim about paths, by default, is applicable only to eigenvalue/eigenvector paths*. We will always specify if we are talking specifically about non-base paths, eigenvalue paths, or eigenvector paths.

In all cases,  $|\mathcal{P}|$  will denote the length of  $\mathcal{P}$  (in other words, the number of edges of  $\mathcal{P}$ ). By  $\text{Cont}(\mathcal{P})$ , we denote the product of weights of all edges of  $\mathcal{P}$ . If a path travels along the same edge multiple times, each of them gives a separate contribution.

If  $\mathcal{Q}$  is an eigenvalue path and  $\mathcal{Q}'$  is a non-base path obtained from  $\mathcal{Q}$  by translating it into a vertex with coordinate  $\mathbf{k}$ , then we have

$$(3.1) \quad \text{Cont}(\mathcal{Q}') = -(V_0 - V_{\mathbf{k}})^{-1} \text{Cont}(\mathcal{Q}),$$

since the weight of the last descending edge in  $\mathcal{Q}'$  has an extra factor  $-(V_0 - V_{\mathbf{k}})^{-1}$ .

**Remark 3.1.** In the above construction,  $\Gamma_j$  has an edge between two points with coordinates  $\mathbf{m}, \mathbf{n}$  on the same sheet if and only if  $\Phi_{\mathbf{nm}} \neq 0$ . In some constructions, it is convenient to assume that has edges between *any* two points on the same sheet, with weights defined in the same way as above. At this stage, it is just a formality since newly added edges will have zero weights and no contribution to any calculations. Later, when one considers  $x$ -dependent hopping terms, it will be convenient that the  $x$ -dependence is only present in the edge weights but not in the graph  $\Gamma$ .

The following is a formulation of the Rayleigh–Schrödinger perturbation theory. See, for example, [1] for a similar result in a different notation and slightly different setting.

**Theorem 3.2.** *Define  $V$  and  $\Phi$  as in (2.1), (2.2), and assume that the algebraic non-resonance condition (2.5) is satisfied at  $\mathbf{n}_0 = \mathbf{0}$ . Define  $\Gamma$  and eigenvalue/eigenvector paths*

on  $\Gamma$  as above. Then the infinite system of equations

$$(3.2) \quad (V + \varepsilon\Phi)(\psi_0 + \varepsilon\psi_1 + \dots) = (\lambda_0 + \varepsilon\lambda_1 + \dots)(\psi_0 + \varepsilon\psi_1 + \dots),$$

treated formally by equating left and right hand side at each power of  $\varepsilon$ , with

$$(3.3) \quad \lambda_s \in \mathbb{R}, \quad \psi_s \in \ell^2(\mathbb{Z}^d), \quad \lambda_0 = V_0, \quad (\psi_s)_0 = 0, \quad s > 0; \quad \psi_0 = e_0,$$

has a unique solution given by

$$(3.4) \quad \lambda_s = \sum_{\mathcal{P}: |\mathcal{P}|=s} \text{Cont}(\mathcal{P}),$$

where the sum is considered over all eigenvalue paths  $\mathcal{P}$  on  $\Gamma$  (with  $|\mathcal{P}| = s$ ), and

$$(3.5) \quad (\psi_s)_k = \sum_{\mathcal{P}: |\mathcal{P}|=s} \text{Cont}(\mathcal{P}), \quad k \neq 0, \quad s > 0,$$

where the sum is considered over all eigenvector paths  $\mathcal{P}$  (again, with  $|\mathcal{P}| = s$ ) between vertices with coordinates  $\mathbf{0}$  and  $\mathbf{k}$  on the base sheet of  $\Gamma$ .

*Proof.* The first equation obtained from (3.2) by considering terms with  $\varepsilon^0$ , is already contained in (3.3). The next equation (corresponding to  $\varepsilon^1$ ) leads to

$$(3.6) \quad V\psi_1 + \Phi e_0 = V_0\psi_1 + \lambda_1\psi_0.$$

We will solve this equation by projecting the left and right hand sides onto  $\text{span}\{e_0\}$  and  $\text{span}\{e_0\}^\perp$ . It will be convenient to use the ‘‘partial inverse’’ to the non-invertible operator  $V_0I - V$ :

$$(V_0 - V)^{-1} := (V_0I - V)^{-1}(1 - \langle e_0, \cdot \rangle e_0),$$

that is, we always assume that  $(V_0 - V)^{-1}$  is extended by zero into  $\ker(V_0I - V)$ . The operator  $(V_0 - V)^{-1}$ , as well as  $V$  itself, is a possibly unbounded multiplication operator. However, we will only apply  $(V_0 - V)^{-1}$  to vectors with finite support, which are always in its domain. With these conventions, (3.6) reduces to

$$\lambda_1 = 0, \quad \psi_1 = (V_0 - V)^{-1}\Phi e_0.$$

The comparison of terms of (3.2) at  $\varepsilon^2$  yields

$$(3.7) \quad V\psi_2 + \Phi\psi_1 = V_0\psi_2 + \lambda_1\psi_1 + \lambda_2e_0.$$

Projecting (3.7) onto  $\text{span}\{e_0\}$  and  $\text{span}\{e_0\}^\perp$  together with the orthogonality condition  $\psi_j \perp e_0 = 0$  for  $j > 0$  lead to

$$\lambda_2 = \langle \Phi\psi_1, e_0 \rangle,$$

$$\psi_2 = (V_0 - V)^{-1}\Phi\psi_1 - \lambda_1(V_0 - V)^{-1}\psi_1 = (V_0 - V)^{-1}\Phi\psi_1$$

In general, the equation at  $\varepsilon^s$

$$V\psi_s + \Phi\psi_{s-1} = V_0\psi_s + \lambda_1\psi_{s-1} + \lambda_2\psi_{s-2} + \dots + \lambda_{s-1}\psi_1 + \lambda_s e_0,$$

after projecting it onto the same subspaces, becomes a system of two equations

$$(3.8) \quad \psi_s = (V_0 - V)^{-1}(\Phi\psi_{s-1} - \lambda_2\psi_{s-2} - \dots - \lambda_{s-1}\psi_1)$$

and

$$(3.9) \quad \lambda_s = \langle \Phi\psi_{s-1}, e_0 \rangle = \langle \Phi(V_0 - V)^{-1}\Phi\psi_{s-2}, e_0 \rangle - \lambda_2 \langle \Phi(V_0 - V)^{-1}\psi_{s-3}, e_0 \rangle \\ - \lambda_3 \langle \Phi(V_0 - V)^{-1}\psi_{s-4}, e_0 \rangle - \dots - \lambda_{s-2} \langle \Phi(V_0 - V)^{-1}\psi_1, e_0 \rangle.$$

One can easily check that, say, for  $s = 1$  and  $s = 2$  this gives the same expressions as (3.4), (3.5). It remains to show that (3.4), (3.5) satisfy recurrent relations similar to (3.9), (3.8).

Consider the first equality in (3.9). In terms of paths, it corresponds to the following fact: each eigenvalue path ends at the origin, and the origin is only connected to the base sheet. Therefore, each eigenvalue path of length  $s$  can be obtained from an eigenvector path of length  $s - 1$  that ends at a vertex with coordinate  $\mathbf{k}$  on the base sheet by adding an extra edge from  $\mathbf{k}$  to  $\mathbf{0}$ . By the induction assumption (3.5) for  $s - 1$ , the total contribution of such eigenvector paths is  $(\psi_{s-1})_{\mathbf{k}}$ . The extra edge from  $\mathbf{k}$  to  $\mathbf{0}$  will produce a factor  $\Phi_{\mathbf{0k}}$ , which results in

$$\lambda_s = \sum_{\mathbf{k} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} \Phi_{\mathbf{0k}} (\psi_{s-1})_{\mathbf{k}},$$

thus arriving to the first equality in (3.9).

To establish (3.8), suppose that  $\mathcal{P}$  is an eigenvector path that ends at  $\mathbf{k}$  on the base sheet. Then there are two possibilities: either it arrives to  $\mathbf{k}$  from some other vertex  $\mathbf{m}$  on the base sheet, or it descends into  $\mathbf{k}$  from the sheet directly above  $\mathbf{k}$ .

In the first case,  $\mathcal{P}$  is obtained from some other eigenvector path  $\mathcal{P}'$  from  $\mathbf{0}$  to  $\mathbf{m}$  of length  $s - 1$ , by adding an edge from  $\mathbf{m}$  to  $\mathbf{k}$  on that sheet. By the induction assumption (3.5) with  $s - 1$ , the total contribution of such paths from  $\mathbf{0}$  to  $\mathbf{m}$  is  $(\psi_{s-1})_{\mathbf{m}}$ . The weight of the extra edge from  $\mathbf{m}$  to  $\mathbf{k}$  is  $\Phi_{\mathbf{km}}(V_0 - V_k)^{-1}$ . After summing over all possible choices of  $\mathbf{m}$ , we arrive to the total contribution of the paths from the first case being

$$\sum_{\mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}} (V_0 - V_k)^{-1} \Phi_{\mathbf{km}} (\psi_{s-1})_{\mathbf{m}},$$

which is equal to the  $\mathbf{k}$ th component of the first term in (3.8).

In the second case note that, in order to be able to descend into the vertex  $\mathbf{k}$  on the base sheet, the path  $\mathcal{P}$  should have, sometime earlier, ascended from the same vertex to the sheet above it. Consider the last time when  $\mathcal{P}$  ascended from  $\mathbf{k}$  on the base sheet and break  $\mathcal{P}$  into two parts: the part  $\mathcal{P}'$  right before that event, and the part  $\mathcal{Q}'$  after it. The part  $\mathcal{Q}'$  is a non-base eigenvalue path that starts and ends at  $\mathbf{k}$  and spends the rest of time in the sheets above  $\mathbf{k}$ . Its contribution can be calculated using (3.1), thus obtaining

$$\text{Cont}(\mathcal{P}) = \text{Cont}(\mathcal{Q}') \text{Cont}(\mathcal{P}') = -(V_0 - V_k)^{-1} \text{Cont}(\mathcal{Q}) \text{Cont}(\mathcal{P}'),$$

where  $\mathcal{Q}$  is the eigenvalue path obtained by translating  $\mathcal{Q}'$  into the origin. By considering all possible choices for  $\mathcal{Q}$ , we arrive to the total contribution of paths in this case being

$$- \sum_{\mathcal{Q}, \mathcal{P}'} (V_0 - V_k)^{-1} \text{Cont}(\mathcal{Q}) \text{Cont}(\mathcal{P}'),$$

where the summation is considered over all eigenvalue paths  $\mathcal{Q}$  with  $|\mathcal{Q}| = j$  and all eigenvector paths  $\mathcal{P}'$  from  $\mathbf{0}$  to  $\mathbf{k}$  of length  $s - j$ , with  $j = 2, 3, \dots, s - 1$ . For a fixed  $j$ , one can factorize the summation and apply the induction assumptions (3.4), (3.5) at stages  $j$  and  $s - j$ , respectively, and obtain that, for each  $j$ , the contribution of paths  $\mathcal{P} = \mathcal{Q}' \mathcal{P}'$  with  $|\mathcal{Q}'| = j$  is equal to

$$-(V_0 - V_k)^{-1} \left( \sum_{\mathcal{Q}: |\mathcal{Q}|=j} \text{Cont}(\mathcal{Q}) \right) \left( \sum_{\mathcal{P}': |\mathcal{P}'|=s-j} \text{Cont}(\mathcal{P}') \right) = -(V_0 - V_k)^{-1} \lambda_j (\psi_{s-j})_{\mathbf{k}},$$

which can now be identified with a term of (3.8).  $\square$

**3.2. Symbolic and graphical representation of paths on  $\Gamma$ .** We will use the following notation in order to describe paths on  $\Gamma$ . Let  $\mathcal{P}$  be an eigenvalue path which only travels along the base sheet of  $\Gamma$  (later, such paths will be called *loops*). Suppose that it starts at the origin, then visits the vertices of the base sheet with coordinates  $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_k \neq \mathbf{0}$  (in this order, and each new visit of the same vertex is accounted for separately), and then returns to the origin. We will use the following notation for this path:

$$\mathcal{P} = (\mathbf{n}_1 \mathbf{n}_2 \dots \mathbf{n}_k)$$

We will also denote eigenvector paths in the same way, but will include the last point in the notation:

$$\mathcal{P} = (\mathbf{n}_1 \mathbf{n}_2 \dots \mathbf{n}_k \mathbf{m}),$$

where  $\mathbf{m} \notin \{\mathbf{n}_k, \mathbf{0}\}$  is the ending point of the path  $\mathcal{P}$ . Whether the string denotes an eigenvalue or an eigenvector path, should be specified in the context, otherwise the notation is ambiguous: for example, the string (1234321) denotes both an eigenvalue path and an eigenvector path that starts at 0 and ends at 1 (in this example, we assumed that all coordinates belong to  $\mathbb{Z}$ , that is,  $d = 1$ , and are not using boldface notation for them).

Whenever a path ascends to a sheet above some vertex, we will write an opening parenthesis, and then continue with coordinates of vertices that  $\mathcal{P}$  visits on the upper sheet. For example, the eigenvalue path

$$(3.10) \quad \mathcal{P} = (12345(12321)5434321)$$

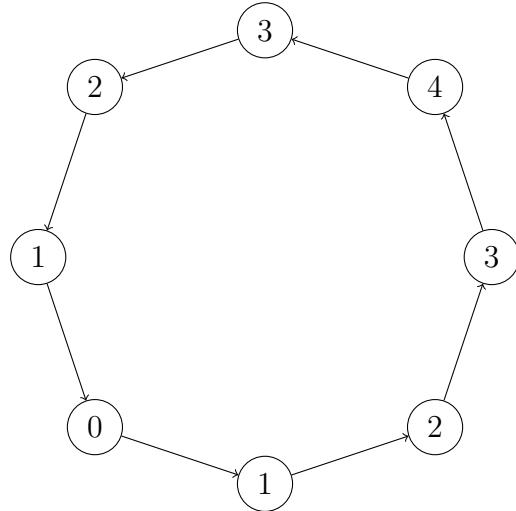
travels from the origin to the vertex with coordinate 5 on the base sheet (here we again assume that all coordinates belong to  $\mathbb{Z}$ , that is,  $d = 1$ ), then uses an ascending edge to ascend the sheet of height 1 directly above the vertex 5, then travels to 3 and back to 1 on that sheet, and then descends to the base sheet and proceeds along the remaining segment 434321. Whenever the path uses a descending edge to descend to a lower sheet, we use the closing parenthesis. One can have multiple levels of parentheses, depending on how high the path climbs. Given the string representing  $\mathcal{P}$ , one can calculate  $\text{Cont}(\mathcal{P})$  using the following rules, which will associate an edge to each pair of consecutive coordinates in that string. Suppose  $v$  and  $v'$  are vertices with coordinates  $\mathbf{n}, \mathbf{n}'$  respectively.

- $\mathbf{n}\mathbf{n}'$  represents the edge from  $v_1$  to  $v_2$  assuming that they are on the same sheet.
- $\mathbf{n}(\mathbf{n}'$  represents the ascending edge from  $v_1$  to  $v_2$  assuming that  $v_2$  is on a sheet directly above  $v_1$ .
- $\mathbf{n})\mathbf{n}'$  represents the descending edge from  $v_1$  to  $v_2$  assuming that  $v_1$  is on a sheet directly above  $v_2$ .
- $(\mathbf{n}$  in the beginning of the string represents the edge from  $\mathbf{0}$  to  $\mathbf{n}$ .
- if  $\mathcal{P}$  is an eigenvalue path, then  $\mathbf{n})$  in the end of the string represents the edge from  $v$  to  $\mathbf{0}$ . In case of an eigenvector path, it does not represent anything (one can consider it as an edge of weight 1 and length zero, meaning that it does not contribute to  $|\mathcal{P}|$  and  $\text{Cont}(\mathcal{P})$ ).

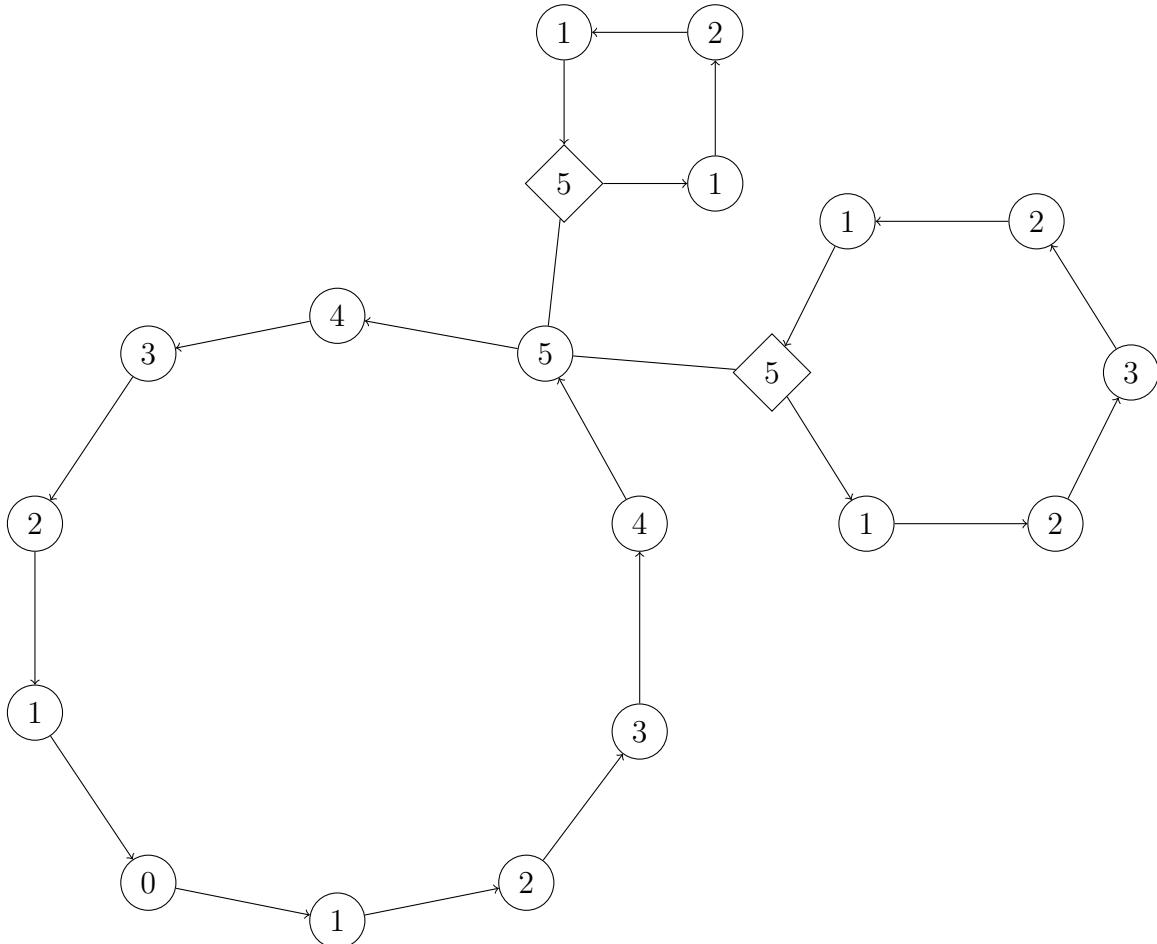
To calculate  $\text{Cont}(\mathcal{P})$ , one can multiply the weights of the edges for each pair of consecutive lattice points in the string, using the rules from the above paragraph and weights from Section 3.1.

In order to better understand the constructions, we will also use graphical representations of the paths on  $\Gamma$ . A simple path (1234321) will be represented graphically as the following

diagram:



A path visiting higher sheets of  $\Gamma$ , such as  $(12345(12321)5(121)54321)$ , corresponds to the diagram

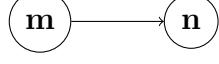
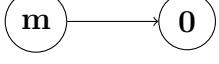
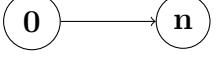
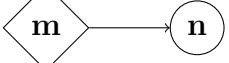


In each diagram, an oriented edge corresponds to an edge of the original path. For the convenience of drawing, each time a path enters a new sheet of  $\Gamma$  directly above a vertex with coordinate  $\mathbf{n}$ , we create a copy of that vertex on the picture and draw the corresponding

part of the path starting from that copy. Such "extra" copy is represented by a diamond-shaped node. The oriented edges connected to that copy are the corresponding ascending and descending edges of  $\Gamma$ . The non-oriented edge connecting  $\mathbf{n}$  and the diamond  $\mathbf{n}$  does not correspond to any edge of  $\Gamma$  and only represents the fact that these nodes correspond to the same vertex of the graph.

In order to recover a path from its diagram, one can start from the node representing the origin (there is only one such node) and then follow the arrows. In the event a node with coordinate  $\mathbf{n}$  is connected with a diamond-shaped node, upon arrival into  $\mathbf{n}$  one must first go to the diamond-shaped node and follow the arrows from that node; they will eventually return back to the diamond-shaped node. Afterwards, one can continue following the arrow coming out of circle-shaped  $\mathbf{n}$ . In case  $\mathbf{n}$  is connected to multiple diamond-shaped nodes, as in the last example, one needs to treat them all in counter-clockwise order before proceeding out of  $\mathbf{n}$ . In particular, this implies that the order in which different attached loops appear on the diagram is important, since the opposite attachment would correspond to a different string  $(12345(121)5(12321)54321)$ .

