Explicit Concentration Inequalities for Eigenvalue-counting Functions in the Anderson Model

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik
der Technischen Universität Dortmund
vorgelegt von

Max Kämper

im Mai 2024

Dissertation
Explicit Concentration Inequalities for Eigenvalue-counting Functions in the Anderson Model

Fakultät für Mathematik
Technische Universität Dortmund

Erstgutachter:	Prof. Dr. Ivan Veselić
Zweitgutachter:	Prof. Dr. Felix Pogorzelski
Tag der mündlichen Prüfung:	16.09.2024

Acknowledgements

First and foremost I would like to express my deepest gratitude to my advisor Prof. Ivan Veselić. The wisdom and knowledge he shared were a tremendous help to me, and I deeply appreciate his guidance.

Next I would like to extend my sincere thanks to Christoph Schumacher and Fabian Schwarzenberger, who taught me a lot about the topic we worked on and contributed many great ideas.

I am thankful to them as well as Albrecht Seelmann and Daniel Rosen for always having an open ear for my ideas and questions, and for the many fruitful discussions and feedback that followed.
Furthermore, I would like to thank them and Nimet Çoşkun for many helpful remarks on early drafts of this thesis.

Thanks should also go to the current and former PhD students of LS IX, in particular Alexander Dicke, Christopher Strothmann, Matthias Täufer and Alexey Schwarzmann for the great discussions, enjoyable coffee breaks and fun time!
I also want to thank Nicole Hufnagel and Kilian Hermann as well as the other members of RTG 2131 for many interesting seminars and learning opportunities.

Lastly I would like to thank my parents Thomas and Christel Kämper for always supporting me and believing in me, as well as Jessica Dreistadt for her encouragement and shared happiness.

Abstract

The Anderson model is a discrete approximation of the Hamiltonian describing the quantum mechanical behaviour of electrons in crystals or metals, featuring a Schrödinger operator on a graph consisting of the sum of the discrete Laplace operator Δ and a random multiplication operator V_ω.
Compressing this operator to the subspace of functions with support in a finite set Λ yields a hermitian matrix with real eigenvalues.
This thesis provides the first explicit quantification of the uniform and almost-sure convergence of the normalized eigenvalue-counting functions N_ω^Λ of these matrix operators to the integrated density of states N of the original Anderson model.
Results are given as concentration inequalities both for the lattice $\mathbb{Z}^d$ as well as general Cayley graphs of finitely generated amenable groups.
Central to these results is a uniform law of large numbers, which is quantified by a bound on the Orlicz norm of a supremum over an empirical process.

For $\mathbb{Z}^d$, $d \geq 3$ and suitable random fields V_ω there exists a universal constant $K < 1186$ and sets $\Omega(n)$ such that

$$\left\| N_\omega^{\Lambda_n} - N \right\|_\infty \leq c(n)$$

for all $\omega \in \Omega(n)$ with explicit $c(n) \sim \frac{1}{\sqrt{n}}$ and

$$\mathbb{P}\left(\Omega(n)\right) \geq 1 - 2\exp\left(- \frac{\sqrt{\lfloor n / \lfloor \sqrt{n} \rfloor \rfloor^d}}{\lfloor \sqrt{n} \rfloor K} \right),$$

where $\Lambda_n = [0, n)^d \cap \mathbb{Z}^d$.
Analogous results for $d = 1, 2$ as well as approximations along monotiling Følner sequences are also given.
For finitely generated amenable groups analogous bounds are shown along Følner sequences based on the Ornstein-Weiss concept of ε-quasitilings.
Treatment for operators with unbound hopping range is also achieved in the case of the Laplace operator on long-range percolation graphs on $\mathbb{Z}^d$.