The contribution  $\text{Cont}(\mathcal{P})$  can be recovered from the picture by using the following rules, representing the fact that each oriented edge of  $\Gamma$  corresponds to an oriented edge on a picture and gives a multiplicative contribution. In the below notation, we assume  $\mathbf{m}, \mathbf{n} \neq \mathbf{0}$ . We also include non-oriented edges in this table for completeness, emphasizing the fact that they do not correspond to edges of  $\Gamma$  and give no contribution.

		
$\frac{\Phi_{nm}}{V_0 - V_n}$	$\Phi_{0m}$	$\frac{\Phi_{n0}}{V_0 - V_n}$
		
$-\frac{\Phi_{0m}}{V_0 - V_n}$	$\frac{\Phi_{n0}}{V_0 - V_n}$	1

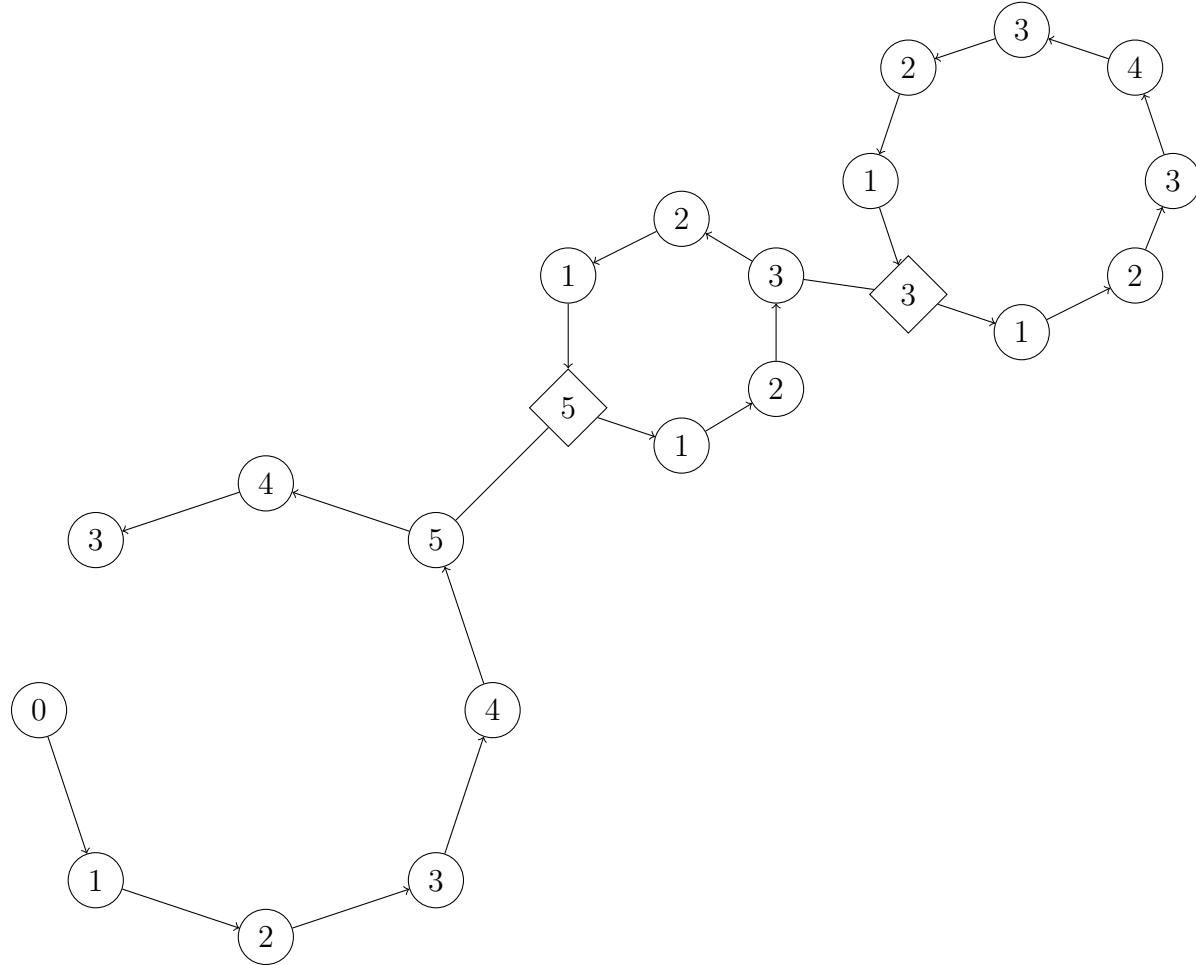
**3.3. Attachment of paths and loops.** Recall that any eigenvalue path can be moved into any point of  $\Gamma$ , where it becomes a non-base path. Suppose that  $\mathcal{P}$  is a path and  $v$  is a vertex on  $\mathcal{P}$ . Let  $\mathcal{Q}$  be an eigenvalue path. Then, one can construct a new path in the following way: take the part of  $\mathcal{P}$  until it reaches  $v$ , then insert a copy of  $\mathcal{Q}$  moved into  $v$  (which becomes a non-base path that starts and ends at  $v$ ) and then, once the copy of  $\mathcal{Q}$  returns to  $v$ , continue following the remaining part of  $\mathcal{P}$ . We will say that the new path is obtained by attaching  $\mathcal{Q}$  to  $\mathcal{P}$  at the vertex  $v$ . In the symbolic notation, suppose that  $\mathbf{n}$  is the coordinate of  $v$ . To attach  $\mathcal{Q}$ , we replace  $\mathbf{n}$  by  $\mathbf{n}\mathcal{Q}\mathbf{n}$  at the corresponding position of  $\mathbf{n}$  in  $\mathcal{P}$ . Note that  $\mathcal{P}$  can visit  $v$  multiple times. In that case, we can attach  $\mathcal{Q}$  at different positions of  $v$  on  $\mathcal{P}$  and, in general, obtain different paths.

By an *eigenvalue/eigenvector loop*, we will denote an eigenvalue/eigenvector path that does not leave the base sheet of  $\Gamma$ . It is easy to see that any path can be obtained from a loop of the same type by finitely many attachments of eigenvalue loops. For example, to

obtain the eigenvector path

$$(12345(123(1234321)321)543),$$

graphically represented as



we start from the base (eigenvector) loop  $(1234543)$ , attach  $(12321)$  at the vertex with coordinate 5, thus obtaining  $(12345(12321)543)$ , and then attach another eigenvalue loop  $(1234321)$  to the vertex with coordinate 3 in the middle. The base loop and all attached loops that were used in building  $\mathcal{P}$  will be called *loops on  $\mathcal{P}$* . Note that the difference between eigenvalue and eigenvector paths is only in the type of the base loop. In both cases, all attached loops must be of the eigenvalue type.

From (3.1), if  $\mathcal{P}'$  is obtained from  $\mathcal{P}$  by attaching an eigenvalue loop  $\mathcal{L}$  at  $\mathbf{n}$ , then

$$\text{Cont}(\mathcal{P}') = \text{Cont}(\mathcal{P})\text{Cont}(\mathcal{L}') = -(V_0 - V_n)^{-1}\text{Cont}(\mathcal{P})\text{Cont}(\mathcal{L}),$$

where  $\mathcal{L}'$  is the non-base loop obtained by moving  $\mathcal{L}$  into  $\mathbf{n}$ .

## 4. SMALL DENOMINATOR EXPANSION

**4.1. Small denominators and resonances.** Fix  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$  and consider the operator (1.1):

$$(H(x_0)\psi)_{\mathbf{n}} = \varepsilon \sum_{\mathbf{m} \in \mathbb{Z}^d} \varphi_{\mathbf{n}-\mathbf{m}} \psi_{\mathbf{m}} + f(x_0 + \mathbf{n} \cdot \omega) \psi_{\mathbf{n}}.$$

Section 2.1 provides a formal power series for eigenvalues and eigenvectors of the operator (1.1). Our goal is to establish convergence of this series, for some class of functions  $f$ . For general non-monotone  $f$  and small  $\varepsilon$ , there are two main obstructions to this convergence: small denominators and resonances. A small denominator appears whenever  $\|\mathbf{n} \cdot \omega\| = \text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z}) \ll \varepsilon$ . If we assume some regularity of  $f$ , this means that  $f(x_0) - f(x_0 + \mathbf{n} \cdot \omega)$  is small. A resonance is a situation when  $f(x_0) - f(x_0 + \mathbf{n} \cdot \omega)$  is small for some other reason: for example, if  $f$  has multiple intervals of monotonicity, it is possible for the values of  $f$  at  $x_0$  and  $x_0 + \mathbf{n} \cdot \omega$  to be close without  $\|\mathbf{n} \cdot \omega\|$  being small. Unlike small denominators, whether or not translation by  $\mathbf{n}$  creates a resonance depends strongly on  $x_0$ . *Except for the last section, we will assume that there are no resonances* (for example, by considering  $f$  that are monotone or locally monotone). Throughout the text, lattice points  $\mathbf{n}$  such that  $\|\mathbf{n} \cdot \omega\| = \text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z})$  is small, will also be called small denominators.

**Remark 4.1.** Note that, if  $\omega$  satisfies some Diophantine properties, then it takes a relatively large number of steps along the lattice to reach a small denominator. If  $\varepsilon$  is small, we can hope to compensate the contribution to  $\text{Cont}(\mathcal{P})$  from the small denominator by the number of  $\varepsilon$  factors that we gain at each step. Moreover, if we are traveling between two different small denominators, similar Diophantine arguments guarantee that the path  $\mathcal{P}$  has to make sufficiently many steps in between. Later in the text, these paths will be called safe. The contributions from safe paths form an absolutely convergent series for which no cancellations are required. However, there are also unsafe situations: suppose that a path visits the same small denominator many times, for example, by going back and forth between the small denominator  $n$  and the point  $n + 1$  (similar to earlier examples, we are considering a quasiperiodic operator on  $\ell^2(\mathbb{Z})$  and not using boldface font to denote lattice points). Then, each of these small trips gains  $\varepsilon^2$  in the numerator and a factor  $f(x_0) - f(x_0 + n\omega)$  in the denominator. If the latter factor is very small, the sum of these contributions diverges. Fortunately, in the quasiperiodic setting, the contribution of this path almost cancels with several other paths of the same length. Since these situations can happen on multiple levels simultaneously, *the goal of this section is to identify which paths are to be grouped together in order to take the biggest advantage of these cancellations*.

**4.2. Canonical marking and the equivalence relation.** To begin with, we introduce the scale of “the smallness” of denominators. We will use the distance function on  $\mathbb{Z}^d$  defined by the hopping matrix  $\Phi$ : for  $\mathbf{k}, \mathbf{k}' \in \mathbb{Z}^d$ , let  $\text{dist}_\varphi(\mathbf{k}, \mathbf{k}')$  be the smallest number  $l$  such that there exists a sequence  $\mathbf{k} = \mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_l = \mathbf{k}'$  with  $\Phi_{\mathbf{k}_{j+1}\mathbf{k}_j} \neq 0$  for  $0 \leq j \leq l - 1$ ; in other words, this is the shortest distance function on the base sheet of the graph  $\Gamma$ . We will say that two functions

$$(4.1) \quad \text{level}: \mathbb{Z}^d \rightarrow \mathbb{Z}_+ \cup \{+\infty\}, \quad \text{safedist}: \mathbb{Z}_+ \rightarrow \mathbb{Z}_+,$$

form a *consistent denominator data* if the following is true:

(c0)  $\text{level}(\mathbf{0}) = +\infty$ ,  $\text{safedist}(0) = 0$ . The function  $\text{safedist}$  is monotone non-decreasing in its argument.

- (c1)  $\text{dist}_\varphi(\mathbf{m}, \mathbf{n}) \geq \min\{\text{safedist}(\text{level}(\mathbf{m})), \text{safedist}(\text{level}(\mathbf{n}))\}$ , for  $\mathbf{m} \neq \mathbf{n}$ .
- (c2) Suppose,  $0 < \text{dist}_\varphi(\mathbf{n}, \mathbf{m}) < \text{safedist}(\text{level}(\mathbf{m}))$ . Then  $\text{level}(\mathbf{n} - \mathbf{m}) = \text{level}(\mathbf{n})$ .

Lattice points  $\mathbf{m} \neq \mathbf{0}$  with  $\text{level}(\mathbf{m}) > 0$  will be called *small denominators*. In order to simplify the notation, we will also extend the function `safedist` into  $\mathbb{Z}^d$  by

$$\text{safedist}(\mathbf{n}) := \text{safedist}(\text{level}(\mathbf{n})); \quad \text{safedist}(\mathbf{0}) := +\infty.$$

Later, we will give specific examples of consistent denominator data in the context of quasiperiodic operators. Throughout the rest of the section, most of the definitions will depend on the choice of a consistent denominator data, and we will assume that some such choice has been fixed.

**Definition 4.2.** Let  $\mathcal{L}$  be an eigenvalue/eigenvector loop. We say that  $\mathcal{L}$  is *safe*, if, between any two visits of a small denominator  $\mathbf{m}$ , it makes at least `safedist`( $\mathbf{m}$ ) steps.

We will now describe an additional structure on the set of eigenvalue/eigenvector paths that will allow us to set up the cancellation procedure. Let  $\mathcal{L}$  be an eigenvalue/eigenvector loop. Suppose,  $\mathbf{m}$  is a small denominator visited by  $\mathcal{L}$  multiple times. Consider a segment of the string that defines  $\mathcal{L}$  between two consecutive visits of  $\mathbf{m}$  by  $\mathcal{L}$ . For each such segment, we now choose either to mark it, in which case we put square parenthesis around that segment, or not to mark it. For example, if  $\mathcal{L} = (12345432321)$  and 3 is the only small denominator on  $\mathcal{L}$ , then there are four possibilities of marking  $\mathcal{L}$ :

$$(12345432321), \quad (123[454]32321), \quad (1234543[2]321), \quad (123[454]3[2]321).$$

The same entry of the string may belong to multiple marked segments. We will require the following: if two marked segments overlap, then one of them must be contained in another. As a consequence, the positions of the square parenthesis uniquely determines which segments are marked and which are not. Note that the marking is always applied to particular segments of  $\mathcal{L}$  rather than to corresponding vertices/edges of  $\Gamma$ . In particular, if a marked segment of  $\mathcal{L}$  contains a vertex  $v$ , it does not automatically imply that all consecutive visits of  $v$  by  $\mathcal{L}$  will be marked.

Under the above assumptions, if  $\mathcal{L}$  is a loop with markings, one can consider a smaller loop obtained from  $\mathcal{L}$  by removing all marked segments from it, that is, replacing each marked segment of the form  $\mathbf{m}[\dots]\mathbf{m}$  by  $\mathbf{m}$  (always starting from the shortest segment). For example, in the above four cases the removal procedure would result in

$$(12345432321), \quad (1232321), \quad (123454321), \quad (12321).$$

Note that a copy of the small denominator also gets removed from the string each time, so that the loop remains well-defined.

We now define *the canonical marking* of a loop  $\mathcal{L}$ . The goal of it is the following: whenever  $\mathcal{L}$  has two visits of a small denominator  $\mathbf{m}$  in a short time, we mark the segment between these visits. By “short”, we mean shorter than `safedist`( $\mathbf{m}$ ). However, if  $\mathcal{L}$  visits another small denominator  $\mathbf{m}'$  of smaller level between visits of  $\mathbf{m}$ , and the segment between visits of  $\mathbf{m}'$  is also short, then we do not want to include it in the calculation of time spent between visits of  $\mathbf{m}$ . More formally, do the following procedure, starting from an eigenvalue/eigenvector loop  $\mathcal{L}$ .

- (1) For each small denominator  $\mathbf{m}$  of level 1 and each segment between two consecutive visits of  $\mathbf{m}$  of length  $< \text{safedist}(1)$ , mark that segment on  $\mathcal{L}$ .
- (2) Denote by  $\mathcal{L}'$  the result of removing all marked segments on  $\mathcal{L}$ .

- (3) For each small denominator  $\mathbf{m}'$  of level 2 on  $\mathcal{L}'$  and each segment between two consecutive visits of  $\mathbf{m}'$  of length  $< \text{safedist}(2)$  on  $\mathcal{L}'$ , mark the corresponding segment on  $\mathcal{L}$ .
- (4) Denote by  $\mathcal{L}''$  the result of removing all previously marked segments on  $\mathcal{L}$ .
- (5) Repeat for levels 3,4,5,...

In other words, we mark every “short” segment between two consecutive visits of each small denominator. The notion of length used in defining “short” should not count already marked segments.

**Remark 4.3.** Assumption (c1) implies that the above procedure will actually produce a loop with markings; in other words, that for any two overlapping marked segments one of them will be contained in the other.

If  $\mathcal{L}$  is safe, then the canonical marking of  $\mathcal{L}$  coincides with  $\mathcal{L}$  and it does not have any marked segments. If one removes all marked segments from the canonical marking of  $\mathcal{L}$ , the resulting (shorter) loop will be safe. We would like to draw the reader’s attention to the following: suppose that  $\mathcal{L}$  is not safe, but the definition of “safe” only fails for one segment between two visits of a denominator, say, of level 1. It is still possible that more than one segment is marked, because the removal of first segment may shorten segments between denominators of higher levels. Therefore, it is important that the algorithm proceeds from lower to higher levels and not the other way around (although it is possible to modify it appropriately).

We will now define the *canonical translation* of a loop. We first define it on the level of strings of symbols and will later check that it defines a path on  $\Gamma$ . Let  $\mathcal{L}$  be an eigenvalue/eigenvector loop, and suppose that it is canonically marked.

- (1) For each vertex  $v$  of  $\mathcal{L}$  with coordinate  $\mathbf{n}$ , consider its position in the canonical marking of  $\mathcal{L}$ . Suppose that  $\mathbf{m}[\dots]\mathbf{m}$  is the smallest marked segment that contains  $v$ . Then replace that entry  $\mathbf{n}$  by  $\mathbf{n} - \mathbf{m}$ . If  $\mathbf{n}$  is not on a marked segment, do nothing.
- (2) Replace all square parenthesis in the string by round parenthesis.

From (c2), it follows that none of the denominators will change their level after translation. As a consequence, no point with non-zero coordinate will be translated into the origin, and therefore the above string will represent a path on  $\Gamma$ . We will denote this path by  $T(\mathcal{L})$  and call it the *canonical translation* of  $\mathcal{L}$ . Unless  $\mathcal{L}$  was safe,  $T(\mathcal{L})$  will no longer be a loop, as it will involve visits to non-base sheets (encoded by the round parenthesis).

In other words, we are replacing each entry of the form  $\mathbf{m}[\dots]\mathbf{m}$  by  $\mathbf{m}(\dots)\mathbf{m}$ , where everything between the parenthesis is translated by  $-\mathbf{m}$ , unless it is inside some smaller marked segment. For example, suppose that 3 and 6 are small denominators, and the canonical marking of  $\mathcal{L}$  is  $\mathcal{L} = (123[456[5]654]321)$ . We will use the same notation for  $\mathcal{L}$  and its canonical marking. Then

$$T(\mathcal{L}) = (123(123(-1)321)321).$$

In this case, the entry 5 in the middle is translated by  $-6$ , and the strings 456 and 654 are each translated by  $-3$ .

We have defined the canonical translation  $T(\mathcal{L})$  of a loop  $\mathcal{L}$ . Since any path on  $\Gamma$  can be obtained from loops using the attachment procedure, we naturally define  $T(\mathcal{P})$  by applying  $T$  to each loop in  $\mathcal{P}$ , and preserving the attachments at the same locations (based on their positions in the string of symbols).

**Remark 4.4.** Let  $\mathcal{L}$  be a loop and  $T(\mathcal{L})$  be its canonical translation. As mentioned above,  $T(\mathcal{L})$  is a path with multiple loops, but every loop of  $T(\mathcal{L})$  is safe. Moreover, the base loop of  $T(\mathcal{L})$  is the result of removing of all marked segments from the canonical marking of  $\mathcal{L}$ .

**Remark 4.5.** One can also define canonical marking and canonical translation directly for paths  $\mathcal{P}$ . If  $\mathcal{P}$  visits a vertex  $v$  on a sheet  $S$  twice, we will say that the visits are consecutive if, between these visits, there are no other visits of  $v$  and there are no visits of the sheet directly below  $S$ . In this case, in order to determine whether or not  $\mathcal{P}$  has made sufficiently many steps between these visits, we do not count previously marked segments and visits of sheets above  $S$ . Similarly, one needs to start from denominators of level 1 and gradually add markings whenever the part of the path between visits is too short. In the translation procedure, for each vertex  $v$ , we determine the smallest marked segment on the same sheet which contains  $v$  and translate that segment by the coordinate of its endpoints.