Contents

1 Introduction

The *Anderson model* is used in physics to describe electrons in alloys or crystals. In this model, the energy of the electrons is governed by a random Schrödinger operator of the form

$$H_\omega = -\Delta + V_\omega$$

acting on functions on $\mathbb{R}^d$, $\mathbb{Z}^d$ or similar spaces. Here, $-\Delta$ is the negative Laplace operator representing the kinetic energy of the electrons and V_ω is a random multiplication operator that models the potential energy of atomic nuclei's electric charges. The randomness should here be understood as a result of the lack of knowledge about the exact composition of a given alloy or crystal, arising for example from impurities or unknown constellations of atoms. An important feature of the model is the fact that although the operator is random, its spectrum is almost surely a specific non-random set under suitable assumptions on V_ω. Thus, at least for the spectrum, the randomness "self-averages" and instead only a macroscopic deterministic effect survives. This motivates investigations of transitions from randomness to non-randomness for related objects, and how and why this transition appears.
In this thesis, we will focus on the Anderson model for functions on $\mathbb{Z}^d$ or similar geometries, and *(normalized) eigenvalue-counting functions* N_ω^Λ. For a given finite subset Λ of $\mathbb{Z}^d$, H_ω is compressed to the subspace of functions with support in Λ, and then $N_\omega^\Lambda(x)$ gives the fraction of eigenvalues of this restricted operator not greater than a real number x. For suitable V_ω, these functions also exhibit a transition to non-randomness if Λ_n is a sequence of larger and larger cubes, with an almost-sure non-random limit function called the *integrated density of states* N. Initially, this convergence was shown point-wise via ergodicity arguments, but later it was also confirmed to hold uniformly in the energy variable x. Schumacher, Schwarzenberger and Veselić [SSV17] showed that the convergence can be quantified as follows: For $n > 2m > 0$, $\Lambda_n = [0,n)^d \cap \mathbb{Z}^d$ and $\kappa > 0$ there is a set $\Omega(\kappa, m, n)$ such that

$$\left\| N_\omega^{\Lambda_n} - N \right\|_\infty \leq c(n,m) + \frac{\kappa}{m^d} \tag{1.1}$$

for all $\omega \in \Omega(\kappa, m, n)$ and

$$\mathbb{P}\left(\Omega(\kappa, m, n)\right) \geq 1 - b(\kappa, m) \exp\left(-a(\kappa, m) \lfloor n/m \rfloor^d\right) \tag{1.2}$$

with explicit $c(n,m)$, but *without* any bound on $a(\kappa, m)$ or $b(\kappa, m)$.

Explicit expressions for these functions are the main goal and result of this thesis, found in Theorem 7.1. As a consequence, the "synthetic" parameters m and κ that do not appear on the left side of (1.1) can be chosen as functions of n. For example, for $d \geq 3$, we will prove that we can choose $m(n) = \lfloor \sqrt{n} \rfloor$ and $\kappa(n) = 1/m(n)$ and resulting from that the existence of a set $\Omega(n)$ such that

$$\left\| N_\omega^{\Lambda_n} - N \right\|_\infty \leq c'(n) \tag{1.3}$$

with the explicit asymptotics $c'(n) \sim \frac{1}{\sqrt{n}}$ and the lower bound

$$\mathbb{P}\left(\Omega(n)\right) \geq 1 - 2\exp\left(- \frac{\sqrt{\lfloor n/\lfloor \sqrt{n} \rfloor \rfloor^d}}{\lfloor \sqrt{n} \rfloor K} \right) \tag{1.4}$$

with a universal constant $K < 1186$, which follows from Corollary 7.3.
Let us outline the content of this work. To set the stage, Chapter 2 gives some further background and concrete definitions for the Schrödinger equation, the Schrödinger operator and the Anderson model, as well as some definitions and notation for probability spaces needed to describe the randomness in the model.
After the model is defined, Chapter 3 recalls basic notions from spectral theory, and discusses them in the context of the random Schrödinger operator of the Anderson operator. In particular, we will reproduce the well-known result of the almost-sure non-randomness of the spectrum of the Schrödinger operator in the Anderson model. This proof already exhibits some of the concepts that are also responsible for convergence of the eigenvalue-counting functions, namely how the involvement of an infinite number of random variables together leads to non-random results. Spectra are also necessary to understand the integrated density of states, which is not just the limit of eigenvalue-counting functions but also has a closed form representation. Both $N_\omega^{\Lambda_n}$ and N are precisely defined in Section 3.5, and previous results of their convergence are also listed there.
Chapter 4 starts the preparation for the proof of the explicit bounds we seek. It establishes the notion of an *admissible function*, which formalizes the required properties of the eigenvalue-counting functions for the following proofs, and Section 4.2 contains an overview of the strategy we will take to obtain the results. In general, there are two distinct important ingredients, one geometric and one probabilistic. First, the structure of $\mathbb{Z}^d$ allows us to approximate an admissible function on some large cube with an averaged sum over the same function on smaller cubes.
This scheme works independent of any random influences and is the main result of Chapter 5, essentially reproducing a method of [SSV17]. Sequences of cubes with growing side-length are of special interest here because their surface grows slower than their volume. The chapter also contains a more general formulation for so called *monotiling Følner sequences* in $\mathbb{Z}^d$, which share this property.
The second important ingredient are explicit uniform concentration inequalities for empirical processes, quantifying the "self-averaging" of the randomness. Chapter 6