**Definition 4.6.** Two eigenvalue/eigenvector paths  $\mathcal{P}$  and  $\mathcal{Q}$  are called *translationally equivalent* if  $T(\mathcal{P}) = T(\mathcal{Q})$ . Denote by  $[\mathcal{P}]$  the equivalence class of  $\mathcal{P}$ , and let

$$\text{Cont}([\mathcal{P}]) := \sum_{\mathcal{P} \in [\mathcal{P}]} \text{Cont}(\mathcal{P}).$$

**Remark 4.7.** Grouping the paths into translational equivalence classes is one of the central ideas of the paper. In all applications, the cancellations will happen inside the equivalence classes. In fact, the proof of the convergence of the perturbation series

$$\sum_{k=0}^{+\infty} \varepsilon^k \left( \sum_{\mathcal{P}: |\mathcal{P}|=k} \text{Cont}(\mathcal{P}) \right)$$

will be based on absolute convergence of the series

$$\sum_{k=0}^{+\infty} \varepsilon^k \left( \sum_{[\mathcal{P}]: |\mathcal{P}|=k} |\text{Cont}([\mathcal{P}])| \right),$$

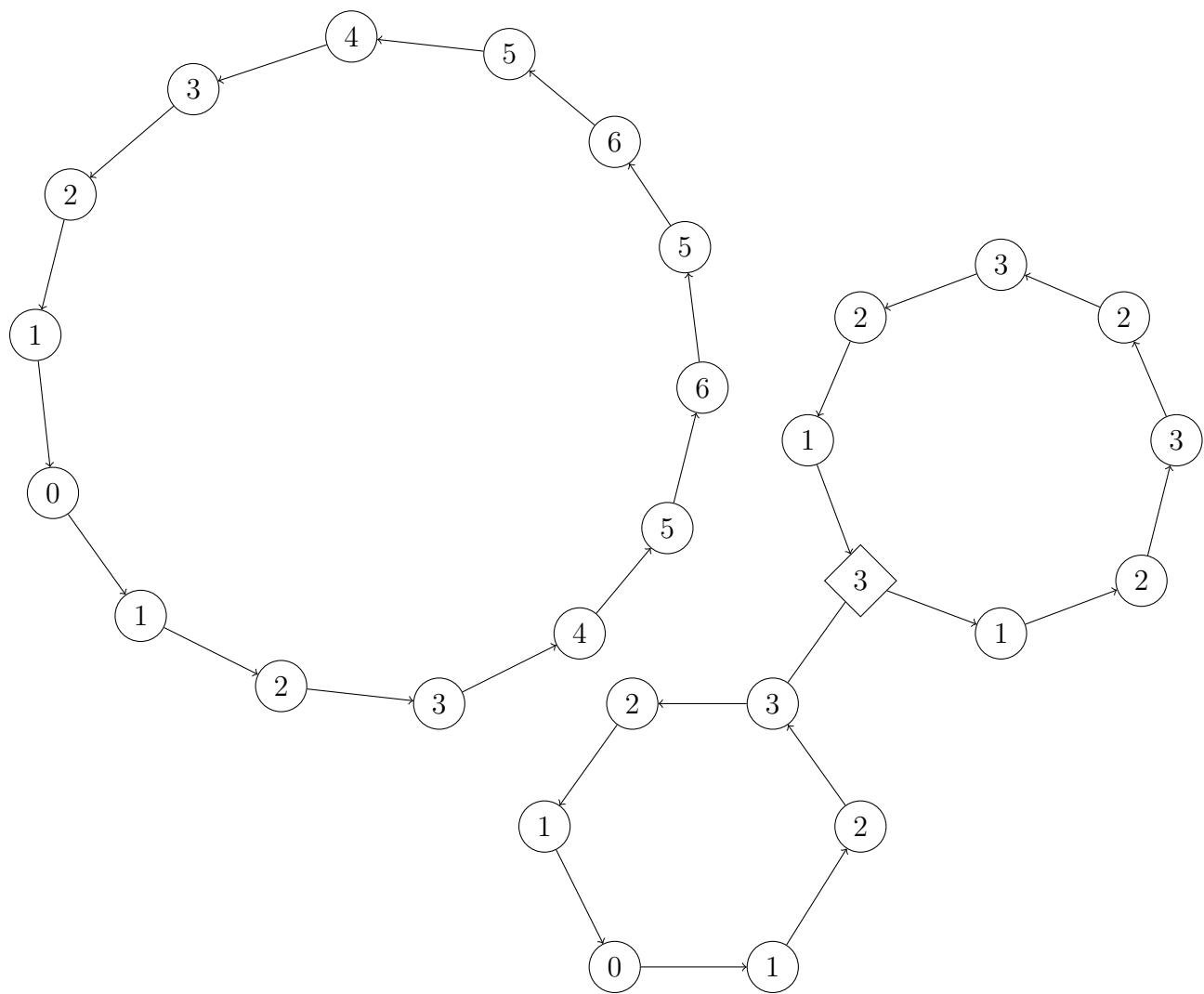
where we assume that each equivalence class is counted once in the summation. It is easy to see (cf. Remark 4.8) that the series for individual loops

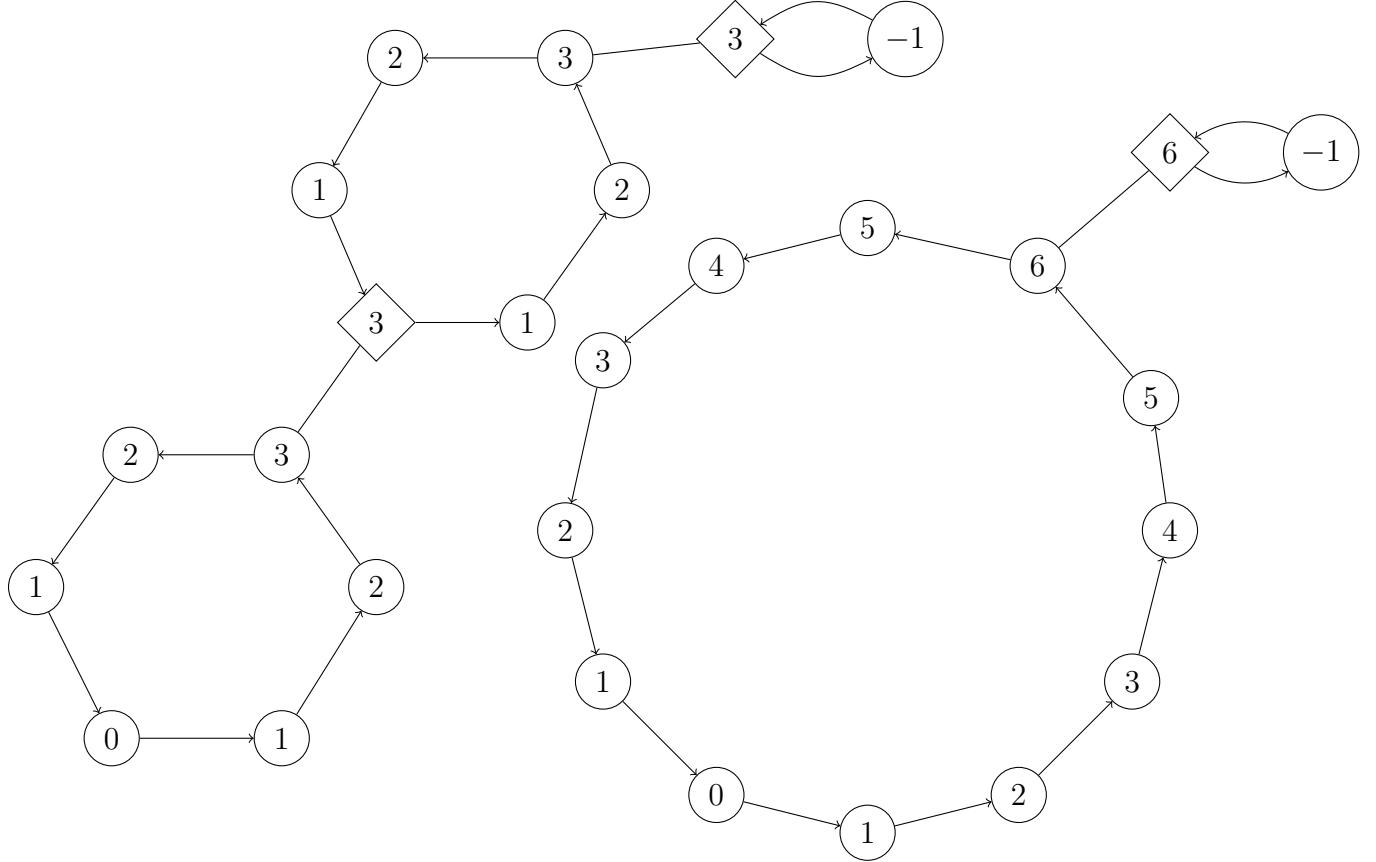
$$(4.2) \quad \sum_{k=0}^{+\infty} \varepsilon^k \left( \sum_{\mathcal{P}: |\mathcal{P}|=k} |\text{Cont}(\mathcal{P})| \right) = \sum_{\mathcal{P}} \varepsilon^{|\mathcal{P}|} |\text{Cont}(\mathcal{P})|.$$

will diverge if  $\lambda_0$  is not an isolated eigenvalue of  $H$  at  $\varepsilon = 0$  and the hopping term is the usual Laplacian (if one considers more general hopping, one needs to assume that  $\Gamma$  is not “too disconnected”).

**4.3. Some examples.** The loop with canonical marking  $\mathcal{L} = (123[456[5]654]321)$  has four elements in its equivalence class:  $(1234565654321)$ ,  $(123(1232321)321)$ ,  $(123(123(-1)321)321)$ ,

and  $(123456(-1)654321)$ , graphically represented by





**Remark 4.8.** As mentioned in Remark 4.7, the cancellation will be achieved inside each equivalence class. Before proceeding with the general construction, let us illustrate the usefulness of these equivalence classes on the following example. Let  $d = 1$ ,  $V_n = f(n\omega)$ , where, say,  $f$  is a smooth 1-periodic function on  $\mathbb{R}$  such that  $f'(x) \geq 1$  in a neighborhood of 0. Let  $\omega \in \mathbb{R} \setminus \mathbb{Q}$  and  $\Phi$  be the usual discrete Laplacian operator:

$$\Phi_{ij} = (\Delta)_{ij} = \delta_{i,j+1} + \delta_{i,j-1}, \quad H = V + \varepsilon \Delta.$$

Suppose that, say,  $\text{dist}(5\omega, \mathbb{Z}) \ll \varepsilon$ , so that  $|V_5| = |f(0) - f(5\omega)| \ll \varepsilon$ . Suppose also that 5 is the only small denominator:

$$\text{level}(5) = 1, \quad \text{level}(1) = \text{level}(2) = \text{level}(3) = \text{level}(4) = 0, \quad \text{safedist}(5) \geq 3.$$

Typically, we would have  $|V_0 - V_5| \approx \varepsilon^{\text{safedist}(5)}$ , which would mean that one needs to make  $\text{safedist}(5) \geq 3$  steps to compensate for the factor  $|f(0) - f(5\omega)|^{-1}$ . Consider the following eigenvalue path:

$$\mathcal{P}_k = (123454545 \dots 454321)$$

where  $\dots$  represents repetition of "45" so that the loop has  $k$  visits of 5. Clearly,

$$\text{Cont}(\mathcal{P}_k) = (V_0 - V_1)^{-1} (V_0 - V_2)^{-1} (V_0 - V_3)^{-1} (V_0 - V_4)^{-(k+1)} (V_0 - V_5)^{-k} (V_0 - V_3)^{-1} (V_0 - V_2)^{-1} (V_0 - V_1)^{-1}.$$

Since the loop  $\mathcal{P}_k$  only makes two steps between several consecutive visits of 5, it is considered non-safe, and its contributions to (4.2) will blow up, even with the factor  $\varepsilon^{|\mathcal{P}_k|}$ , as  $k \rightarrow \infty$ . The canonical marking of  $\mathcal{P}$  will be as follows:

$$\mathcal{P} = (12345[4]5[4]5 \dots [4]54321),$$

with the canonical translation

$$T(\mathcal{P}) = (12345(-1)5(-1)5 \dots (-1)54321).$$

Here, the vertices in round brackets denote vertices on the sheet above 5. After the translation, the vertex 4 becomes  $-1$ . In both cases, there are  $k-1$  times when a sheet of height 1 is visited. The whole translation equivalence class consists of  $2^{k-1}$  elements, since there are  $k-1$  choices between [4] and  $(-1)$ . One can factor the total contribution of the equivalence class as follows:

$$\begin{aligned} \text{Cont}([\mathcal{P}]) &= \text{Cont}(123454321)(V_0 - V_5)^{-(k-1)}((V_0 - V_4)^{-1} - (V_0 - V_{-1})^{-1})^{k-1} \\ &= \text{Cont}(123454321) \frac{((f(0) - f(4\omega))^{-1} - (f(0) - f(-\omega))^{-1})^{k-1}}{(f(0) - f(5\omega))^{k-1}}, \end{aligned}$$

since the change from [4] to  $(-1)$  replaces one of the terms  $-V_4$  by  $V_{-1}$ . Note that, since  $\|5\omega\|$  is small, we have  $f(-\omega) \approx f(4\omega)$ . Therefore,

$$|(f(0) - f(4\omega))^{-1} - (f(0) - f(-\omega))^{-1}| \approx \|5\omega\| \frac{|f'(4\omega)|}{(f(0) - f(4\omega))^2}$$

and

$$|f(0) - f(5\omega)| \approx f'(0)\|5\omega\| \geq \|5\omega\|,$$

which implies that

$$|\text{Cont}(\mathcal{P})| \approx \text{Cont}(123454321) \frac{|f'(4\omega)|^{k-1}}{f'(0)^{k-1}|f(0) - f(4\omega)|^{2k-2}}.$$

Since  $f$  does not vanish around  $4\omega$ , we can see that the “bad” factor  $\|5\omega\|^{-1}$  enters  $\text{Cont}(\mathcal{P})$  with power 1 rather than  $k$ .

Unfortunately, applying this general idea to more complicated situations requires tedious bookkeeping, which we will now describe.

**4.4. Loop stacks.** In order to describe the general case, motivated by the last example, we will use the following definition.

**Definition 4.9.** A *loop stack* is an eigenvalue/eigenvector path  $\mathcal{P}$  on  $\Gamma$  satisfying the following:

- (1) All loops of  $\mathcal{P}$  are safe.
- (2) For any loop  $\mathcal{L}$  on  $\mathcal{P}$  that is attached to a vertex with coordinate  $\mathbf{n}$ , we have  $|\mathcal{L}| < \text{safedist}(\mathbf{n})$ . In particular, the loops on  $\mathcal{P}$  can only be attached to small denominators.

We can now describe the translational equivalence class of any path  $\mathcal{P}$ . Let  $T(\mathcal{P})$  be the canonical translation of  $\mathcal{P}$ . We will say that a loop  $\mathcal{L}$  on  $T(\mathcal{P})$  is *short*, if it is attached to a vertex with coordinate  $\mathbf{n}$ , with  $|\mathcal{L}| < \text{safedist}(\mathbf{n})$ . The base loop is never considered short. Since  $T(\mathcal{P}) \in [\mathcal{P}]$ , we can choose  $T(\mathcal{P})$  as the “natural representative” of the equivalence class of  $\mathcal{P}$ .

- Let  $\mathcal{P}$  be an eigenvalue/eigenvector path on  $\Gamma$ , and suppose that  $T(\mathcal{P})$  has  $s$  short loops. Then  $\#[\mathcal{P}] = 2^s$ . The elements of  $[\mathcal{P}]$  are obtained by replacing some short loops on  $T(\mathcal{P})$  by marked segments:

$$\mathbf{m}(\dots)\mathbf{m} \mapsto \mathbf{m}[\dots + \mathbf{m}]\mathbf{m},$$

where  $(\dots)$  denotes a short loop, and  $\dots + \mathbf{m}$  denotes translation of all coordinates by  $\mathbf{m}$ .

- All loops of  $T(\mathcal{P})$  are safe, and  $T(\mathcal{P})$  is the only element of  $[\mathcal{P}]$  with this property.
- One can introduce an attachment procedure for loop stacks similarly to loops: if  $\mathcal{P}_0$  is an eigenvalue/eigenvector path that has a vertex  $v$  with coordinate  $\mathbf{n}$  and  $\mathcal{P}_1$  is an eigenvalue loop stack, one can attach  $\mathcal{P}_1$  to  $\mathcal{P}_0$  by replacing

$$\mathbf{n} \mapsto \mathbf{n}\mathcal{P}_1\mathbf{n}$$

in the string that describes  $\mathcal{P}_0$ . Any path  $\mathcal{P}$  with  $T(\mathcal{P}) = \mathcal{P}$  can be obtained from a base (eigenvalue/eigenvector) loop stack by a sequence of the above attachment operations, at the points of the base stack or already attached loop stacks. Each attached loop stack becomes a non-base loop stack on  $\mathcal{P}$ .

Now, given a path  $\mathcal{P} = T(\mathcal{P})$ , we would like to break it into loop stacks in some natural way. In general, there are multiple ways of doing so: for example, one can treat each loop as a separate loop stack. The opposite of this would be a decomposition with smallest possible number of loop stacks. Note that any stack may contain at most one non-short loop (which can only be its base loop). As a consequence, each non-short loop  $\mathcal{L}$  of  $\mathcal{P}$ , including the base loop of  $\mathcal{P}$ , must be the base loop of some stack. For each such loop, we can form the largest possible stack that starts from that loop using the following “greedy algorithm”: consider all short loops attached to  $\mathcal{L}$ ; now, consider all short loops attached to the loops considered before, and repeat the process until we run out of attached short loops. It is easy to see that these stacks will not overlap and will ultimately contain each loop on  $\mathcal{P}$ . We will call these stacks *maximal*.

- Suppose that the above procedure results in the path  $\mathcal{P} = T(\mathcal{P})$  being decomposed into maximal stacks  $\mathcal{P}_0, \dots, \mathcal{P}_n$ , where  $\mathcal{P}_0$  contains the base loop of  $\mathcal{P}$ , and  $\mathcal{P}_j$  is attached to a small denominator  $\mathbf{n}_j$  on one of the previous stacks  $\mathcal{P}_0, \dots, \mathcal{P}_{j-1}$ . In particular, the base loops of  $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_n$  are all non-short loops on  $\mathcal{P}$ . Then

$$(4.3) \quad \text{Cont}([\mathcal{P}]) = \text{Cont}(\mathcal{P}_0) \prod_{s=1}^n (V_{\mathbf{n}_s} - V_0)^{-1} \text{Cont}([\mathcal{P}_s]).$$

Our next objective is to obtain a bound of the form  $\varepsilon^{|\mathcal{P}_s|} |\text{Cont}([\mathcal{P}_s])| \lesssim \varepsilon^{c|\mathcal{P}_s|}$  by taking advantage of cancellations inside  $[\mathcal{P}_s]$ . In other words, we expect  $[\mathcal{P}_s]$  to behave as a safe loop of comparable length, thanks to the base loop of  $\mathcal{P}_s$  being non-short. Since  $|\mathcal{P}_s| \geq \text{safedist}(\mathbf{n}_s)$ , we can absorb the  $(V_{\mathbf{n}_s} - V_0)^{-1}$  factors into  $\text{Cont}([\mathcal{P}_s])$  if we have bounds of the form

$$\varepsilon^{c \cdot \text{safedist}(\mathbf{n}_s)} \ll |V_0 - V_{\mathbf{n}_s}|.$$

In the quasiperiodic case, this will be achieved by imposing a Diophantine condition on the frequency, combined with choosing a function `safedist` that does not grow too slowly, see Section 5 and, in particular, Theorem 5.10.

## 5. CONVERGENCE OF THE PERTURBATION SERIES

**5.1. Regular points of monotone potentials.** The cancellation procedure described above will be applied to quasiperiodic operators, that is, when

$$V_{\mathbf{n}} = f(x + \mathbf{n} \cdot \omega),$$

$f$  is an 1-periodic function on  $\mathbb{R}$ , and  $\omega = (\omega_1, \dots, \omega_d)$  is a vector with rationally independent components. Theorem 3.2 gives an asymptotic series for an eigenvalue which would have been an analytic continuation of the eigenvalue  $V_0 = f(x)$  of the operator at  $\varepsilon = 0$ , if that eigenvalue was isolated. In our case, it is not isolated; however, as long as the algebraic non-resonance condition

$$f(x + \mathbf{n} \cdot \omega) \neq f(x) \quad \text{for } \mathbf{n} \neq \mathbf{0}$$

is satisfied, each term of the perturbation series is well-defined as discussed in Section 3.

We will formulate the convergence results locally, that is, for fixed  $x$ , assuming that  $f$  is monotone and one-to-one in a neighborhood of  $x$ . We will also provide sufficient conditions for obtaining convergent expansions for eigenvectors for all  $x \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ . In all cases, we will consider the eigenvector associated with the origin (that is, a perturbation series that starts from  $\psi_0 = e_0$ ,  $\lambda_0 = f(x)$ ). However, if we obtain such eigenfunctions for all  $H(x)$  with  $x \in \{x_0 + \mathbf{n}_0 \cdot \omega : \mathbf{n}_0 \in \mathbb{Z}^d\}$ , their translations will form a complete system of eigenvectors of  $H(x_0)$  sufficient to establish Anderson localization for that operator.

We will always assume the following:

(f1)  $f: (-1/2, 1/2) \rightarrow \mathbb{R}$  is continuous,  $f(-1/2 + 0) = -\infty$ ,  $f(1/2 - 0) = +\infty$ , and is extended by 1-periodicity into  $\mathbb{R} \setminus (\mathbb{Z} + 1/2)$ .

Suppose,  $f$  satisfies (f1). Let

$$C_{\text{reg}} > 0, \quad x_0 \in (-1/2, 1/2)$$

We say that  $f$  is  $C_{\text{reg}}$ -regular at  $x_0$ , if

(cr0) The pre-image  $f^{-1}((f(x_0) - 2, f(x_0) + 2)) \cap (-1/2, 1/2)$  is an open interval (denoted by  $(a, b)$ ), and  $f|_{(a,b)}$  is a one-to-one map between  $(a, b)$  and  $(f(x_0) - 2, f(x_0) + 2)$ .  
 (cr1) Let  $D_{\min}(x_0) := \inf_{x \in (a,b)} f'(x)$ . Then,

$$(5.1) \quad D_{\min}(x_0) \leq f'(x) \leq C_{\text{reg}} D_{\min}(x_0), \quad \forall x \in (a, b).$$

At the points where  $f'(x)$  does not exist, we require that the inequalities hold for all derivative numbers (the inf is also taken over the set of all derivative numbers).

(cr2) Define  $(a_1, b_1) = f^{-1}(f(x_0) - 1, f(x_0) + 1) \subset (a, b)$ , and

$$g(x) = \frac{1}{f(x_0) - f(x)}, \quad x \in (b_1, a_1 + 1),$$

extended by continuity to  $g(0) = g(1) = 0$  (recall that we also assume  $f(x + 1) = f(x)$ , so that the interval  $(b_1, a_1 + 1)$  is essentially  $(-1/2, 1/2) \setminus (a_1, b_1)$  together with the point  $0 = 1 \bmod 1$ ). Then, under the same conventions on the existence of derivatives,

$$|g'(x)| \leq C_{\text{reg}} D_{\min}(x_0), \quad x \in (b_1, a_1 + 1).$$

**Remark 5.1.** Suppose,  $f$  satisfies (f1) and

$$f'(x) < C f'(y) (1 + |f(x) - f(y)|^2) \quad \text{for all } x, y \in (-1/2, 1/2).$$

Then  $f$  is  $C_{\text{reg}}$ -regular on  $(-1/2, 1/2)$  (with the value of  $C_{\text{reg}}$  depending on  $C$ ). In particular, any meromorphic function satisfying (f1), such that  $f'(x) > 0$  on  $(-1/2, 1/2)$ , has this property.

**Remark 5.2.** One can see Condition (cr1) as a statement that  $f'$  does not oscillate too much on intervals of length determined by the change of the values of  $f$ . Note that the interval  $(a, b)$  will shrink as one of its endpoints approaches  $\{-1/2, 1/2\}$ . Condition (cr2) is a “regularity at infinity” condition, which will be important to account for non-small denominators.

**Remark 5.3.** In some later results, it will be convenient to consider a “rescaled” version of  $C_{\text{reg}}$ -regularity. To obtain the definition of  $C_{\text{reg},\nu}$ -regularity, replace the interval  $(f(x_0) - 2, f(x_0) + 2)$  in (cr0) by  $(f(x_0) - 2\nu, f(x_0) + 2\nu)$  and the interval  $(f(x_0) - 1, f(x_0) + 1)$  in (cr2) by  $(f(x_0) - \nu, f(x_0) + \nu)$ .

**5.2. Consistent level functions for monotone quasiperiodic potentials.** We will start by constructing a consistent denominator data on  $\mathbb{Z}^d$ , suitable for use with locally monotone quasiperiodic operators. Recall that we need to define two functions `safedist` and `level`, satisfying

- (c0) `level(0) = +∞`, `safedist(0) = 0`. The function `safedist` is monotone non-decreasing in its argument.
- (c1)  $\text{dist}_\varphi(\mathbf{m}, \mathbf{n}) \geq \min\{\text{safedist}(\mathbf{m}), \text{safedist}(\mathbf{n})\}$ , for  $\mathbf{m} \neq \mathbf{n}$ .
- (c2) Suppose  $0 < \text{dist}_\varphi(\mathbf{n}, \mathbf{m}) < \text{safedist}(\mathbf{m})$ . Then `level(n - m) = level(n)`.

Let  $\omega \in (-1/2, 1/2)^d$  be a Diophantine frequency vector, that is,

$$(5.2) \quad \|\mathbf{n} \cdot \omega\| = \text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z}) \geq C_{\text{dio}} |\mathbf{n}|^{-\tau}, \quad \forall \mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}.$$

We will always assume that some  $\tau > 1$  is fixed.

**Theorem 5.4.** Fix  $C_{\text{safe}} > 0$  and, for  $\beta > 0$ , define

$$(5.3) \quad \beta_k := \lfloor \beta^{-k-1} \rfloor^{-1}, \quad k \in \mathbb{Z}_+; \quad \beta_{-1} := +\infty.$$

Define the level and safe distance functions as follows:

$$(5.4) \quad \text{level}(\mathbf{n}) = k \quad \text{iff} \quad \beta_k \leq \|\mathbf{n} \cdot \omega\| < \beta_{k-1}, \quad \mathbf{n} \in \mathbb{Z}^d \setminus \{\mathbf{0}\}; \quad \text{level}(\mathbf{0}) = +\infty;$$

$$(5.5) \quad \text{safedist}(\mathbf{n}) = \lceil C_{\text{safe}} \text{level}(\mathbf{n})^3 \rceil, \quad \mathbf{n} \in \mathbb{Z}^d.$$

Then, for

$$(5.6) \quad 0 < \beta < \beta_{\max}(C_{\text{safe}}, C_{\text{dio}}, \text{dist}_\varphi, \tau)$$

the constructed functions `level(n)` and `safedist(n)` satisfy (c1) and (c2), and therefore define a consistent denominator data on  $\mathbb{Z}^d$ .

*Proof.* Suppose,  $\min\{\text{level}(\mathbf{n}), \text{level}(\mathbf{m})\} = k \geq 1$ . The Diophantine condition implies that, if  $\mathbf{n} \neq \mathbf{m}$ ,

$$(5.7) \quad \text{dist}_\varphi(\mathbf{m}, \mathbf{n}) \geq c_1(C_{\text{dio}}, \tau, \text{dist}_\varphi) \beta^{-k/\tau}.$$

One can obtain (c1) as long as  $\beta_{\max}$  is chosen to satisfy

$$C_{\text{safe}} k^3 \leq c_1(C_{\text{dio}}, \tau, \text{dist}_\varphi) \beta^{-k/\tau},$$

that is,

$$\beta \leq c_2(C_{\text{dio}}, \tau, \text{dist}_\varphi)^{\tau/k} C_{\text{safe}}^{-\tau/k} k^{-3\tau/k}, \quad \forall k \geq 1.$$

The worst case is, essentially, either  $k = 1$  or  $k \rightarrow \infty$  (up to a factor that can be absorbed into  $c_2$ ), and hence (5.6) implies (c1), under an appropriate choice of  $\beta_{\max}(C_{\text{safe}}, C_{\text{dio}}, \text{dist}_\varphi, \tau)$ .

Let us establish (c2). Let  $\mathbf{m}, \mathbf{n} \in \mathbb{Z}^d$ . By the Diophantine condition, for  $k \geq 0$ ,

$$(5.8) \quad \|\|(\mathbf{n} - \mathbf{m}) \cdot \omega\| - \beta_k\| \geq \lfloor \beta^{-k-1} \rfloor^{-1} \|\lfloor \beta^{-k-1} \rfloor (\mathbf{n} - \mathbf{m}) \cdot \omega\| \geq C_{\text{dio}} \beta_k^{1+\tau} |\mathbf{n} - \mathbf{m}|^{-\tau}.$$

If

$$(5.9) \quad C_{\text{dio}} |\mathbf{n} - \mathbf{m}|^{-\tau} \geq 2\beta_k,$$

then one can also obtain a better bound

$$(5.10) \quad \|\|(\mathbf{n} - \mathbf{m}) \cdot \omega\| - \beta_k\| \geq \|\|(\mathbf{n} - \mathbf{m}) \cdot \omega\| - \beta_k\| \geq \frac{1}{2} C_{\text{dio}} |\mathbf{n} - \mathbf{m}|^{-\tau}.$$

If (5.10) cannot be obtained, then the opposite of (5.9) holds, that is,

$$(5.11) \quad |\mathbf{n} - \mathbf{m}| > 2^{-1/\tau} C_{\text{dio}}^{1/\tau} \beta_k^{-1/\tau}.$$

Now, suppose that  $0 < \text{dist}_\varphi(\mathbf{n}, \mathbf{m}) < \text{safedist}(\mathbf{m})$ . Then (5.5) yields

$$\text{level}(\mathbf{m}) \geq C_{\text{safe}}^{-1/3} \text{dist}_\varphi(\mathbf{n}, \mathbf{m})^{1/3} \geq C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) |\mathbf{n} - \mathbf{m}|^{1/3} > 0,$$

which, combined with (5.4), implies

$$(5.12) \quad \|\mathbf{m} \cdot \omega\| \leq 2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) |\mathbf{n} - \mathbf{m}|^{1/3} \rceil}.$$

We would like to show that addition of  $\mathbf{m}$  will not change the level of  $\mathbf{n} - \mathbf{m}$ . This would follow from

$$\|\mathbf{m} \cdot \omega\| < \min\{\|(\mathbf{n} - \mathbf{m}) \cdot \omega\| - \beta_k, \beta_{k-1} - \|(\mathbf{n} - \mathbf{m}) \cdot \omega\|\}, \quad \text{where } k = \text{level}(\mathbf{n} - \mathbf{m}).$$

We will verify that the above inequality follows from the lower bounds (5.8), (5.10) and the upper bound (5.12). Suppose, (5.10) holds. Then we need

$$(5.13) \quad 2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) |\mathbf{n} - \mathbf{m}|^{1/3} \rceil} < \frac{1}{2} C_{\text{dio}} |\mathbf{n} - \mathbf{m}|^{-\tau}.$$

Suppose now that (5.10) does not hold and we have to use (5.8). Then one needs to establish

$$(5.14) \quad 2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) |\mathbf{n} - \mathbf{m}|^{1/3} \rceil} < C_{\text{dio}} \beta_k^{1+\tau} |\mathbf{n} - \mathbf{m}|^{-\tau}.$$

Using (5.11), one can replace  $|\mathbf{n} - \mathbf{m}|$  in the left hand side by

$$\delta_{k,k} := \frac{1}{2} |\mathbf{n} - \mathbf{m}| + 2^{-1-1/\tau} C_{\text{dio}}^{1/\tau} \beta_k^{-1/\tau}$$

and reduce to the inequality

$$(5.15) \quad 2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) \delta_{k,k}^{1/3} \rceil} < C_{\text{dio}} \beta_k^{1+\tau} |\mathbf{n} - \mathbf{m}|^{-\tau}.$$

Both (5.13) and (5.15) can be obtained by taking a sufficiently small  $\beta$ , as the bound in the left hand side is exponentially better both in  $|\mathbf{n} - \mathbf{m}|$  and in  $\beta_k$ . If  $k \geq 1$ , one also needs a similar bound for  $\|\|(\mathbf{n} - \mathbf{m}) \cdot \omega\| - \beta_{k-1}\|$ , which reduces to

$$2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) |\mathbf{n} - \mathbf{m}|^{1/3} \rceil} < \frac{1}{2} C_{\text{dio}} |\mathbf{n} - \mathbf{m}|^{-\tau}$$

if  $C_{\text{dio}} |\mathbf{n} - \mathbf{m}|^{-\tau} \geq 2\beta_{k-1}$ , and otherwise to

$$2\beta^{\lceil C_{\text{safe}}^{-1/3} c(\text{dist}_\varphi) \delta_{k,k-1}^{1/3} \rceil} < C_{\text{dio}} \beta_{k-1}^{1+\tau} |\mathbf{n} - \mathbf{m}|^{-\tau},$$

where

$$\delta_{k,k-1} := \frac{1}{2} |\mathbf{n} - \mathbf{m}| + 2^{-1-1/\tau} C_{\text{dio}}^{1/\tau} \beta_{k-1}^{-1/\tau}.$$

In both cases, the inequalities can be obtained in the same way as before.  $\square$

**5.3. Lipschitz bounds for loop stacks.** In this subsection, we will obtain the central inductive estimates on  $\text{Cont}([\mathcal{P}])$ , where  $\mathcal{P}$  is a loop stack, by taking advantage of the cancellations. It will be convenient to introduce some notation in the beginning. In the remainder of the section, we will assume that  $\beta \in (0, 1)$  is fixed and the denominators are leveled according to Theorem 5.4. The main results will be obtained under additional assumptions on  $\beta$  being small enough. Let  $\mathcal{P}$  be an eigenvalue/eigenvector loop stack.

- Denote by  $\text{height}(\mathcal{P})$  the height of  $\mathcal{P}$ , which is the maximal height of sheets of  $\Gamma$  visited by  $\mathcal{P}$ .
- Let  $\text{height}(k)$  and  $\text{maxlevel}(k)$ , respectively, be the maximal possible value of  $\text{height}(\mathcal{P})$  and the maximal level of a denominator on  $\mathcal{P}$ , considered over all loop stacks  $\mathcal{P}$  with  $|\mathcal{P}| = k$ .
- Let  $m \in \mathbb{R}$ . For each loop of  $\mathcal{P}$ , calculate the maximal level of denominators on that loop. Among these numbers, consider only those that are  $\geq m$  and add them together. Denote the resulting number by  $\text{level}(\mathcal{P}, m)$ . Let also  $\text{level}(\mathcal{P}) = \text{level}(\mathcal{P}, 0)$ .
- Let  $\text{den}(\mathcal{P}, m)$  be the total number of denominators on  $\mathcal{P}$  of level  $\geq m$ . Here, each denominator is counted as many times as the corresponding lattice point is visited by  $\mathcal{P}$ ; however, the contributions from the descending edges are not counted. Let also  $\text{totalllevel}(\mathcal{P}, m)$  denote the sum of levels of all these denominators.
- Let  $\text{loops}(\mathcal{P}, m)$  be the number of loops of  $\mathcal{P}$  that contain a denominator of level  $\geq m$ . We have  $\text{den}(\mathcal{P}, 0) + \text{loops}(\mathcal{P}, 0) = |\mathcal{P}|$ .
- Let  $\text{nbloops}(\mathcal{P}, m)$  and  $\text{nblevel}(\mathcal{P}, m)$  denote the same quantities as above ( $\text{loops}$  and  $\text{level}$ ), but the base loop is not counted.
- Let also  $\text{downedges}(\mathcal{P}, m)$  be the number of descending edges on  $\mathcal{P}$  that lead to denominators of level  $\geq m$ . We have  $\text{downedges}(\mathcal{P}, 0) = \text{loops}(\mathcal{P}, 0) - 1$ , since  $\mathcal{P}$  has to exit each non-base loop once.

**Lemma 5.5.** *In the above notation, the following bounds hold:*

$$(5.16) \quad \text{maxlevel}(k) \leq \max \left\{ c(C_{\text{dio}}, \tau, \text{dist}_\varphi) \frac{\log(k)}{\log(\beta^{-1})}, 0 \right\}$$

$$(5.17) \quad \text{height}(k) \leq 1 + \text{height}(\text{safedist}(\text{maxlevel}(k))) = 1 + \text{height}(\lceil C_{\text{safe}} \text{maxlevel}(k)^3 \rceil).$$

*Proof.* Let  $\mathcal{P}$  be an eigenvalue/eigenvector loop stack with  $|\mathcal{P}| = k$ . Estimate (5.16) follows directly from the Diophantine property. To show (5.17), note that any loop on  $\mathcal{P}$  that is directly attached to the base loop, can be of length at most  $\text{safedist}(\text{maxlevel}(k))$ . Then one can consider the stack that starts from that loop and apply the induction assumption.  $\square$

**Remark 5.6.** In order to obtain a meaningful bound, one needs to choose the parameters to satisfy  $C_{\text{safe}} \text{maxlevel}(k)^3 \ll k$ , in which case  $\text{height}(k)$  will be a very slowly growing function. The bound can be achieved either by choosing small  $C_{\text{safe}}$  or small  $\beta$ .

Recall that, for a fixed  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ , we have

$$V_{\mathbf{n}} = f(x_0 + \mathbf{n} \cdot \omega).$$

In order to describe loop translations, it will be convenient to introduce an additional parameter  $t$  as follows:

$$V_{\mathbf{n}}(t) := \begin{cases} f(t + x_0 + \mathbf{n} \cdot \omega), & \mathbf{n} \neq \mathbf{0} \\ f(x_0), & \mathbf{n} = \mathbf{0}. \end{cases}$$

so that  $V_{\mathbf{n}}(0) = V_{\mathbf{n}}$ ,  $V_{\mathbf{0}}(t) \equiv V_{\mathbf{0}}$ . Denote by  $\text{Cont}(\mathcal{P}, t)$  the function obtained by replacing all  $V_{\mathbf{n}}$  in  $\text{Cont}(\mathcal{P})$  with  $V_{\mathbf{n}}(t)$ . We will only consider this function in a neighborhood of the origin small enough (cf. (5.19)) so that  $V_{\mathbf{n}}(t) \neq V_{\mathbf{0}}$  for  $\mathbf{n} \neq \mathbf{0}$ , thus avoiding zero denominators. Clearly,  $\text{Cont}(\mathcal{P}) = \text{Cont}(\mathcal{P}, 0)$ . Let us also extend the above definition to the equivalence class:

$$\text{Cont}([\mathcal{P}], t) := \sum_{\mathcal{P}' \in [\mathcal{P}]} \text{Cont}(\mathcal{P}', t).$$

**Lemma 5.7.** *Suppose,  $f$  is  $C_{\text{reg}}$ -regular at  $x_0$ . Let  $\mathbf{n} \in \mathbb{Z}^d \setminus \{0\}$ ,  $|t| \leq \frac{1}{4}\|\mathbf{n} \cdot \omega\|$ . Then*

$$\begin{aligned} (1) \quad & \left| \frac{1}{V_{\mathbf{0}} - V_{\mathbf{n}}(t)} \right| \leq \max \left\{ \frac{4}{D_{\min}(x_0)\beta^{\text{level}(\mathbf{n})+1}}, 1 \right\}; \\ (2) \quad & \left| \frac{V'_{\mathbf{n}}(t)}{(V_{\mathbf{0}} - V_{\mathbf{n}}(t))^2} \right| \leq C_{\text{reg}} \cdot \max \left\{ \frac{4}{\beta^{\text{level}(\mathbf{n})+1}}, D_{\min}(x_0) \right\} \cdot (\text{r.h.s of (1)}). \end{aligned}$$

*Proof.* First, let us note that, if  $\{t + \mathbf{n} \cdot \omega\}$  is large enough so that

$$x_0 + t + \mathbf{n} \cdot \omega \notin (a, b) := f^{-1}((f(x_0) - 2, f(x_0) + 2)) \cap (-1/2, 1/2),$$

then one can use (cr0) and bound the denominator by 1 in (1) and use (cr2) in (2). Otherwise, one can use the lower bound in (cr1) (5.1) combined with  $\|t + \mathbf{n} \cdot \omega\| \geq \frac{3}{4}\|\mathbf{n} \cdot \omega\|$ .  $\square$

If  $D_{\min}(x_0)$  is large, some denominators of small levels may actually be not very small. It is convenient to introduce an extra parameter which will indicate the minimal level at which a denominator can in principle be smaller than 1. Denote by

$$(5.18) \quad M_{\beta} = \frac{\log(4/D_{\min}(x_0))}{\log \beta} - 1,$$

that is,

$$\beta^{M_{\beta}+1} = \frac{4}{D_{\min}(x_0)}.$$

In the sequel, it will be convenient to use the notation  $\mathbf{n} \in \mathcal{L}$  if the loop  $\mathcal{L}$  visits a lattice point  $\mathbf{n}$ .