consists of this ingredient and its main result, two concentration inequalities where one implies a faster convergence but has a restricted applicability and one implies a slower convergence that can be applied in more cases.
Both of these chapters are relatively independent of each other and are combined in Chapter 7 to achieve the stated goal of explicit expressions in (1.2), as well as some of its corollaries such as (1.3) and (1.4).
Apart from the specific case of cubes, results for the more general case of monotiling Følner sequences in $\mathbb{Z}^d$ are also given here.
Chapter 8 contains two further generalizations of the achieved results. The first is a generalization of the geometric arguments in Chapter 5 from $\mathbb{Z}^d$ to Cayley graphs of finitely generated amenable groups. The second is an explicit quantification for the uniform convergence of the eigenvalue-counting functions of the discrete Laplace operator on long-range percolation graphs in $\mathbb{Z}^d$, which shows that the achieved results are not restricted to just the Anderson model.

2 Setting and physical background

The Schrödinger equation and the Schrödinger operator are two fundamental parts of quantum physics, named after physicist Erwin Schrödinger [Sch26], who postulated them in 1926 as part of the effort to understand the possible energy levels of electrons in hydrogen atoms. Inspiration for the formula came from the numerous discoveries about the nature of light at the time that led to the establishment of quantum physics. We will give a short summary in this chapter, for in-depth accounts of this development and the physical background see for example [Hal13] or [GGAGP12].

One of these new discoveries was the photoelectric effect, the emission of electrons from metal induced by light. Increasing the intensity of the light produces more electrons, but raising the frequency of the light produces electrons with higher energy instead. Thus, the energy E of light is linearly dependent on its frequency ω with a factor of proportionality that is called $\hbar$, i.e. $E = \hbar\omega$. On the other hand, the wave nature of light was also known for a long time from the Maxwell equations, which result in wave equations for electric and magnetic waves and therefore light as well.

A function $\Psi : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ with

$$\Psi(x,t) := \exp\left(\mathrm{i}(kx - \omega t)\right)$$

with a wave vector k and frequency ω is a simple plane wave. The relation between frequency and energy for light then means that

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi(x,t) = E\Psi(x,t), \tag{2.1}$$

so knowing the energy of a light wave means knowing its time evolution. The discoveries about light led Louie de Broglie to theorize that massive particles could also be described as wave functions, and experiments by Davisson and Germer [DG27] showed that electrons can indeed be scattered on crystal lattices, which is a phenomenon that was previously only known for waves such as sound waves, water waves and light.

2.1 The Schrödinger equation

The experimental verification of a wave-like behaviour of electrons inspired the development of an equivalent of (2.1) for massive particles, the general Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi = H\Psi$$

for the time evolution of a wave function Ψ, where H is the Hamilton operator which relates to the energy of Ψ.
The wave function at each time t has to be from a Hilbert space, usually the space $L^2(\mathbb{R}^d)$ of functions $f : \mathbb{R}^d \to \mathbb{C}$ that have a finite L^2-norm

$$\|f\|_{L^2(\mathbb{R}^d)} = \sqrt{\int_{\mathbb{R}^d} |f(x)|^2 \,\mathrm{d}x} < \infty. \tag{2.2}$$

Wave functions that are associated to particles are assumed to be normalized at every time t, i.e. $\|\Psi(t)\|_{L^2(\mathbb{R}^d)} = 1$, where we used a characterization as a function $\Psi : \mathbb{R} \to L^2(\mathbb{R}^d)$. The exact meaning of the wave function was hotly debated at the time the Schrödinger equation was developed. The prevailing consensus now is the so called *Born rule*, which states that for every time t, $|\Psi(t)(x)|^2$ is the probability density function for finding the particle at point x.
If there is some operator $U(t)$ describing the time evolution, i.e. $\Psi(t) = U(t)\Psi(0)$ for wave functions Ψ, then the probability to find the particle anywhere at time t still needs to be one, thus $\|\Psi(t)\|_{L^2(\mathbb{R}^d)} = \|U(t)\Psi(0)\|_{L^2(\mathbb{R}^d)} = 1$ and $U(t)$ needs to be unitary for all t. As will be shown in Chapter 3.1, this requires (under suitable conditions) H to be self-adjoint. This is also supported by the fact that energy is a scalar real quantity that can be measured experimentally and is therefore an *observable* like position and momentum. Quantum physics postulates that the operators associated to observables are self-adjoint.
The next task is thus to find a fitting self-adjoint Hamilton operator for a given system.