**Lemma 5.8.** *Let  $\mathcal{L}$  be an eigenvalue/eigenvector loop. Suppose,  $f$  is  $C_{\text{reg}}$ -regular at  $x_0$  and*

$$(5.19) \quad |t| \leq \frac{1}{4} \min_{\mathbf{n} \in \mathcal{L}} \|\mathbf{n} \cdot \omega\|.$$

*Define  $M_{\beta}$  by (5.18). Then*

- (1)  $|\text{Cont}(\mathcal{L}, t)| \leq \|\varphi\|_{\infty}^{|\mathcal{L}|} \left( \frac{4}{D_{\min}(x_0)} \right)^{\text{den}(\mathcal{L}, M_{\beta})} \beta^{-\text{totalevel}(\mathcal{L}, M_{\beta}) - \text{den}(\mathcal{L}, M_{\beta})}.$
- (2) *The function  $\text{Cont}(\mathcal{L}, \cdot)$  is Lipschitz continuous:*

$$\left| \frac{d}{dt} \text{Cont}(\mathcal{L}, t) \right| \leq |\mathcal{L}| \cdot C_{\text{reg}} \max \left\{ \frac{4}{\beta^{\text{level}(\mathcal{L})+1}}, D_{\min}(x_0) \right\} \cdot (\text{r.h.s of (1)}).$$

*Proof.* Both estimates directly follow from Lemma 5.7 and counting the contributions from each denominator.  $\square$

**Remark 5.9.** The previous lemma does not assume  $\mathcal{L}$  to be safe. If it is safe, then one can estimate the number of denominators of level  $\geq M \geq 1$  on  $\mathcal{L}$  by  $\lceil c(\text{dist}_\varphi) \frac{|\mathcal{L}|}{M^3 C_{\text{safe}}} \rceil$ . As a consequence,  $\text{totalllevel}(\mathcal{L}, 0) \leq \frac{C_{\text{dist}}}{C_{\text{safe}}} |\mathcal{L}|$ . Here,  $C_{\text{dist}}$  denotes a constant that only depends on the distance function.

The following theorem is the main technical result of the section. It establishes a bound on the class of equivalence of a single loop stack, thus providing a bound on each  $\text{Cont}([\mathcal{P}_s])$  in (4.3).

**Theorem 5.10.** *Let  $\mathcal{P}$  be an eigenvalue/eigenvector loop stack. Suppose,  $f$  is  $C_{\text{reg}}$ -regular at  $x_0$  and  $\mathcal{L}$  is the base loop of  $\mathcal{P}$ . Define  $M_\beta$  by (5.18). Let*

$$|t| \leq \frac{1}{4} \min_{\mathbf{n} \in \mathcal{L}} \|\mathbf{n} \cdot \omega\|, \quad 0 < \beta < \beta_{\max}(\text{dist}_\varphi, C_{\text{dio}}, \tau, C_{\text{safe}}).$$

Then

$$(1) \quad \begin{aligned} |\text{Cont}([\mathcal{P}], t)| &\leq (C_{\text{dist}} \|\varphi\|_\infty)^{|\mathcal{P}|} \left( \frac{4}{D_{\min}(x_0)} \right)^{\text{den}(\mathcal{P}, M_\beta)} \beta^{-\text{totalllevel}(\mathcal{P}, M_\beta) - \text{den}(\mathcal{P}, M_\beta)} \times \\ &\times C_{\text{reg}}^{\text{downedges}(\mathcal{P}, M_\beta)} \left( \frac{4}{D_{\min}(x_0)} \right)^{\text{nbloops}(\mathcal{P}, M_\beta)} \beta^{-\text{nblevel}(\mathcal{P}, M_\beta) - \text{nbloops}(\mathcal{P}, M_\beta)}. \end{aligned}$$

(2) *The function  $\text{Cont}(\mathcal{P}, \cdot)$  is Lipschitz continuous with the bound on the derivative*

$$\left| \frac{d}{dt} \text{Cont}([\mathcal{P}], t) \right| \leq C_{\text{reg}} \cdot \max \left\{ \frac{4}{\beta^{\text{level}(\mathcal{L})+1}}, D_{\min}(x_0) \right\} \cdot (\text{r.h.s of (1)}).$$

*Proof.* Before starting the proof, we have a few remarks on the structure of the terms. The first line in the right hand side of (1) is the “safe” part of the contribution, similarly to (1) in Lemma 5.8. This is what the estimate would be like if we ignored all extra factors appearing in the attachment points. As in the example following Definition 4.6, the extra attachment factors will be cancelled with small numerators that appear as the result of subtracting a loop contribution term and its translation. Each such cancellation will introduce a derivative factor, similar to the extra factor (2) in Lemma 5.8. We only need to do this if the attachment denominator  $V_{\mathbf{n}_j}(t) - V_0$  is small enough; that is,  $\text{level}(\mathbf{n}_j) \geq M_\beta$  or, equivalently, that we cannot bound  $|V_{\mathbf{n}_j}(t) - V_0|$  by 1 from below. The number of such attachments is equal to  $\text{downedges}(\mathcal{P}, M_\beta)$ . For each of these attachments, we apply the induction assumption (2), which ultimately reduces to part (2) of Lemma 5.8. In this case, there are two possibilities: the level of the attached loop is still  $\geq M_\beta$ , or it becomes less than  $M_\beta$  (note that it must be smaller than  $\text{level}(\mathbf{n}_j)$  due to the shortness condition in the loop stack structure). In both cases, we get  $C_{\text{reg}}$  in the derivative bound, but the rest will depend on which argument of the max function is bigger. The number of times we use the first factor is, ultimately, controlled by  $\text{nblevel}$  and  $\text{nbloops}$ .

An important observation is that, in estimating the derivative factor, one can always use the loop at the lowest level, since it will always contain the smallest denominator. The contribution from differentiating other loops is additive, and can hence be absorbed into a combinatorial factor  $C_{\text{dist}}^{|\mathcal{P}|}$ , where  $C_{\text{dist}}$  only depends on the distance function.

Both claims are proved by induction in  $\text{height}(\mathcal{P})$  at the same time. The case  $\text{height}(\mathcal{P}) = 0$  is contained in Lemma 5.8. Any  $\mathcal{P}$  satisfying the assumptions of the theorem can be split

into the base loop  $\mathcal{L}$  (which is eigenvalue/eigenvector loop, depending on the type of  $\mathcal{P}$ ) and several eigenvalue loop stacks  $\mathcal{P}_j$ , where  $\mathcal{P}_j$  is attached to a vertex  $v_j$  on  $\mathcal{L}$  with coordinate  $\mathbf{n}_j$ , which is a small denominator. Then, since the loop translation operation is equivalent to replacing  $t$  by  $t + \mathbf{n}_j \cdot \omega$ , we have

$$(5.20) \quad \text{Cont}([\mathcal{P}], t) = \text{Cont}(\mathcal{L}, t) \prod_j (V_0 - V_{\mathbf{n}_j}(t))^{-1} (\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)).$$

Clearly,

$$\begin{aligned} \text{level}(\mathcal{P}, M_\beta) &= \text{level}(\mathcal{L}, M_\beta) + \sum_j \text{level}(\mathcal{P}_j, M_\beta), \quad |\mathcal{P}| = |\mathcal{L}| + \sum_j (1 + |\mathcal{P}_j|), \\ \text{loops}(\mathcal{P}) &= \text{loops}(\mathcal{L}) + \sum_j \text{loops}(\mathcal{P}_j), \\ \text{den}(\mathcal{P}, M_\beta) &= \text{den}(\mathcal{L}, M_\beta) + \sum_j \text{den}(\mathcal{P}_j, M_\beta). \\ \text{totalllevel}(\mathcal{P}, M_\beta) &= \text{totalllevel}(\mathcal{L}, M_\beta) + \sum_j \text{totalllevel}(\mathcal{P}_j, M_\beta). \end{aligned}$$

For each  $\mathbf{n}_j$ , calculate the estimate of  $|V_0(x) - V_{\mathbf{n}_j}(x)|^{-1}$  provided by Lemma 5.7. The cases will be based on which of the arguments of the max function in that lemma is larger. If the lemma provides  $|V_0 - V_{\mathbf{n}_j}(t)|^{-1} \leq 1$  (the second argument of max), use the direct estimate

$$(5.21) \quad |\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)| \leq |\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega)| + |\text{Cont}([\mathcal{P}_j], t)|$$

and the induction assumption (1) (we will verify the necessary prerequisite shortly). This means that  $\beta$  is small enough so that the denominator  $(V_0 - V_{\mathbf{n}_j}(t))^{-1}$  does not need to be treated as small, and no derivative factors appear. Each time the estimate is applied, we gain the bound from the previous induction step and an extra factor of 2, which can be absorbed into the combinatorial factor.

In the case  $|V_0 - V_{\mathbf{n}_j}(t)|$  is smaller (meaning that the best that Lemma 5.7 can provide is  $|V_0 - V_{\mathbf{n}_j}(t)|^{-1} \leq B$  with some  $B > 1$ , using the first argument of the max function), we apply the induction assumption (2) and estimate  $|\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)|$  using the derivative of  $\mathcal{P}_j$ .

Let us verify the prerequisites for applying the induction assumptions on the range of  $t$ . Since  $|t| \leq \frac{1}{4} \|\mathbf{n}_j \cdot \omega\|$ , the assumption would follow from

$$(5.22) \quad \|\mathbf{n}_j \cdot \omega\| \leq \frac{1}{8} \min_{\mathbf{n} \in \mathcal{L}_j} \|\mathbf{n} \cdot \omega\|,$$

where  $\mathcal{L}_j$  is the base loop of  $\mathcal{P}_j$ . Recall that the denominator leveling constructed in Theorem 5.4 implies

$$(5.23) \quad \|\mathbf{n}_j \cdot \omega\| \leq 2\beta^{\text{level}(\mathbf{n}_j)}, \quad \|\mathbf{n} \cdot \omega\| \geq \beta^{\text{level}(\mathcal{L}_j)+1},$$

where  $\mathbf{n}$  is from the right hand side of (5.22). In the special case  $\text{level}(\mathbf{n}_j) = 1$ , we have  $|\mathcal{L}_j| < C_{\text{safe}}$ , and the second inequality can be improved to

$$\|\mathbf{n} \cdot \omega\| \geq C_{\text{dio}} |C_{\text{dist}} C_{\text{safe}}|^{-\tau}.$$

Thus, one can achieve (5.22) by choosing a small  $\beta$  (depending on  $C_{\text{safe}}$ ) and applying the first inequality in (5.23).

Let us now consider the case  $\text{level}(\mathbf{n}_j) \geq 2$ . In this case, one can reduce (5.22) to

$$\beta^{\text{level}(\mathbf{n}_j)-1} \leq \frac{1}{16} \beta^{\text{level}(\mathcal{L}_j)}.$$

Recall that  $\mathcal{P}$  is a loop stack, and therefore  $\mathcal{L}_j$  is a short loop:  $|\mathcal{L}_j| < \text{safedist}(\mathbf{n}_j)$ . The bound (5.16) implies

$$(5.24) \quad \text{level}(\mathcal{L}_j) \leq \text{maxlevel}(|\mathcal{L}_j|) \leq \text{maxlevel}(\text{safedist}(\mathbf{n}_j)) \ll \text{level}(\mathbf{n}_j),$$

assuming again that  $\beta_{\max}$  is small enough (depending on  $C_{\text{safe}}$  and other parameters specified in the statement). Thus, in both cases we have (5.22). If we are applying the induction assumption (2) to each  $\mathcal{P}_j$ , we use the bound

$$(5.25) \quad |(V_0 - V_{\mathbf{n}_j}(t))^{-1}| \leq \frac{1}{D_{\min} \|\mathbf{n}_j \cdot \omega\|}.$$

The difference  $|\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)|$  can be estimated by the bound (2) on the derivative, times  $\|\mathbf{n}_j \cdot \omega\|$ . As a consequence, we have

$$(5.26) \quad \left| \frac{\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)}{V_0 - V_{\mathbf{n}_j}(t)} \right| \leq C_{\text{reg}} \max \left\{ \frac{4}{D_{\min} \beta^{\text{level}(\mathcal{L}_j)+1}}, 1 \right\} \cdot (1)_j,$$

where  $(1)_j$  denotes the right hand side of (1) for the induction assumption for  $\mathcal{P}_j$ , and  $\mathcal{L}_j$  is the base loop of  $\mathcal{P}_j$ ; we have also cancelled  $\|\mathbf{n}_j \cdot \omega\|$ . We see that we obtain  $C_{\text{reg}}$  each time the differentiation happens, and we do not gain  $D_{\min}$  in the numerator. The latter is important since, in the high energy region,  $D_{\min}$  can be very large.

To complete the proof of (1), note that the factor (5.26) appears  $\text{downdges}(\mathcal{P}, M_{\beta})$  times. Out of these, the first argument in the max function is chosen  $\text{nbloops}(\mathcal{P}, M_{\beta})$  times, and the second argument appears  $\text{downdges}(\mathcal{P}, M_{\beta}) - \text{nbloops}(\mathcal{P}, M_{\beta})$  times.

To show (2), consider the derivative of (5.20), and suppose that the product runs over  $j = 1, \dots, m$ . As a result of differentiating the product, we obtain  $2m+1$  terms. The estimate from differentiating  $\text{Cont}(\mathcal{L}, t)$  or  $(V_0 - V_{\mathbf{n}_j}(t))^{-1}$  follows from Lemma 5.8 or Lemma 5.7, plus the induction assumption (1) for the remaining factors combined with (5.21).

In the case of differentiating the last factor  $\text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \text{Cont}([\mathcal{P}_j], t)$  in (5.20), we estimate the derivatives of both terms by absolute value:

$$\left| \frac{d}{dt} \text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) - \frac{d}{dt} \text{Cont}([\mathcal{P}_j], t) \right| \leq \left| \frac{d}{dt} \text{Cont}([\mathcal{P}_j], t + \mathbf{n}_j \cdot \omega) \right| + \left| \frac{d}{dt} \text{Cont}([\mathcal{P}_j], t) \right|.$$

Each time we do the last operation, we gain a factor of 2. Ultimately, the power of 2 is equal to the number of attachment points and can therefore be absorbed into  $16^k$ . We also get a factor  $2m+1 < 5k$  each time we apply the induction. The total contribution from the last factors is bounded by  $(5k)^{\text{height}(k)}$ , which can ultimately also be absorbed into  $16^k$  (recall that  $\text{height}(k)$  is an extremely slowly growing function).

□

**5.4. The main result.** Suppose that  $\mathcal{P}$  is a loop stack,  $|\mathcal{P}| = k$ . Then we have the following:

$$\text{den}(\mathcal{P}, M_{\beta}) \leq \frac{k}{\max\{1, C_{\text{dist}} C_{\text{safe}} M_{\beta}^2\}};$$

$$\text{nbloops}(\mathcal{P}, M_{\beta}) \leq \text{loops}(\mathcal{P}, M_{\beta}) \leq \frac{k}{2 \exp\{c(\text{dist}_{\varphi}, C_{\text{dio}}, \tau) \max\{M_{\beta}, 0\}\}};$$

$$\mathbf{nblevel}(\mathcal{P}, M_\beta) \leq \mathbf{totalllevel}(\mathcal{P}, M_\beta) \leq \frac{C_{\text{dist}} k}{C_{\text{safe}} \max\{M_\beta, 1\}}.$$

$$\mathbf{downedges}(\mathcal{P}, M_\beta) \leq k/2.$$

Here, as above,  $C_{\text{dist}}$  denotes some constant that only depends on the distance function and can be different from one estimate to another. Note that the bound on  $\mathbf{downedges}(\mathcal{P}, M_\beta)$  cannot be improved much, since once the base loop of  $\mathcal{P}$  reaches a small denominator of level  $M$ , we can attach an unlimited amount of shortest possible loops (that is, of length 2) to that denominator. Recall now the main bound from Theorem 5.10

$$|\mathbf{Cont}([\mathcal{P}], t)| \leq (C_{\text{dist}} \|\varphi\|_\infty)^k \left( \frac{4}{D_{\min}} \right)^{\mathbf{den}(\mathcal{P}, M_\beta)} \beta^{-\mathbf{totalllevel}(\mathcal{P}, M_\beta) - \mathbf{den}(\mathcal{P}, M_\beta)} \times$$

$$\times C_{\text{reg}}^{\mathbf{downedges}(\mathcal{P}, M_\beta)} \left( \frac{4}{D_{\min}} \right)^{\mathbf{nbloops}(\mathcal{P}, M_\beta)} \beta^{-\mathbf{nblevel}(\mathcal{P}, M_\beta) - \mathbf{nbloops}(\mathcal{P}, M_\beta)}.$$

Combining it with the previous estimates, we obtain (assuming additionally  $D_{\min} \geq 1$ )

$$|\mathbf{Cont}([\mathcal{P}], t)| \leq C_{\text{reg}}^{k/2} (C_{\text{dist}} \|\varphi\|_\infty)^k \beta^{-\frac{3k}{2} \left(1 + \frac{C_{\text{dist}}}{C_{\text{safe}}}\right)}.$$

Recall that these bounds rely on  $\beta$  being sufficiently small in order to satisfy the consistency condition (5.6):

$$0 < \beta < \beta_{\max}(C_{\text{safe}}, C_{\text{dio}}, \text{dist}_\varphi, \tau)$$

and a bound of the form

$$C_{\text{safe}} \mathbf{maxlevel}(k)^3 \ll k,$$

which is required for (5.17) and (5.24) and is possible due to (5.16) (see also Remark 5.6). If  $M_\beta > 0$ , then all the bounds only become better. Note that  $D_{\min}$  can be arbitrarily large, and it is important that it is always in the denominator. We arrive at the following main result.

**Theorem 5.11.** *Suppose,  $f$  is  $C_{\text{reg}}$ -regular at  $x_0 \in (-1/2, 1/2) \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$  and  $D_{\min} \geq 1$ . Consider the operator (1.1):*

$$(H(x_0)\psi)_{\mathbf{n}} = \varepsilon \sum_{\mathbf{m} \in \mathbb{Z}^d} \varphi_{\mathbf{n}-\mathbf{m}} \psi_{\mathbf{m}} + f(x_0 + \mathbf{n} \cdot \omega) \psi_{\mathbf{n}}.$$

Let  $V_{\mathbf{n}} = f(x_0 + \mathbf{n} \cdot \omega)$ , and consider the perturbation series

$$(5.27) \quad \lambda_0 + \varepsilon \lambda_1 + \varepsilon^2 \lambda_2 + \dots,$$

$$\psi = \psi_0 + \varepsilon \psi_1 + \varepsilon^2 \psi_2 + \dots$$

with  $\lambda_0 = f(x)$  and  $\psi_0 = e_0$ , constructed in Theorem 3.2. There exists  $\gamma = \gamma(\tau, C_{\text{dio}}, \text{dist}_\varphi)$  such that the the terms of the eigenvalue series satisfy

$$(5.28) \quad |\lambda_k| \leq C_{\text{reg}}^{k/2} (C_{\text{dist}} \|\varphi\|_\infty)^k \gamma^{-k}.$$

Suppose, in addition, that  $\varepsilon < C_{\text{reg}}^{-1/2} (C_{\text{dist}} \|\varphi\|_\infty)^{-1} \gamma$ , so that the series converges. Then

$$(5.29) \quad |\psi_{\mathbf{n}}| = |\langle e_{\mathbf{n}}, \psi \rangle| \leq \left( \frac{C_{\text{dist}} \|\varphi\|_\infty \sqrt{C_{\text{reg}}}}{\gamma} \right)^{\text{dist}_\varphi(\mathbf{n}, \mathbf{0})} \varepsilon^{\text{dist}_\varphi(\mathbf{n}, \mathbf{0})},$$

so that  $\psi \in \ell^2(\mathbb{Z}^d)$  and it satisfies the eigenvalue equation

$$H(x_0)\psi = \lambda\psi.$$

*Proof.* Recall that the coefficient at  $\varepsilon^k$  at the eigenvalue series is a sum of the following terms:

$$\text{Cont}([\mathcal{P}]) = \text{Cont}(\mathcal{P}_0) \prod_{s=1}^m (V_{\mathbf{n}_s} - V_0)^{-1} \text{Cont}([\mathcal{P}_s]),$$

Each  $\text{Cont}([\mathcal{P}_s])$  factor can be estimated using Theorem 5.10. Due to the Diophantine condition, extra denominators  $V_{\mathbf{n}_s}$  can be absorbed into the bound.

For the eigenvector series, the same argument is sufficient to obtain the convergence of the series for each component of the eigenvector. However, note also that the most “efficient” way to contribute into  $\psi_{\mathbf{n}}$  is to consider a path that only has a safe base loop, since each attachment “resets” the route, and, due to cancellations, we do not need to worry about multiple attachments at small denominators. The bound (5.29) follows from the safe loop bound similar to Lemma 5.8.  $\square$

**Corollary 5.12.** *Suppose that  $f$  is (uniformly)  $C_{\text{reg}}$ -regular on an interval  $(a, b) \subset (-1/2, 1/2)$ . Denote by  $\lambda(x)$ , where  $x \in (a, b) \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ , the result of applying Theorem 5.11 to the operator  $H(x)$ . Then, the function  $\lambda(x)$  extends to a continuous strictly monotone function on  $(a, b)$ .*

*Proof.* We have, from (5.27),

$$\lambda(x) = f(x) + \varepsilon \lambda_1(x) + \varepsilon^2 \lambda_2(x) + \dots$$

The first term is strictly monotone in  $x$  with a positive lower bound on the derivative. The derivatives of the remaining terms can be estimated in a manner similar to (5.28), using part (2) of Theorem 5.10. Therefore, after possibly choosing a smaller  $\varepsilon$ ,  $\lambda(x)$  will have the same lower bound on the derivative (which can be made arbitrarily close to that of  $f$ ).  $\square$

**Corollary 5.13.** *Suppose,  $f$  is  $C_{\text{reg}}$ -regular on  $(-1/2, 1/2)$  and  $D_{\min} \geq 1$ . Let  $x_0 \in (-1/2, 1/2) \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$  and  $\psi[\mathbf{n}]_{\mathbf{m}} := \psi(x_0 + \mathbf{n} \cdot \omega)_{\mathbf{m}}$ . Here,  $\lambda(x)$  is defined in the same way as in Corollary 5.12, and  $\psi(x)$  as in Theorem 5.11. Then*

$$(5.30) \quad H(x_0) \psi[\mathbf{n}] = \lambda_{\mathbf{n}} \psi[\mathbf{n}],$$

where  $\lambda_{\mathbf{n}} = \lambda(x_0 + \mathbf{n} \cdot \omega)$ . The spectrum of  $H(x_0)$  is pure point, and the above eigenvectors form a complete system.

*Proof.* The eigenfunction equation (5.30) follows from a standard translational covariance computation. Consider the operator  $U$  with columns  $\psi[\mathbf{n}]$ . Clearly,

$$\|UU^* - I\| \leq c\varepsilon, \quad \|U^*U - I\| \leq c\varepsilon,$$

where  $c = c(C_{\text{safe}}, C_{\text{dio}}, \text{dist}_{\varphi}, \|\varphi\|_{\infty}, \tau)$ . Therefore, the span of the eigenvectors  $\psi[\mathbf{n}]$  is dense in  $\ell^2(\mathbb{Z}^d)$ . Corollary 5.12 implies that the spectrum is simple.  $\square$

**Remark 5.14.** The inverse function  $\lambda^{-1}(E)$  is equal to  $N(E) - 1/2$ , where  $N(E)$  is the integrated density of states of the ergodic operator family  $H(x)$ . The easiest way to see that is to define IDS as the expectation value of the spectral measure.

**Remark 5.15.** One can easily check that if  $f$  is a meromorphic function in a neighborhood of  $\mathbb{R}$ , then both  $\lambda(x)$  and  $N(E)$  are also meromorphic.

**Remark 5.16.** Under the assumptions of Corollary 5.13, one can normalize the eigenfunctions  $\psi[\mathbf{n}]$  in  $\ell^\infty(\mathbb{Z}^d)$  in order to satisfy *uniform localization*:

$$(5.31) \quad \psi[\mathbf{n}]_{\mathbf{n}} = 1; \quad |\psi[\mathbf{n}]_{\mathbf{m}}| \leq c\varepsilon^{\lfloor \frac{|\mathbf{n}-\mathbf{m}|}{c_{\text{dist}}} \rfloor}, \quad \mathbf{m} \neq \mathbf{n},$$

where  $c$  and  $c_{\text{dist}}$  can be picked uniformly in the region of applicability of Theorem 5.11, that is, whenever  $f$  is  $C_{\text{reg}}$ -regular at  $\{x_0 + \mathbf{n} \cdot \omega\}$ . In other words, each eigenvector is a small perturbation of  $e_{\mathbf{n}}$ , and its components decay exponentially in the distance from the “localization center”  $\mathbf{n}$ , with an upper bound that is uniform for all eigenvectors.

**Remark 5.17.** The fact that  $\lambda(x)$  extends continuously into  $(-1/2, 1/2)$  has the following meaning: suppose that  $x \in \mathbb{Z} + \mathbf{n} \cdot \omega + 1/2$ . Then the potential becomes infinite at  $\mathbf{n}$ . As usual, one can understand it as a Dirichlet boundary condition imposed at  $\mathbf{n}$ . For the usual Schrödinger operator on  $\mathbb{Z}$  with nearest neighbor hopping, the problem thus splits into a direct sum of two half-line operators, each having pure point spectrum.

**5.5. The local result.** Theorem 5.10 only relies on  $C_{\text{reg}}$ -regularity of  $f$  at one point  $x$ . In particular, it does not require monotonicity of  $f$  outside the interval  $(a, b)$  (although it still needs the “regularity at infinity” condition (cr2)). As a consequence, at any such point  $x$  the perturbation series for the eigenvalue and the eigenfunction will converge. If  $f$  is  $C_{\text{reg}}$ -regular on an interval, it implies convergence of the perturbation series for all eigenvalues and eigenfunctions whose (unperturbed) energies fall into the image of that interval. The following results state that, if one slightly decreases the size of the interval, then these eigenfunctions will actually exhaust the spectral projection of  $H$  in the smaller interval. In other words, regularity of  $f$  on an interval implies Anderson localization for  $H$  on the same interval.

**Lemma 5.18.** Suppose that  $f$  is  $C_{\text{reg}}$ -regular on  $(\alpha, \beta)$ , where  $-1/2 < \alpha < \beta < 1/2$  and  $D_{\min} \geq 1$ . Fix  $x_0 \in (-1/2, 1/2) \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$  and apply Theorem 5.11 for all  $x = x_0 + \mathbf{n} \cdot \omega \in (\alpha, \beta)$ . In the notation of Corollary 5.13, denote

$$\Pi_{(\alpha, \beta)} = \overline{\text{span}\{\psi[\mathbf{n}]: \{x_0 + \mathbf{n} \cdot \omega\} \in (\alpha, \beta), \mathbf{n} \in \mathbb{Z}^d\}}.$$

There exist  $c = c(C_{\text{reg}}, C_{\text{dio}}, \text{dist}_\varphi) > 0$  and  $\varepsilon_0 = \varepsilon_0(C_{\text{reg}}, C_{\text{dio}}, \text{dist}_\varphi, \|\varphi\|_\infty)$  such that, for  $0 < \varepsilon < \varepsilon_0$  and all  $\psi \perp \Pi_{(\alpha, \beta)}$ , we have

$$\left\| \left( H(x_0) - \frac{f(\alpha) + f(\beta)}{2} I \right) \psi \right\| \geq \frac{f(\beta) - f(\alpha)}{2} (1 - c\varepsilon) \|\psi\|.$$

*Proof.* Denote by  $E$  the orthogonal projection onto the subspace

$$\text{Ran } E = \overline{\text{span}\{e_{\mathbf{n}}: \{x_0 + \mathbf{n} \cdot \omega\} \in (\alpha, \beta), \mathbf{n} \in \mathbb{Z}^d\}}.$$

Let also  $U: \ell^2(\mathbb{Z}^d) \rightarrow \ell^2(\mathbb{Z}^d)$  be defined on the standard basis by

$$U e_{\mathbf{n}} = \begin{cases} \psi[\mathbf{n}], & \{x_0 + \mathbf{n} \cdot \omega\} \in (\alpha, \beta) \\ 0, & \{x_0 + \mathbf{n} \cdot \omega\} \notin (\alpha, \beta). \end{cases}$$

One can check, using (5.31), that

$$(5.32) \quad \|UU^* - E\| \leq c_1 \varepsilon,$$

where  $c_1$  depends on  $c$  and  $c_{\text{dist}}$  from (5.31). Let  $\psi \perp \Pi_{(\alpha, \beta)}$ . Then  $U^* \psi = 0$ , and (5.32) implies  $\|E\psi\| \leq c_1 \varepsilon \|\psi\|$ . Recall that  $H(x_0) = V(x_0) + \varepsilon \Phi$ , where  $V(x_0)$  is an operator of

multiplication by  $f(x_0 + \mathbf{n} \cdot \omega)$ , and let  $\psi' = E\psi$ ,  $\psi'' = (1 - E)\psi$ . From the definition of  $E$  it follows that

$$\begin{aligned} \left\| \left( V(x_0) - \frac{f(\beta) + f(\alpha)}{2} \right) \psi'' \right\| &\geq \frac{f(\beta) - f(\alpha)}{2} \|\psi''\| \geq \frac{f(\beta) - f(\alpha)}{2} (1 - c_1 \varepsilon) \|\psi\|, \\ \left\| \left( V(x_0) - \frac{f(\beta) + f(\alpha)}{2} \right) \psi' \right\| &\leq \frac{f(\beta) - f(\alpha)}{2} \|\psi'\| \leq \frac{f(\beta) - f(\alpha)}{2} c_1 \varepsilon \|\psi\|. \end{aligned}$$

By combining the previous bounds, we obtain

$$\begin{aligned} &\left\| H(x)\psi - \frac{f(\alpha) + f(\beta)}{2} \psi \right\| \\ &\geq \left\| \left( V(x) - \frac{f(\beta) + f(\alpha)}{2} \right) \psi'' \right\| - \varepsilon \|\Phi\psi\| - \left\| \left( V(x) - \frac{f(\beta) + f(\alpha)}{2} \right) \psi' \right\| \\ &\geq \left( \frac{f(\beta) - f(\alpha)}{2} (1 - c_1 \varepsilon) - \varepsilon \|\Phi\| - \frac{f(\beta) - f(\alpha)}{2} c_1 \varepsilon \right) \|\psi\|, \end{aligned}$$

from which the claim follows.  $\square$

The original definition of  $C_{\text{reg}}$ -regularity may be too restrictive if the interval  $(f(\alpha), f(\beta))$  is small, since it requires monotonicity of  $f$  in a large neighborhood of that interval. It is more convenient to formulate a general result using the rescaled version of regularity in the sense of 5.3:

**Theorem 5.19.** *Suppose that  $f$  is  $C_{\text{reg},\nu}$ -regular on  $(\alpha, \beta)$ , where  $-1/2 < \alpha < \beta < 1/2$  and  $D_{\min}(x) \geq D_0 > 0$  on  $(\alpha, \beta)$ . There exists  $\varepsilon_0 = \varepsilon_0(C_{\text{reg}}, C_{\text{dio}}, \text{dist}_\varphi, \|\varphi\|_\infty, \nu, D_0)$  such that, for  $0 < \varepsilon < \varepsilon_0$  and all  $x_0 \in (-1/2, 1/2) \setminus (\mathbb{Z} + \mathbb{Z}^d \cdot \omega)$ , the spectrum of  $H(x_0)$  in the interval  $[f(\alpha), f(\beta)]$  is pure point, and the eigenfunctions  $\psi[\mathbf{n}]$ , whose energies are in that interval, form a complete system in the range of the spectral projection onto  $[f(\alpha), f(\beta)]$ .*

*Proof.* Since  $f$  is  $C_{\text{reg},\nu}$ -regular on  $(\alpha, \beta)$ , we can assume that  $f$  is  $C_{\text{reg},\nu/2}$ -regular in a slightly larger interval  $(\alpha', \beta')$ . Now the result follows from applying Lemma 5.18 to the rescaled function  $(\nu/2)^{-1}f$  and choosing an appropriately small  $\varepsilon$ . The case of arbitrary  $D_0$  can also be treated by a similar rescaling.  $\square$

## 6. GENERALIZATIONS: THE LONG-RANGE CASE AND NON-CONSTANT HOPPING TERMS

**6.1. Non-constant long range hopping terms.** A *quasiperiodic hopping matrix* is, by definition, a matrix with elements of the following form

$$(6.1) \quad \Phi_{\mathbf{m}\mathbf{n}}(x) = \varphi_{\mathbf{m}-\mathbf{n}}(x + (\mathbf{m} + \mathbf{n}) \cdot \omega/2), \quad \mathbf{m}, \mathbf{n} \in \mathbb{Z}^d,$$

where  $\varphi_{\mathbf{m}}: \mathbb{R} \rightarrow \mathbb{C}$  are Lipschitz 1-periodic functions, satisfying the self-adjointness condition:

$$\varphi_{\mathbf{m}} = \overline{\varphi_{-\mathbf{m}}}.$$

Let also

$$\|\varphi\|_\infty = \sup_{\mathbf{k}} \|\varphi_{\mathbf{k}}\|_\infty, \quad \|\varphi'\|_\infty = \sup_{\mathbf{k}} \|\varphi'_{\mathbf{k}}\|_\infty.$$

Define  $\text{Range}(\Phi)$  to be the smallest number  $L \geq 0$  such that  $\Phi_{\mathbf{n}\mathbf{m}} \equiv 0$  for  $|\mathbf{m} - \mathbf{n}| > L$ . We will only consider hopping matrices of finite range. Note that (6.1) can be reformulated as the following covariance property:

$$\Phi_{\mathbf{m}+\mathbf{a}, \mathbf{n}+\mathbf{a}}(x) = \Phi_{\mathbf{m}\mathbf{n}}(x + \mathbf{a} \cdot \omega), \quad \mathbf{m}, \mathbf{n}, \mathbf{a} \in \mathbb{Z}^d.$$

Fix some  $R \in \mathbb{N}$ , and suppose that  $\Phi^1, \Phi^2, \dots$  is a family of quasiperiodic hopping matrices with  $\text{Range}(\Phi_k) \leq kR$ , defined by a family of functions  $\varphi_{\mathbf{m}}^1, \varphi_{\mathbf{m}}^2, \dots$ . A more general class of operators we would like to consider will be of the following form:

$$(6.2) \quad H = V + \varepsilon\Phi^1 + \varepsilon^2\Phi^2 + \dots, \quad 0 \leq \varepsilon < 1,$$

where, as previously,

$$(V\psi)_{\mathbf{n}} = V_{\mathbf{n}}\psi_{\mathbf{n}} = f(x_0 + \mathbf{n} \cdot \omega)\psi_{\mathbf{n}}.$$

One can easily check that, assuming

$$\|\varphi\|_{\infty} = \sup_j \|\varphi^j\|_{\infty} = \sup_{j, \mathbf{m}} \|\varphi_{\mathbf{m}}^j\|_{\infty} < +\infty, \quad 0 \leq \varepsilon < 1,$$

the part  $\varepsilon\Phi^1 + \varepsilon^2\Phi^2 + \dots$  defines a bounded operator on  $\ell^2(\mathbb{Z}^d)$ . Similarly to the previous sections, we are dealing with the equations

$$(V + \varepsilon\Phi^1 + \varepsilon^2\Phi^2 + \dots)(\psi_0 + \varepsilon\psi_1 + \dots) = (\lambda_0 + \varepsilon\lambda_1 + \dots)(\psi_0 + \varepsilon\psi_1 + \dots),$$

and the initial conditions

$$\lambda_0 = V_{\mathbf{0}} = f(x_0), \quad \psi_0 = e_{\mathbf{0}}, \quad \psi_j \perp \psi_0 \text{ for } j > 0.$$

Formal equalizing of terms at  $\varepsilon^k$  leads to

$$V\psi_k + \Phi^1\psi_{k-1} + \dots + \Phi^{k-1}\psi_1 + \Phi^k e_{\mathbf{0}} = \lambda_0\psi_k + \lambda_1\psi_{k-1} + \lambda_2\psi_{k-2} + \dots + \lambda_{k-1}\psi_1 + \lambda_k e_{\mathbf{0}}$$

which, after projecting onto  $\text{span}\{e_{\mathbf{0}}\}$  and  $\text{span}\{e_{\mathbf{0}}\}^{\perp}$ , is reduced to

$$(6.3) \quad \psi_k = (V_{\mathbf{0}} - V)^{-1} (\Phi^1\psi_{k-1} + \dots + \Phi^{k-1}\psi_1 + \Phi^k e_{\mathbf{0}} - \lambda_1\psi_{k-1} - \lambda_2\psi_{k-2} - \dots - \lambda_{k-1}\psi_1);$$

$$(6.4) \quad \lambda_k = \langle \Phi^1\psi_{k-1} + \Phi^2\psi_{k-2} + \dots + \Phi^{k-1}\psi_1, e_{\mathbf{0}} \rangle.$$

In order to establish an analogue of Theorem 3.2, one needs to construct an appropriate version of the graph  $\Gamma$ . It will formalize the following construction: each time a path makes a jump between two vertices, it can “choose” which term of the kinetic energy will be used to perform that jump. Naturally, the use of the term  $\Phi^j$  will cost  $\varepsilon^j$ ; in other words, the corresponding graph edge will contribute  $j$  to the length of the path. In the construction from Section 3, all edges had length 1.

- Take  $\Phi = \Phi^j$  and denote the graph constructed in Section 3.1 using only that  $\Phi$ , by  $\Gamma_j = (\mathcal{E}, \mathcal{V}_j)$ ; here  $\mathcal{E}$  is the set of vertices (same for each  $\Gamma_j$ ), and  $\mathcal{V}_j$  is the set of edges. Let

$$\Gamma := (\mathcal{E}, \bigsqcup_j \mathcal{V}_j),$$

where  $\bigsqcup$  denotes the disjoint union. In other words,  $\Gamma$  is a graph on the same set of vertices as each of the  $\Gamma_j$ , and includes all edges of each  $\Gamma_j$ , considered separately (thus obtaining a multi-graph). To each edge of  $\Gamma$ , we associate two numbers: the length (an integer number  $j$  for an edge inherited from  $\Gamma^j$ ) and the weight (defined in Section 3.1 with  $\Phi = \Phi^j$ ). We will also be following the convention from Remark 3.1 for future convenience.

- As a result, a path in the graph  $\Gamma$  is no longer determined by a sequence of vertices, and requires additional data: for each jump (by a jump we will mean moving from one vertex in our sequence to the next one), one needs to specify the length of the edge used for that jump. Any edge of  $\Gamma$  is uniquely determined by its length and starting and ending vertices.

- The loop translation is now defined in a way that, whenever a piece of a path gets translated, every edge gets translated into an edge of the same length as it was before translation. The convention in Remark 3.1 guarantees that the translated path will exist. However, since  $\Phi_{\mathbf{m}\mathbf{n}}^j$  depends on  $x$  and is no longer translation invariant, a path with a non-zero contribution may be translated into a path with zero contribution.

If  $\mathcal{P}$  is an eigenvalue/eigenvector path on  $\Gamma$ , we will denote by  $|\mathcal{P}|$  the total length of edges of  $\mathcal{P}$ . With the above conventions in place, one can reformulate the main results of Section 2:

$$\lambda_s = \sum_{\mathcal{P}, |\mathcal{P}|=s} \text{Cont}(\mathcal{P}),$$

where the sum is considered over all eigenvalue loop configurations  $\mathcal{P}$  (with  $|\mathcal{P}| = s$ ), and a similar expression for the eigenvector

$$(\psi_s)_{\mathbf{k}} = \sum_{\mathcal{P}, |\mathcal{P}|=s} \text{Cont}(\mathcal{P}), \quad \mathbf{k} \neq 0, \quad s > 0;$$

The definition of a safe loop follows Definition 4.2, with the addition that the number of jumps is replaced by the total length of the edges. In the definition of  $\text{Cont}(\mathcal{P}, t)$ , in addition to replacing  $V_{\mathbf{n}} = f(x_0 + \mathbf{n} \cdot \omega)$  by  $V_{\mathbf{n}}(t) = f(x_0 + t + \mathbf{n} \cdot \omega)$  for  $\mathbf{n} \neq \mathbf{0}$ , we also replace  $\Phi_{\mathbf{m}\mathbf{n}}^j(x_0)$  by

$$\Phi_{\mathbf{m}\mathbf{n}}^j(x_0, t) := \Phi_{\mathbf{m}\mathbf{n}}^j(x_0 + t) = \varphi_{\mathbf{m}-\mathbf{n}}(x_0 + t + (\mathbf{m} + \mathbf{n})\omega/2).$$

The results of Section 4 can be repeated without significant changes. It is helpful to observe the following: one can treat a jump from  $\mathbf{m}$  to  $\mathbf{n}$ , say, with  $\text{dist}_{\Phi}(\mathbf{m}, \mathbf{n}) = 3$ , as three jumps of distance 1, where only one of these jumps makes an actual contribution to  $\text{Cont}(\mathcal{P})$ , and the contributions of two other jumps are equal to 1. The resulting bounds are only better (since there are less denominators to care about). Since  $\varphi$  is no longer a constant, each time we are estimating a derivative we have a choice of differentiating one of the  $\varphi$  factors. As a consequence, the factor  $\|\varphi\|_{\infty}$  in front has to be replaced by  $\|\varphi\|_{\infty} + \|\varphi'\|_{\infty}$ . An extra combinatorial factor can be absorbed into  $C_{\text{dist}}^k$ .