2.2 The Schrödinger operator

The classical energy of a massive particle in a conservative force field consists of the kinetic energy $T = \frac{1}{2}mv^2 = \frac{p^2}{2m}$ where $v \in \mathbb{R}^3$ is the velocity, $p \in \mathbb{R}^3$ the momentum and m the mass of the particle, and the potential energy $V(x)$ that only depends on the location $x \in \mathbb{R}^3$ of the particle. By replacing the momentum in the j-th direction $p^{(j)}$ with $-\mathrm{i}\hbar\frac{\partial}{\partial x^{(j)}}$ in analogy to the energy, Schrödinger proposed the Schrödinger operator

$$H = T + V = \frac{-\hbar^2}{2m}\Delta + V$$

with the Laplace operator $\Delta = \sum_{j=1}^{3} \frac{\partial^2}{\partial^2 x^{(j)}}$ and the multiplication operator V with $V\Psi(x) = V(x)\Psi(x)$ as the Hamilton operator for one electron in the presence of a proton, which models a Hydrogen atom. Here, $V(x) = \frac{-e^2}{|x|}$ is the electric potential of a single proton with the Coulomb charge e. This model was able to explain the

emergence of the different energy levels of the hydrogen atom and is still used as one model for Hamilton operators.
One way to expand this Schrödinger operator is to add more potentials to simulate more atoms. A finite number of potentials can be used to model molecules, and a potential of the type

$$V(x) = \sum_{n=1}^{\infty} q_n f(x - y_n)$$

can be used to describe an idealized infinitely big metal or crystal. Here, q_n are charges of different atomic nuclei, y_n are their positions, and f is a function describing the decay of the potential, for example $f(z) = \frac{e^{-|z|}}{|z|}$ to account for screening effects of multiple nuclei and to keep the potential finite. Choosing different q_n allows to describe alloys or impurities in substances, since different elements have varying numbers of protons and therefore varying electric potentials. Choosing different y_n allows for different spacial arrangements of the nuclei.
However, it should be noted that this is only a limited model for real materials. The nuclei themselves are not part of the wave function and are assumed to be static, the wave function only considers a single electron, and the potentials of nuclei are more complex than just a function of the distance. Nevertheless, it is an important model allowing the study of some effects of systems with many atoms. A particular case of such systems are crystals or metals, where the positions of the atomic nuclei form a lattice.

2.3 The Schrödinger operator on a lattice

For materials with a very regular atomic structure it is useful to formalize this structure in the form of a "periodic" **graph** $\Gamma = (\mathcal{V}, \mathcal{E})$ consisting of a set of vertices $\mathcal{V}$ and a set $\mathcal{E}$ of edges $(x, y),\ x, y \in V$ between them. The lattice $\mathbb{Z}^d$ already forms a graph where the vertices are the elements of $\mathbb{Z}^d$ and there is an edge between $x, y \in \mathbb{Z}^d$ if they are neighbours, i.e. if $\|y - x\|_1 = \sum_{j=1}^{d} \left|y^{(j)} - x^{(j)}\right| = 1$, where $y^{(j)}$ is the $j - th$ coordinate of y. Despite this model being already less complex than one without structure, its complexity is still high. There is another step of approximation that helps to handle some of it: changing the underlying space from a continuous space like $\mathbb{R}^d$ to the graph as well. Instead of looking at normalized functions on $\mathbb{R}^d$ with L^2-norm one, the relevant functions are now functions $f : \mathbb{Z}^d \to \mathbb{C}$ that are normalized with respect to the ℓ^2-norm

$$\|f\|_{\ell^2(\mathbb{Z}^d)} = \sqrt{\sum_{x \in \mathbb{Z}^d} |f(x)|^2} \tag{2.3}$$