**Proposition 6.1.** *Let  $\mathcal{P}$  be an eigenvalue/eigenvector loop stack with  $k$  edges. Suppose,  $f$  is  $C_{\text{reg}}$ -regular at  $x_0$  and  $\mathcal{L}$  is the base loop of  $\mathcal{P}$ . Denote  $M_{\beta}$  by (5.18), and let*

$$|t| \leq \frac{1}{4} \min_{\mathbf{n} \in \mathcal{L}} \|\mathbf{n} \cdot \omega\|, \quad 0 < \beta < \beta_{\max}(\text{dist}_{\varphi}, C_{\text{dio}}, \tau_{\text{dio}}, C_{\text{safe}}).$$

Then

$$(1) \quad \begin{aligned} |\text{Cont}([\mathcal{P}], t)| &\leq C_R^{|\mathcal{P}|} (\|\varphi\|_{\infty} + \|\varphi'\|_{\infty})^k \left( \frac{4\gamma}{D_{\min}} \right)^{\text{den}(\mathcal{P}, M_{\beta})} \beta^{-\text{totallevel}(\mathcal{P}, M_{\beta}) - \text{den}(\mathcal{P}, M_{\beta})} \times \\ &\times C_{\text{reg}}^{\text{downedges}(\mathcal{P}, M_{\beta})} \left( \frac{4}{D_{\min}} \right)^{\text{nbloops}(\mathcal{P}, M_{\beta})} \beta^{-\text{nlevel}(\mathcal{P}, M_{\beta}) - \text{nbloops}(\mathcal{P}, M_{\beta})}. \end{aligned}$$

(2) *The function  $\text{Cont}([\mathcal{P}], \cdot)$  is Lipschitz continuous with the bound on the derivative*

$$\left| \frac{d}{dt} \text{Cont}([\mathcal{P}], t) \right| \leq C_{\text{reg}} \max \left\{ \frac{4\gamma}{\varepsilon^{\text{level}(\mathcal{L}) + \delta_{\text{lev}}}}, D_{\min} \right\} \cdot (\text{r.h.s of (1)}).$$

**Remark 6.2.** The combinatorial factor now depends only on  $R$ , since the distance function and the range of  $\Phi^j$  is defined by  $R$ . Note that, despite multiple choice of edges, the total number of paths of length  $|\mathcal{P}|$  with non-zero contributions is still bounded by  $C_R^{|\mathcal{P}|}$ . The power of  $\|\varphi\|_\infty$  now depends only on  $k$ , which is the number of edges. The fact that one also needs to differentiate  $\varphi$  only affects the result by adding  $\|\varphi'\|_\infty$  into the factor  $(\|\varphi\|_\infty + \|\varphi'\|_\infty)$ , since differentiating  $\varphi$  means that we are using  $\|\varphi'\|_\infty \cdot (\text{r.h.s of (1)})$  instead of  $(\text{r.h.s of (2)})$ . These possible extra  $\|\varphi'\|_\infty$  can be absorbed into the factors in front of (1).

**Remark 6.3.** The reader can easily formulate appropriate versions of the results of Section 4, starting from Theorem 5.11. The only differences are the inclusion of  $\|\varphi'\|_\infty$  and different notation for the distance function.

**6.2. The case of monotone potentials with small derivatives.** The results of Section 6.1 can easily be extended to the case where the terms  $\Phi^j$  of (6.2) have some additional dependence on  $\varepsilon$ , as long as the estimates of  $\|\varphi\|_\infty$  and  $\|\varphi'\|_\infty$  are uniform in  $\varepsilon$  in the range of parameters under consideration. The same arguments can be applied to the dependence of  $f$  on  $\varepsilon$  if one considers both upper and lower bounds on the derivatives described in the definitions of regularity in Section 5.1. In particular, we always assume that  $D_{\min} \geq 1$  (although rescaling allows to replace 1 by any positive  $\varepsilon$ -independent quantity). In some of the applications, these assumptions are not flexible enough. In particular, the operators considered in Section 7, for which  $f$  is originally not strictly monotone, can be transformed into those with monotone  $f$ , but the derivative will only admit a lower bound by  $c\varepsilon^2$  on some intervals. The bounds outlined above, essentially, contain a factor  $D_{\min}^{-k}$ , leading to a coefficient of the order  $\varepsilon^{-2k}$  at  $\varepsilon^k$ , not sufficient to guarantee convergence of the series for such  $f$ . In addition to that,  $C_{\text{reg}}$  will also contain negative powers of  $\varepsilon$ .

In order to deal with these problems, we introduce some refinements to our general scheme which we will now describe. They will involve additional restrictions on the hopping terms and  $f$ . Assume the following:

- (sing0)  $I_1, \dots, I_k \subset (-1/2, 1/2)$  are open intervals whose closures are mutually disjoint.
- (sing1)  $f$  satisfies (f1) from Section 3. Additionally,  $f$  is  $C_{\text{reg}}$ -regular on  $(-1/2, 1/2) \setminus (I_1 \cup \dots \cup I_k)$  with  $D_{\min}(x_0) \geq 1$  for all  $x_0 \in (-1/2, 1/2) \setminus (I_1 \cup \dots \cup I_k)$ . For each  $I_j = (a_j, b_j)$ , let  $D_{j,+} := \max\{D_{\min}(a_j), D_{\min}(b_j)\}$ ,  $D_{j,-} := \min\{D_{\min}(a_j), D_{\min}(b_j)\}$ .
- (sing2) For some  $\delta > 0$ , we have  $c_{j,-}\delta^{\mu_j} \leq f'(x) \leq c_{j,+}$  on  $I_j$ . As usual, we assume that the inequalities hold for all derivative numbers in the case  $f$  is not differentiable. As a consequence,  $f$  is strictly monotone on  $(-1/2, 1/2)$ .
- (sing3) For each  $x_0 \in I_j$ , we have (cr2) with  $D_{\min}(x_0)$  replaced by  $c_{j,+}$ .

Since  $f$  is regular on  $(-1/2, 1/2) \setminus (I_1 \cup \dots \cup I_k)$ , the results of the previous subsection apply for  $x_0 \in (-1/2, 1/2) \setminus (I_1 \cup \dots \cup I_k)$  outside of these intervals.

If  $x_0 \in I_j$ , we can estimate the denominators by considering three cases:

- (1) *Singular* small denominators:  $t+x_0 \in I_j$ ,  $t+x_0+\mathbf{n} \cdot \omega \in I_j$  for all  $t$  under consideration. In this case, we have to use  $c_{j,-}\delta^{\mu_j}$  as the lower bound for the derivative of  $f$ , and can use  $c_{j,+}$  as an upper bound.
- (2) *Regular* small denominators:  $x_0 \in I_j$ , but either  $t+x_0$  or  $t+x_0+\mathbf{n} \cdot \omega$  is not in  $I_j$ , for some  $t$  under consideration. In this case, while  $f$  is not regular at  $x_0$ , we can treat  $x_0$  as a small perturbation of  $t+x_0$  or  $t+x_0+\mathbf{n} \cdot \omega$ . These denominators can be considered in the same way as the regular case, under some modifications of the constants: since some of the points are outside of  $I_j$ , one cannot use  $c_{j,+}$  as an upper

bound on the derivative but can use  $C_{\text{reg}} D_{j,+}$  as an upper bound and  $D_{j,-}$  as a lower bound.

(3) For non-small denominators, in view of (sing3), one can use  $C_{\text{reg}} c_{j,+}$  as an upper bound on the derivative; lower bounds are not needed since the absolute value of the denominator is at least 1.

For simplicity, let us also assume that  $\beta$  is chosen to be small enough so that all denominators of level 1 or higher are actually small. In other words,  $M_\beta(x_0) = 0$  for all  $x_0 \in I_j$ . We will drop  $M_\beta$  from the notation: for example,  $\text{den}(\mathcal{L})$  will denote  $\text{den}(\mathcal{L}, 0)$ . For a loop  $\mathcal{L}$ , denote by  $\text{singden}(\mathcal{L})$  the total number of singular small denominators visited by  $\mathcal{L}$ .

**Lemma 6.4.** *Under the above assumptions, let  $\mathcal{L}$  be an eigenvalue/eigenvector loop with  $k$  edges. Let*

$$x_0 \in I_j, \quad |t| \leq \frac{1}{4} \min_{\mathbf{n} \in \mathcal{L}} \|\mathbf{n} \cdot \omega\|.$$

Then

$$(1) \quad |\text{Cont}(\mathcal{L}, t)| \leq \|\varphi\|_\infty^k \left( \frac{4}{D_{j,-}} \right)^{\text{den}(\mathcal{L})} \beta^{-\text{totallevel}(\mathcal{L}) - \text{den}(\mathcal{L})} \left( \frac{D_{j,-}}{\delta^{\mu_j} c_{j,-}} \right)^{\text{singden}(\mathcal{L})}.$$

(2) *The function  $\text{Cont}(\mathcal{L}, \cdot)$  is Lipschitz continuous:*

$$\left| \frac{d}{dt} \text{Cont}(\mathcal{L}, t) \right| \leq |\mathcal{L}| \cdot \max \left\{ \frac{4\delta^{-\mu_j} c_{j,+}}{c_{j,-} \beta^{\text{level}(\mathcal{L})+1}}, \frac{4C_{\text{reg}} D_{j,+}}{D_{j,-} \beta^{\text{level}(\mathcal{L})+1}}, C_{\text{reg}} c_{j,+}, \right\} \cdot (\text{r.h.s of (1)}).$$

*Proof.* As described above, we consider three types of denominators. In the case of estimating (1), non-small denominators do not give any contribution (we use a lower bound by 1 for each of them). The “regular” small denominators are treated in the same way as in the regular case, with the same contributions. In the case of singular small denominators, we replace  $D_{j,-}$  by  $(\delta^{c_{j,-}})^{\mu_j}$  which produces the last factor in (1) (“replacement factor”).

In the case of (2), we argue similarly to Lemma 5.8; that is, a differentiation would add an extra factor  $|\mathcal{L}| \left| \frac{V'_n(t)}{V_0 - V_n(t)} \right|$  in the upper bound, and we estimate the ratio in the worst possible case, again, by considering the three types of denominators.  $\square$

In order to make further estimates less cumbersome, we assume that  $\beta$  and  $\delta$  are small enough, depending on  $C_{\text{reg}}$ ,  $D_{j,\pm}$ ,  $c_{j,\pm}$ , so that in the conclusion (2) of Lemma 6.4 we have

$$(6.5) \quad \frac{4\delta^{-\mu_j} c_{j,+}}{c_{j,-} \beta^{\text{level}(\mathcal{L})+1}} \geq \frac{4C_{\text{reg}} D_{j,+}}{D_{j,-} \beta^{\text{level}(\mathcal{L})+1}} \geq C_{\text{reg}} c_{j,+}.$$

It will be convenient to introduce some additional notation.

- Let  $\text{singdownedges}(\mathcal{P})$  denote the number of descending edges of  $\mathcal{P}$  which lead to a singular small denominator.
- Let also  $\text{singden}(\mathcal{P})$  be the total number of singular small denominators visited by  $\mathcal{P}$ .

**Theorem 6.5.** *Under the above assumptions on  $f$  and  $\Phi$ , let  $\mathcal{P}$  be an eigenvalue/eigenvector loop stack with  $k$  edges. Suppose that  $x_0 \in I_j$  and  $\mathcal{L}$  is the base loop of  $\mathcal{P}$ . Let*

$$|t| \leq \frac{1}{4} \min_{\mathbf{n} \in \mathcal{L}} \|\mathbf{n} \cdot \omega\|, \quad 0 < \beta < c(\text{dist}_\varphi, C_{\text{dio}}, \tau_{\text{dio}}, C_{\text{safe}}, c_{j,\pm}, D_{j,\pm}, C_{\text{reg}})$$

$$0 < \delta < \delta_0(\text{dist}_\varphi, C_{\text{dio}}, \tau_{\text{dio}}, C_{\text{safe}}, c_{j,\pm}, D_{j,\pm}, C_{\text{reg}}).$$

Then we have

$$|\text{Cont}([\mathcal{P}], t)| \leq C_R^{|\mathcal{P}|} (\|\varphi\|_\infty + \|\varphi'\|_\infty)^k \left( \frac{4}{D_{j,-}} \right)^{\text{den}(\mathcal{P})} \beta^{-\text{totalevel}(\mathcal{P}) - \text{den}(\mathcal{P})} \times \left( \frac{D_{j,-}}{\delta^{\mu_j} c_{j,-}} \right)^{\text{singden}(\mathcal{P})} \times \\ \times \left( \frac{4C_{\text{reg}} D_{j,+}}{D_{j,-}^2} \right)^{\text{downdedges}(\mathcal{P})} \beta^{-\text{nblevel}(\mathcal{P}) - \text{downdedges}(\mathcal{P})} \left( \frac{c_{j,+} D_{j,-}^2}{\delta^{2\mu_j} c_{j,-}^2 D_{j,+}} \right)^{\text{singdowndedges}(\mathcal{P})}.$$

*Proof.* The proof essentially repeats the proof of Theorem 5.10, with modified constants. The first factor comes from considering the individual loops on  $\mathcal{P}$  and applying Lemma 6.4 to each of them.

The remaining factor comes from considering the attachment factors and the differentiation factors. For example, in the factor  $C_{\text{reg}}^{\text{downdedges}(\mathcal{P}, M_\beta)} \left( \frac{4}{D_{\min(x_0)}} \right)^{\text{nbloops}(\mathcal{P}, M_\beta)}$  from Theorem 5.10, we note that  $\text{downdedges}(\mathcal{P}) = \text{nbloops}(\mathcal{P})$ , replace  $D_{\min}(x_0)$  by  $D_{j,-}$  and  $C_{\text{reg}}$  by  $\frac{C_{\text{reg}} D_{j,+}}{D_{j,-}^2}$ , thus obtaining the factor  $\left( \frac{4C_{\text{reg}} D_{j,+}}{D_{j,-}^2} \right)^{\text{downdedges}(\mathcal{P})}$ . Together with  $\beta^{-\text{nblevel}(\mathcal{P}) - \text{nbloops}(\mathcal{P})}$ , this would complete the bounds in the case of no singular small denominators.

For each singular small denominators, instead of the above procedure, we need to replace  $C_{\text{reg}}$  by  $\frac{c_{j,+}}{c_{j,-} \delta^{\mu_j}}$ , and  $D_{\min}$  by  $c_{j,-} \delta^{\mu_j}$ , which produces the last replacement factor.  $\square$

In order to formulate the main result of this section, let us state an additional condition on  $f$ . We would like to emphasize that  $\delta$  is allowed to depend on  $\varepsilon$ .

(sing4) There exists  $\beta_{\max} = \beta_{\max}(C_{\text{reg}}, I_j, D_{j,\pm}, c_{j,\pm}, \omega) > 0$  such that, for  $0 < \beta < \beta_{\max}$  there exist

$$\varepsilon_0 = \varepsilon_0(\beta, C_{\text{reg}}, I_j, D_{j,\pm}, c_{j,\pm}, \omega) > 0, \quad 0 < r = r(\beta, C_{\text{reg}}, I_j, D_{j,\pm}, c_{j,\pm}, \omega) < 1,$$

$$C = C(\beta, C_{\text{reg}}, I_j, D_{j,\pm}, c_{j,\pm}, \omega) > 0$$

such that, for any loop stack  $\mathcal{P}$  and for any  $j$  we have

$$\delta^{\mu_j \text{singden}(\mathcal{P}) + 2\mu_j \text{singdowndedges}(\mathcal{P})} \leq C^{|\mathcal{P}|} \varepsilon^{r|\mathcal{P}|}.$$

We can now state the last result of this section. Its proof, after the above preparations, repeats the corresponding steps of the proof of Theorem 5.11.

**Theorem 6.6.** *Suppose that  $f$  satisfies (sing0)–(sing3) with  $C_{\text{reg}}$ ,  $D_{j,\pm}$ ,  $c_{j,\pm}$ ,  $I_j$  independent of  $\varepsilon$ . Under the assumptions of (sing4), and further assuming*

$$0 < \varepsilon < \varepsilon_1(\beta, C_{\text{reg}}, I_j, D_{j,\pm}, c_{j,\pm}, \omega, C, c, \varepsilon_0, r),$$

*the perturbation series for  $H(x_0)$  converges for all  $x_0 \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ .*

## 7. AN EXAMPLE: MONOTONE POTENTIAL WITH A FLAT SEGMENT

**7.1. The class of operators.** Fix an interval  $(a - h, a + h) \subset (-1/2, 1/2)$ , where

$$-1/2 < a - h < a < a + h < 1/2.$$

Consider a continuous non-decreasing function  $f: (-1/2, 1/2) \rightarrow \mathbb{R}$  and a frequency vector  $\omega$  with the following properties:

- $f(-1/2 + 0) = -\infty$ ,  $f(1/2 - 0) = +\infty$ .
- $f$  is  $C_{\text{reg}}$ -regular uniformly for  $x \in (-1/2, a - h) \cup (a + h, 1/2)$  for some  $C_{\text{reg}} > 0$ .

- $f'(x) \geq 1$  for all  $x \in (-1/2, a-h) \cup (a+h, 1/2)$  and  $0 \leq f'(x) \leq c_+$  for all  $x \in [a-2h, a+2h]$  for some  $c_+ > 0$ .
- $\omega \in (-1/2, 1/2)^d$  is a Diophantine frequency vector. In addition,  $\|\mathbf{n} \cdot \omega\| \geq 6h$ , for all  $\mathbf{n}$  with  $1 \leq |\mathbf{n}|_\infty \leq c_1$ . Note that this imposes an additional requirement on the smallness of  $h$ . Here,  $c_1$  is a sufficiently large absolute constant ( $c_1 = 6$  is sufficient for the main result of this section; however, we did not intend to optimize it and believe that a small modification would allow to use  $c_1 = 3$ ).

**Remark 7.1.** This class of functions covers the main example that we have in mind: that is, the case when  $f$  is equal to a constant on some sub-interval in  $(a-h_1, a+h_1) \subset (a-h, a+h)$ . Note that  $C_{\text{reg}}$ -regularity outside  $(a-h, a+h)$  requires  $f$  to be strictly monotone near  $a-h$  and  $a+h$ , which implies that the inclusion must be strict. However, by introducing additional rescaling that depends on  $h-h_1$ , one can consider examples with arbitrarily small values of  $h-h_1$ .

**7.2. Outline of the procedure.** Let us consider the following operator family on  $\ell^2(\mathbb{Z}^d)$ :

$$(7.1) \quad (H(x)\psi)_{\mathbf{n}} = \varepsilon(\Delta\psi)_{\mathbf{n}} + f(x + \mathbf{n} \cdot \omega)\psi_{\mathbf{n}},$$

where  $x \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ , and  $f$  is extended into  $\mathbb{R} \setminus (\mathbb{Z} + 1/2)$  by 1-periodicity. Operator (7.1) does not satisfy the assumptions of Theorem 5.11, as  $f$  is not strictly monotone. However, it is possible to replace it by a unitarily equivalent operator with non-constant hopping terms which will satisfy the criteria described in Section 6. We will describe the corresponding unitary transformations in a general form. Suppose that

$$(H\psi)_{\mathbf{m}} = V_{\mathbf{m}}\psi_{\mathbf{m}} + \varepsilon \sum_{\mathbf{m}' \in \mathbb{Z}^d} \Phi_{\mathbf{m}\mathbf{m}'}\psi_{\mathbf{m}'}.$$

To begin with, we will define the basic ‘‘building block’’ that will be used in the construction of the required unitary transformation. Let  $\mathbf{n} \neq \mathbf{0}$ . Suppose that  $V_{\mathbf{0}} = H_{\mathbf{00}} = 0$ ,  $V_{\mathbf{n}} = H_{\mathbf{nn}} > 0$ , and

$$H_{\mathbf{0n}} = \overline{H_{\mathbf{n0}}} = \varepsilon\Phi_{\mathbf{0n}} = \varepsilon\overline{\Phi_{\mathbf{n0}}}.$$

Put

$$D := \sqrt{V_{\mathbf{n}}^2 + \varepsilon^2|\Phi_{\mathbf{0n}}|^2}.$$

Consider the operator  $U$  with the following matrix elements:

$$\begin{aligned} U_{\mathbf{00}} &= U_{\mathbf{nn}} = \frac{V_{\mathbf{n}}}{D}; \\ U_{\mathbf{0n}} &= \varepsilon \frac{\overline{\Phi_{\mathbf{0n}}}}{D}, \quad U_{\mathbf{n0}} = -\varepsilon \frac{\Phi_{\mathbf{n0}}}{D} = -\varepsilon \frac{\overline{\Phi_{\mathbf{0n}}}}{D}; \\ U_{\mathbf{mm}} &= 1 \text{ for } \mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}, \mathbf{n}\}; \end{aligned}$$

and all other matrix elements of  $U$  are zero. One can check that  $U$  is a unitary operator on  $\ell^2(\mathbb{Z}^d)$ . Let us calculate the conjugation of  $H$  by  $U$ :

$$U^*HU = \tilde{V} + \varepsilon\tilde{\Phi},$$

where

$$\begin{aligned} \tilde{V}_{\mathbf{0}} &= -\varepsilon^2 \frac{V_{\mathbf{n}}|\Phi_{\mathbf{0n}}|^2}{D^2}, \quad \tilde{V}_{\mathbf{n}} = \frac{V_{\mathbf{n}}^3 + 2\varepsilon^2|\Phi_{\mathbf{0n}}|^2V_{\mathbf{n}}}{D^2}, \quad \tilde{V}_{\mathbf{m}} = V_{\mathbf{m}} \text{ for } \mathbf{m} \in \mathbb{Z}^d \setminus \{\mathbf{0}, \mathbf{n}\}; \\ \tilde{\Phi}_{\mathbf{0n}} &= -\varepsilon^2 \frac{|\Phi_{\mathbf{0n}}|^2\Phi_{\mathbf{n0}}}{D^2}, \quad \tilde{\Phi}_{\mathbf{n0}} = -\varepsilon^2 \frac{|\Phi_{\mathbf{0n}}|^2\Phi_{\mathbf{0n}}}{D^2}, \end{aligned}$$

$$\tilde{\Phi}_{0m} = \frac{V_n}{D} \Phi_{0m} + \varepsilon \frac{\Phi_{n0}}{D} \Phi_{nm}, \quad \tilde{\Phi}_{nm} = \frac{V_n}{D} \Phi_{nm} - \varepsilon \frac{\Phi_{0n}}{D} \Phi_{0m} \text{ for } m \in \mathbb{Z}^d \setminus \{\mathbf{0}, \mathbf{n}\},$$

and  $\tilde{\Phi}_{m_1 m_2} = \Phi_{m_1 m_2}$  for  $\mathbf{m}_1, \mathbf{m}_2 \in \mathbb{Z}^d \setminus \{\mathbf{0}, \mathbf{n}\}$ .

**Remark.** Clearly, we can apply this transform even if  $V_0 \neq 0$ , as long as  $V_n \neq V_0$ : we represent  $H$  as  $H = V_0 I + \hat{H}$  and transform  $\hat{H}$  using the procedure described above. Similarly, one can apply the procedure to any pair of lattice points  $\{\mathbf{k}, \mathbf{k} + \mathbf{n}\}$  through adding  $\mathbf{k}$  to all indices in the above procedure: in particular,  $D = \sqrt{(V_{n+k} - V_k)^2 + \varepsilon^2 |\Phi_{k,n+k}|^2}$ .

This procedure will be called *elimination of the entry*  $\Phi_{0n}$ . Note that, strictly speaking, it only eliminates it up to the order  $\varepsilon^2$ . As a result, new entries can also be created, such as  $\tilde{\Phi}_{mn}$  in the above calculations. Using translation, one can describe elimination of any entry  $\Phi_{k,k+n}$  under similar assumptions.

It is natural to consider the elimination procedure in the context of long range operators from Section 7.1. For example, if one starts from an operator  $H = V + \varepsilon \Phi^1$ , the result of the transformation can be represented as

$$U^* H U = \tilde{V} + \varepsilon \tilde{\Phi}^1 + \varepsilon^2 \tilde{\Phi}^2 + \varepsilon^3 \tilde{\Phi}^3.$$

Here, we will ignore the dependence of  $D$  on  $\varepsilon$  and group the correction terms based on the number of  $\varepsilon$  factors in front of them: for example, in the earlier notation,  $\tilde{\Phi}_{n0}^3 = -\frac{|\Phi_{0n}|^2 \Phi_{0n}}{D^2}$ . In application, the dependence on  $\varepsilon$  through  $D$  will be slow enough so that the upper required bounds on the off-diagonal terms and their derivatives will be uniform in  $\varepsilon$ . On the other hand, one will need to deal more carefully with the diagonal terms, since the presence of the  $\varepsilon^2$  terms will be crucial for establishing monotonicity.

In this new notation, one can state that the procedure completely eliminates  $\Phi_{0n}^1$ , since the correction to this term goes into  $\tilde{\Phi}^3$ . One can repeat this procedure, say, for the pair of lattice points  $\{\mathbf{0}, \mathbf{m}\}$ . In this case, we will use the operator  $U$  constructed from  $\tilde{V} + \varepsilon \tilde{\Phi}^1$ , without including further terms. As a result, we will eliminate both  $\Phi_{0m}^1$  and  $\Phi_{0n}^1$ . This may introduce some new higher order terms. The exact result of the operations will, in particular, depend on the order in which  $\Phi_{0m}^1$  and  $\Phi_{0n}^1$  are eliminated, but that dependence will only affect higher order terms.

We intend to apply a sequence of transformations of the above kind to the operator (7.1), where  $f$  satisfies the bullet points in the beginning of the section. The goal of these transformations will be to construct a new operator, unitarily equivalent to  $H(x)$ , satisfying the assumptions made in Section 6.2. We will start by considering  $H(x)$  as a “long range” operator  $H_0(x) = V_0(x) + \varepsilon \Phi_0^1(x)$ , where  $\Phi_0^1(x) = \Delta$ . During later steps, the operator will have non-constant off-diagonal terms of higher range than  $\Delta$ . In the notation  $\Phi_k^j(x)_{mn}$  above and below,  $k$  is the step of the procedure,  $j$  is the order of the perturbation, and  $\mathbf{mn}$  indicate the matrix element under consideration. The transformation procedure will be performed in two steps as follows.

- For every  $\mathbf{m} \in \mathbb{Z}^d$  such that  $x + \mathbf{m} \cdot \omega \in [a - h, a + h]$ , eliminate all entries  $\Phi_0^1(x)_{mn}$ . This procedure can be performed in a covariant way, so that the resulting operator family  $H_1(x)$  will still be quasiperiodic. After this first step, the family  $H_1(x) = V_1(x) + \varepsilon \Phi_1^1(x) + \varepsilon^2 \Phi_1^2(x) + \varepsilon^3 \Phi_1^3(x)$  will satisfy two additional properties: the diagonal entries will have lower bounds on the derivatives of the order  $\varepsilon^2$ , and for every  $\mathbf{m} \in \mathbb{Z}^d$  such that  $x + \mathbf{m} \cdot \omega \in [a - h, a + h]$ , all edges starting at  $\mathbf{m}$  will have weight at least two (in other words,  $\Phi_1^1(x)_{mn} = 0$ ).

- On the second step, eliminate all entries  $\Phi_1^2(x)_{\mathbf{m}\mathbf{n}}$  for all  $\mathbf{m} \in \mathbb{Z}^d$  with  $x + \mathbf{m} \cdot \omega \in [a - h, a + h]$  (again, one needs to do it in a covariant way described below in more detail).

Afterwards, one needs to check that the new operator family  $H_2(x)$  satisfies the assumptions (sing0) – (sing4) in Subsection 7.2.

**7.3. Step 1.** For  $x \in [a - h, a + h]$  and  $\mathbf{m} \in \mathbb{Z}^d$  with  $|\mathbf{m}|_1 = 1$ , denote by  $U_{1,\mathbf{m}}(x)$  the unitary operator that eliminates the off-diagonal entry  $\Phi_0^1(x)_{\mathbf{m}\mathbf{n}} = (\Delta)_{\mathbf{0}\mathbf{m}}(x) = 1$  from  $H(x)$ . Let

$$U_1(x) = \prod_{\mathbf{m} \in \mathbb{Z}^d: |\mathbf{m}|_1=1} U_{1,\mathbf{m}}(x).$$

Let  $U_2(x) = U_1(x)$  for  $x \in [a - h, a + h]$ ,  $U_2(x) = I$  for  $x \in [-1/2, 1/2] \setminus [a - 2h, a + 2h]$ , and “unitarily interpolated” in between:

$$U_2(a - 2h + th) = ((1 - t)I + tU_1(a - h)) |(1 - t)I + tU_1(a - h)|^{-1}, \quad t \in [0, 1],$$

and similarly on  $[a + h, a + 2h]$ . Additionally, extend  $U_2(x)$  from  $[-1/2, 1/2]$  to  $\mathbb{R}$  by 1-periodicity and continuity in the variable  $x$ .

Next, we would like to spread this transformation to all lattice points by considering

$$U_3(x) = \prod_{\mathbf{n} \in \mathbb{Z}^d} T_{\mathbf{n}} U_2(x + \mathbf{n} \cdot \omega) T_{-\mathbf{n}},$$

where  $T_{\mathbf{n}}$  denotes the lattice translation operator. Note that the factors in the product commute and the product is well-defined in the strong operator topology: for any vector  $\psi \in \ell^2(\mathbb{Z}^d)$  with finite support, all factors of  $U_3(x)$ , except for finitely many, will act on  $\psi$  as an identity operator, and the remaining finitely many will commute with each other.

Define

$$H_1(x) := U_3(x)^* H(x) U_3(x) = V_1(x) + \varepsilon \Phi_1^1(x) + \varepsilon^2 \Phi_1^2(x) + \varepsilon^3 \Phi_1^3(x).$$

Here, we will use several assumptions discussed above:  $\Phi_1^j(x)$  may depend on  $\varepsilon$  through the  $D$  factors. We will also absorb all higher order terms into  $\Phi_1^3(x)$ . Let  $f_1(x) := V_1(x)_{\mathbf{0}\mathbf{0}}$ .

**Lemma 7.2.** *Suppose that  $H(x)$  satisfies the assumptions from Subsection 7.1. Construct  $U_3(x)$  and  $H_1(x)$  as above. There exists  $\varepsilon_0 = \varepsilon_0(f, \omega) > 0$  and  $c_1 = c_1(f, \omega) > 0$ ,  $c_2(f, \omega) > 0$ ,  $c_3(d) > 0$  such that, for  $0 < \varepsilon < \varepsilon_0$  we have the following:*

- (1)  $f_1$  is  $(C_{\text{reg}} - c_2 \varepsilon)$ -regular on  $[-1/2, a - h] \cup [a + h, 1/2]$ .
- (2)  $f'_1(x) \geq 1 - c_2 \varepsilon$  for  $x \in [-1/2, a - h] \cup [a + h, 1/2]$ .
- (3)  $f'_1(x) \leq D + c_2 \varepsilon$  for  $x \in [a - 2h, a + 2h]$ .
- (4)  $\|(\Phi_1^j)_{\mathbf{m}\mathbf{n}}\|_{C^1} \leq c_2$ .
- (5) The range of  $\Phi_1^2(x)$  is at most 2; in other words,  $\Phi_1^2(x)_{\mathbf{m}\mathbf{n}} \neq 0$  can only happen for  $|\mathbf{m} - \mathbf{n}|_1 \leq 2$ . The range of  $\Phi_1^3(x)$  is bounded by  $c_3(d)$ .
- (6) For  $x + \mathbf{m} \cdot \omega \in [a - h, a + h] + \mathbb{Z}$ , we have  $\Phi_1^1(x)_{\mathbf{m}\mathbf{n}} = 0$ .
- (7) For  $x \in [a - h, a + h] + \mathbb{Z}$ , we have  $f'_1(x) \geq c_1 \varepsilon^2$ .

*Proof.* Properties (1) – (4) follow from the fact that all matrices  $U_j(x)$  involved in the process have finite range are small perturbations of identity with  $\|(I - U_j(x))_{\mathbf{m}\mathbf{n}}\|_{C^1} \leq c(\omega, f) \varepsilon$ . The latter follows from the fact that the entries of  $U_j(x)$  are obtained from  $\frac{1}{f(x) - f(x + \mathbf{n} \cdot \omega)}$ , where  $\text{dist}(\mathbf{n} \cdot \omega, \mathbb{Z}) \geq 2h$  (and therefore  $f$  is regular either at  $x$  or at  $x + \mathbf{n} \cdot \omega$ ). Property (5) follows from the construction of the operators  $U$ : one can note that any edge created between  $\mathbf{n}$

and  $\mathbf{m}$  will have length at least  $|\mathbf{m} - \mathbf{n}|$ . Note that, since we absorb all lower order terms into  $\Phi_1^3$ , we cannot argue that its range is bounded by 3. However, it is clearly bounded by some constant that only depends on the number of matrix multiplications which, in turn, only depends on the dimension. Property (6) directly follows from the construction of  $U$ . Property (7) is the most important property which follows from the calculations of Section 7.2. For  $x \in [a - h, a + h]$ , we have

$$f_1(x) = f(x) + \varepsilon^2 \sum_{j=1}^d \left( \frac{1}{f(x) - f(x + \omega_j)} + \frac{1}{f(x) - f(x - \omega_j)} \right) + O(\varepsilon^3),$$

where the last term admits a  $C^1$ -uniform upper bound in  $x$  on  $[a - h, a + h]$ . In particular, we have  $f'_1(x) \geq c\varepsilon^2$  on  $[a - h, a + h]$ , where  $c = c(f, \omega) > 0$ ; note that both  $x + \omega_j$  and  $x - \omega_j$  are outside of  $[a - h, a + h] \bmod \mathbb{Z}$ , and therefore the derivatives of  $f$  are bounded from below in that region.  $\square$

**7.4. Step 2.** The operator  $H_1(x)$  is a “borderline” case which just misses the assumptions of (sing4) and thus Theorem 6.6. Indeed, we have only one singular interval with  $\mu_1 = 2, \nu_1 = 1$ . The shortest possible loop starting from a singular small denominator has length 4 (two steps of length two). By attaching multiple copies of this loop, we can have  $\text{singdownedges}(\mathcal{P}) \sim |\mathcal{P}|/4$ , which fall short of satisfying (sing4).

The above issue can be resolved by performing another step of eliminating diagonal entries. For each  $\mathbf{n} \in \mathbb{Z}^d$  with  $|\mathbf{n}|_1 \leq 2$  and  $x \in [a - h, a + h]$ , let  $U_{3,\mathbf{n}}(x)$  be the unitary operator which eliminates the entry  $\Phi_1^2(x)_{\mathbf{0}\mathbf{n}}$  in the operator  $V_1(x) + \varepsilon^2 \Phi_1^2(x)$ . Note that, due to (6) in Lemma 7.2, we already have  $\Phi_1^1(x)_{\mathbf{0}\mathbf{n}}$  eliminated for  $x$  under consideration. Let  $U_3(x)$  be the operator obtained from  $U_{3,\mathbf{n}}(x)$  in the same way as  $U_3$  was obtained from  $U_1$  (that is, interpolation with the identity and extension by covariance).

**Lemma 7.3.** *Under the assumptions of Lemma 7.2 and with the above construction of  $U_3(x)$ , let*

$$H_2(x) := U_3(x)^* H_1(x) U_3(x) = V_2(x) + \varepsilon \Phi_2^1(x) + \varepsilon^2 \Phi_2^2(x) + \varepsilon^3 \Phi_2^3(x).$$

*Then, the operator  $H_2 = V_2 + \varepsilon \Phi_2^1 + \varepsilon^2 \Phi_2^2 + \varepsilon^3 \Phi_2^3$  satisfies (1)–(7) from Lemma 7.2. In addition,  $\Phi_2^2(x)_{\mathbf{m}\mathbf{n}} = 0$  for  $x + \mathbf{m} \cdot \omega \in [a - h, a + h] + \mathbb{Z}$ .*

*Proof.* The conclusions of Lemma 7.2 remain true under perturbations of the diagonal entries by  $O(\varepsilon^3)$  and off-diagonal entries of  $O(\varepsilon^2)$ , which is clearly the case if one conjugates  $H_1$  using  $U_3$ . The additional claim follows from the construction of  $U_3$ , since it completely eliminates  $\Phi_1^2(x)_{\mathbf{m}\mathbf{n}}$  with  $x + \mathbf{m} \cdot \omega \in [a - h, a + h] + \mathbb{Z}$ , possibly creating additional terms of order  $\varepsilon^3$  or higher.  $\square$

**Theorem 7.4.** *Let  $H(x)$  be an operator of the class defined in Section 7.1. Construct the operator  $H_2(x)$  using  $U_1(x) - U_3(x)$  as above, and consider the loop expansion for  $H_2(x)$  as in Section 7.1. There exist  $\beta_{\max} = \beta_{\max}(f, \omega) > 0$ ,  $C(f, \omega) > 0$  and  $\varepsilon_0 = \varepsilon_0(f, \omega) > 0$  such that, for*

$$0 < \varepsilon < \varepsilon_0, \quad 0 < \beta < \beta_{\max},$$

*the operator  $H_2(x)$  satisfies (sing4) from Section 7.3 and, as a consequence, the perturbation series for  $H_2(x)$  converges and both  $H(x)$  and  $H_2(x)$  satisfy Anderson localization for all  $x \in \mathbb{R} \setminus (\mathbb{Z} + 1/2 + \mathbb{Z}^d \cdot \omega)$ .*

*Proof.* In order to check (sing4) for  $H_2(x)$ , let  $\mathcal{P}$  be a loop stack. For a non-base loop  $\mathcal{L}$  on  $\mathcal{P}$ , let us consider the total number of  $\delta$  factors produced by this loop (including the factor from its attachment). The last bullet point in the beginning of Section 7.1 implies that, starting from the interval  $[a - h, a + h]$ , the total length of steps to return to that interval without ending the loop will be at least 7. As a consequence, *any loop  $\mathcal{L}$  with  $|\mathcal{L}| \leq 6$  does not have any singular small denominators on it*. Therefore, the contribution of this loop is of the order  $\delta^{-2} \sim \varepsilon^{-4}$ , which gives a total contribution of at most  $O(\varepsilon^{-2k_1/3})$ , where  $k_1$  is the total length of these loops.

Let  $k_2$  be the total length of loops  $\mathcal{L}$  on  $\mathcal{P}$  with  $|\mathcal{L}| \geq 7$  (note that  $H_2(x)$  may contain off-diagonal entries of odd ranges). The total contribution of the attachment factors from these loops is  $\varepsilon^{-4k_2/7}$ . Additionally, each such loop can have at most  $\lfloor |\mathcal{L}|/6 \rfloor$  singular small denominators (as the result of Lemma 7.3). As a consequence, the total contribution from  $\delta$ -factors from these loops will be  $O(\varepsilon^{-4k_2/7-2k_2/6}) = O(\varepsilon^{-19k_2/21})$ . By combining all bounds, we arrive to (conv4) with  $r = 19/21$ .  $\square$

**Remark 7.5.** Suppose that  $f$  indeed has a flat piece. One can check that then the integrated density of states of our operator  $H$  has a lower derivative number of order  $\sim \varepsilon^{-2}$  on an interval of length  $\sim \varepsilon^2$ ; outside this interval, derivative numbers are bounded. Thus, if the ‘non-integrated density of states’ exists, it has a pike of order  $\sim \varepsilon^{-2}$  on an interval of length  $\sim \varepsilon^2$  and is bounded otherwise.

**Remark 7.6.** The assumptions on  $f$  listed at the beginning of this section are not optimal. First of all, we can assume, for example, that  $f$  is non-decreasing when  $x \in [a - h, a + h]$  and  $f$  is Lipschitz monotone (meaning  $f'(x) > 1$ ) outside of this interval. This allows  $f$  to have continuous derivative (or even be infinitely smooth). The assumptions on the ‘number of steps’ we need to make to return from the flat piece back to the flat piece can also be improved. We plan to come back to this class of examples in a subsequent publication and discuss other effects occurring here.

## REFERENCES

- [1] Arnol'd V., *Remarks on perturbation theory for problems of Mathieu type* (Russian), Uspekhi Mat. Nauk 38 (1983), no. 4 (232), 189 – 203.
- [2] Bellissard J., Lima R., Scoppola E., *Localization in  $\nu$ -dimensional incommensurate structures*, Comm. Math. Phys. 88 (1983), no. 4, 465 – 477.
- [3] Eliasson H., *Absolutely convergent series expansions for quasi periodic motions*, Math. Phys. Electron. J. 2 (1996), Paper 4, 33 pp.
- [4] Figotin A., Pastur L., *An exactly solvable model of a multidimensional incommensurate structure*, Comm. Math. Phys. 95 (1984), no. 4, 401 – 425.
- [5] Grempel D., Fishman S., Prange R., *Localization in an incommensurate potential: an exactly solvable model*, Phys. Rev. Lett. 49 (1982), 833 – 836.
- [6] Simon B., *Almost periodic Schrödinger operators. IV. The Maryland model*, Ann. Phys. 159 (1985), no. 1, 157 – 183.
- [7] Jitomirskaya S., Liu W., *Arithmetic spectral transitions for the Maryland model*, Comm. Pure Appl. Math. 70 (2017), no. 6, 1025 – 1051.
- [8] Jitomirskaya S., Kachkovskiy I., *All couplings localization for quasiperiodic operators with Lipschitz monotone potentials*, J. Eur. Math. Soc. 21 (2019), no. 3, 777 – 795.
- [9] Jitomirskaya S., Yang F., *Pure point spectrum for the Maryland model: a constructive proof*, Ergodic Theory Dynam. Systems (2021), no. 1, pp. 283 – 294.
- [10] Jitomirskaya S., Yang F., *Singular continuous spectrum for singular potentials*, Comm. Math. Phys. 351 (2017), no. 3, 1127 – 1135.

- [11] Kachkovskiy I., *Localization for quasiperiodic operators with unbounded monotone potentials*, J. Funct. Anal 277 (2019), no. 10, 3467 – 3490.
- [12] Kachkovskiy I., Parnovski L., Shterenberg R., Krymski S., *Perturbative diagonalisation for Maryland-type quasiperiodic operators with flat pieces*, J. Math. Phys. 62 (2021), no. 6, 063509.
- [13] Gentile G., Mastropietro V., *Anderson Localization for the Holstein Model*, Comm. Math. Phys. 215 (2000), 69 – 103.

DEPARTMENT OF MATHEMATICS, MICHIGAN STATE UNIVERSITY, WELLS HALL, 619 RED CEDAR RD, EAST LANSING, MI, 48824, USA

*Email address:* ikachkov@msu.edu

DEPARTMENT OF MATHEMATICS, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, WC1E 6BT, UK

*Email address:* leonid@math.ucl.ac.uk

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALABAMA, BIRMINGHAM, CAMPBELL HALL, 1300 UNIVERSITY BLVD, BIRMINGHAM, AL, 35294, USA

*Email address:* shterenb@math.uab.edu