on the space of square-summable sequences of numbers indexed by points in $\mathbb{Z}^d$.
This is also a Hilbert space with scalar product

$$\langle f, g\rangle = \sum_{x\in\mathbb{Z}^d} \overline{f(x)}g(x). \tag{2.4}$$

Just as in the continuous case, the wave functions need to have norm one and for every time t, $|\Psi(t)(x)|^2$ can be thought of as the probability density to find the electron at the site x.
The Schrödinger operator needs to be modified to work on a graph. The potential V is restricted to the vertices, so instead of the overlap of many different potentials f with varying offsets and strengths it is easier to just define V as having a value $v(x)$ for each $x \in \mathbb{Z}^d$.
Furthermore, the Laplace operator needs to be replaced with something else. Since there are no other points with distance less than one around any lattice point, derivatives like on $L^2(\mathbb{R}^d)$ are not possible on $\ell^2(\mathbb{Z}^d)$. However, it is possible to construct an analogous operator as follows:
For a function $f : \mathbb{R}^d \to \mathbb{C}$ that is twice partially differentiable in each coordinate, the definition of the second derivative means for all $x \in \mathbb{R}^d$ and small enough $h \in \mathbb{R}$ that

$$f(x+he_j) = f(x) + \frac{\partial}{\partial x^{(j)}} f(x)h + \frac{1}{2}\frac{\partial^2}{\partial^2 x^{(j)}} f(x)h^2 + o(h^2)$$
$$f(x-he_j) = f(x) - \frac{\partial}{\partial x^{(j)}} f(x)h + \frac{1}{2}\frac{\partial^2}{\partial^2 x^{(j)}} f(x)h^2 + o(h^2)$$

where e_j is the unit vector in the i-th direction and $x^{(j)}$ is the j-th coordinate. From this follows

$$\lim_{h\to 0} \frac{f(x+he_j) + f(x-he_j) - 2f(x)}{h^2} = \frac{\partial^2}{\partial^2 x^{(j)}} f(x)$$

for each $j \in \{1, ..., d\}$. Adding the partial derivatives then leads to

$$\Delta f(x) = \sum_{j=1}^{d} \frac{\partial^2}{\partial^2 x^{(j)}} f = \lim_{h\to 0} \sum_{j=1}^{d} \frac{f(x+he_j) + f(x-he_j) - 2f(x)}{h^2}.$$

If f is now an element of $\ell^2(\mathbb{Z}^d)$, this limit can not be taken, but stopping at the closest point $h = 1$ leads to the **discrete Laplacian** $\Delta : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$,

$$\Delta f(x) = \sum_{i=1}^{d} (f(x+e_i) + f(x-e_i) - 2f(x)) = \sum_{y\in\mathbb{Z}^d:\ \|y-x\|_1=1} (f(y) - f(x)).$$

In order to check whether this is a useful replacement of the continuous Laplacian, we can compare the continuous Fourier transform

$$(\mathcal{F}\phi)(p) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{e}^{-ipx} \phi(x)\mathrm{d}x$$

of the negative continuous Laplacian to the discrete Fourier transform

$$\tilde{\mathcal{F}}\phi(k) = \sum_{x\in\mathbb{Z}^d} \phi(x)\mathrm{e}^{\mathrm{i}x\cdot k}.$$

of the negative discrete Laplacian.
In the continuous case, we get $(\mathcal{F}(-\Delta(\mathcal{F}^{-1}\phi)))(p) = p^2\phi(p)$ for suitable ϕ, and as is shown in Example 3.10 in the discrete case we have

$$(\tilde{\mathcal{F}}(-\Delta(\tilde{\mathcal{F}}^{-1}\phi)))(p) = \left(2d - 2\sum_{j=1}^{d} \cos(p^{(j)})\right)\phi(p) = \left(p^2 + o(p^4)\right)\phi(p).$$

The action of the operators agrees at lowest order in p, so there is indeed some connection.
All together, it is possible to define a Schrödinger operator for a bounded potential $v(x)$ on the lattice as $H : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$,

$$H\Psi(x) = -\Delta\Psi(x) + v(x)\Psi(x) = \sum_{y\in\mathbb{Z}^d:\ \|y-x\|_1=1} (\Psi(x) - \Psi(y)) + v(x)\Psi(x).$$

Here, the prefactor $\frac{-\hbar^2}{2m}$ has been dropped for ease of notation. This is possible since the numerical value of this prefactor depends on the chosen physical units, so there is no loss of generality by setting it to one.
Unbounded potentials can also be used, but in that case the domain of the operator needs to be restricted. Further details on this will be given in Chapter 3.

2.4 Probability spaces

To fully specify a Hamilton operator, we still need to determine what the potential v looks like. Irregularities and introduction of intentional impurities (also called doping) of the crystal can lead to a lot of variation in the potential at different points and measuring the exact values is difficult. And even if the potential of one crystal is measured, another similar crystal could have roughly the same distribution of potentials at its sites, but in a completely different order.
One approach to this is to assume that the potential is not a fixed given function but a random one. To describe this function, we will need some definitions for the random component of the operator, as well as some other definitions and properties we will use in later chapters. This section will only cover the basic definitions and no proofs, for more details see for example [Kle20] or [Kal21].

Definition 2.1 (Probability and measure)
*A **measurable space** $(\Omega, \mathcal{A})$ consists of a set Ω, the **sample set**, and a σ-algebra $\mathcal{A}$, whose elements are called **events**.*
*A σ**-algebra** $\mathcal{A}$ is a class of subsets of Ω that satisfies the following three conditions:*

- $\Omega \in \mathcal{A}$,
- *If $A \in \mathcal{A}$, then $A^C \in \mathcal{A}$, where $A^C = \Omega \setminus A$ is the complement of A,*
- *If $(A_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$*

*If Ω is a general topological space such as $\mathbb{R}^d$ or $\mathbb{C}$, then the **Borel σ-algebra** $\mathcal{B}(\Omega)$ is defined as the smallest σ-algebra that contains all open sets.*
*A map $P : \mathcal{A} \to [0, \infty]$ is called a **measure** if*

- *$P(\emptyset) = 0$, where $\emptyset = \Omega^C$ is the empty set,*
- *P is σ-additive, that is for every sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ of **pairwise disjoint** sets, meaning $A_i \cap A_j = \emptyset$ for all $i \neq j$, the map P satisfies*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

*If there is a sequence of subsets $(U_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ such that $\bigcup_{n\in\mathbb{N}} U_n = \Omega$ and $P(U_n) < \infty$ for all $n \in \mathbb{N}$, then P is called σ**-finite**.*
*A space $(\Omega, \mathcal{A}, P)$ consisting of a set Ω, a σ-algebra $\mathcal{A}$ and a measure P is called a **measure space**. If there is an event A such that $P(\Omega \setminus A) = 0$, then A is said to contain P**-almost all** or **almost all** (if the measure is clear from context) elements of Ω.*
*If $P(\Omega) = 1$, then P is called a **probability measure** and is usually written with the symbol $\mathbb{P}$, and a measure space $(\Omega, \mathcal{A}, \mathbb{P})$ with a probability measure $\mathbb{P}$ is called a **probability space**. An event A with $\mathbb{P}(A) = 1$ is called **almost sure**.*

Similar to σ-algebras are **topologies**, another class of subsets of a given set Ω. Every topology $\mathcal{T}$ needs to satisfy

- $\Omega, \emptyset \in \mathcal{T}$,
- If $A, B \in \mathcal{T}$, then $A \cap B \in \mathcal{T}$,
- If $\mathcal{I} \subset \mathcal{T}$, then $\bigcup_{A\in\mathcal{I}} A \in \mathcal{T}$.

A **topological space** $(\Omega, \mathcal{T})$ consists of a set Ω with an associated topology $\mathcal{T}$.

For $\Omega = \mathbb{R}$ and $\Omega = \mathbb{R}^d$ the canonical topologies are the families of all open subsets. If $(\Omega, \mathcal{T})$ is a topological space, then the **Borel σ-algebra $\mathcal{B}(\Omega)$** is defined as the smallest σ-algebra that contains $\mathcal{T}$.

> If not indicated otherwise, the σ-algebra for $\Omega = \mathbb{R}$ and $\Omega = \mathbb{R}^d$ are the Borel σ-algebras $\mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\mathbb{R}^d)$, the smallest σ-algebra that contain all open intervals and all open d-dimensional cubes, respectively.
> The σ-algebra for $\Omega = \mathbb{R}^{\mathbb{Z}^d}$ is the product Borel σ-algebra of $\mathcal{B}(\mathbb{R})$ (see also Definition 2.4) unless indicated otherwise.

Topologies are used to characterise continuity, since a function $f : (\Omega_1, \mathcal{T}_1) \to (\Omega_2, \mathcal{T}_2)$ is continuous if (and only if) $f^{-1}(A) \in \mathcal{T}_1$ for all $A \in \mathcal{T}_2$. The same condition applied to σ-algebras leads to the concept of measurable functions.

Definition 2.2 (Measurable functions and random variables)
*A function $X : (\Omega, \mathcal{A}) \to (Y, \mathcal{B})$ from one measurable space to another is called a **measurable function** if $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.*
*For every measurable function $X : (\Omega, \mathcal{A}, P) \to (Y, \mathcal{B})$ there is also an **image measure** $P_X : \mathcal{B} \to [0, \infty]$ defined by $P_X(B) = P(X^{-1}(B))$.*
*A measurable function from a probability space to a measurable space is called a **random variable**.*

Remark. A function between measurable spaces equipped with Borel σ-algebras is measurable if it is continuous, but for example indicator functions of measurable sets are measurable but not continuous.

Definition 2.3 (Expected values and integrals)
*Let $(\Omega, \mathcal{A}, P)$ be a measure space. The **integral (with respect to P)** for real functions on this space is defined by the following three typical steps. If s is a non-negative **step-function**, that is a measurable function that only takes a finite number of non-negative values, then s is of the form*

$$s = \sum_{n=1}^{k} c_n \mathbb{1}_{A_n} \tag{2.5}$$

where $c_n \in (0, \infty)$ for all $1 \leq n \leq k$ and $(A_n)_{1 \leq n \leq k}$ is a finite sequence of pairwise disjoint sets in $\mathcal{A}$. The integral of a step function is then defined as

$$\int_\Omega s \; \mathrm{d}P = \int_\Omega s(\omega) \mathrm{d}P(\omega) := \sum_{i=1}^{k} c_i P(A_i),$$

which is well-defined with regard to the representation (2.5)*. The integral can then be extended to non-negative measurable functions* $X\colon (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *via*

$$\int_\Omega X \,\mathrm{d}P := \sup\left\{\int_\Omega s \,\mathrm{d}P \mid 0 \le s(\omega) \le X(\omega)\forall\omega \in \Omega, s \text{ is a step function}\right\}$$

but here the integral might be infinite. The integral can then be further extended to all measurable functions $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *by*

$$\int_\Omega X \,\mathrm{d}P = \int_\Omega X_+ \,\mathrm{d}P - \int_\Omega X_- \,\mathrm{d}P$$

where $X_+ := \max\{X, 0\}$*,* $X_- := -\min\{X, 0\}$*, provided at most one of the terms on the right side is infinite. If* $\int_\Omega |X| \,\mathrm{d}P < \infty$ *then* X *is called* ***(P-)integrable****. By integrating real and imaginary parts separately, the integral can also be extended to complex measurable functions.*
If $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space and* X *a random variable, the integral* $\int_\Omega X \mathrm{d}\mathbb{P}$ *is called the* ***expected value*** *of* X *and often written as* $\mathbb{E}(X)$*. The integral is linear, and for* $1 \le p < \infty$ *the maps*

$$\|X\|_{L^p(\Omega,\mathcal{A},P)} := \left(\int_\Omega |X|^p \,\mathrm{d}P\right)^{1/p}$$

and

$$\|X\|_{L^\infty(\Omega,\mathcal{A},P)} := \inf\{M \ge 0 \mid P(\{|X| > M\}) = 0\}$$

are semi-norms on the vector space

$$\mathcal{L}^p(\Omega, \mathcal{A}, P) := \{X\colon (\Omega, \mathcal{A}, P) \to \mathbb{C} \mid \ X \text{ is measurable and } \|X\|_{L^p(\Omega,\mathcal{A},P)} < \infty\}.$$

Thus the quotient space

$$L^p(\Omega, \mathcal{A}, P) = \mathcal{L}^p(\Omega, \mathcal{A}, P)/\mathcal{N}$$

with

$$\mathcal{N} = \{X : (\Omega, \mathcal{A}, P) \to \mathbb{C} \mid \ X \text{ is measurable and } X(\omega) = 0 \text{ for } P\text{-almost all } \omega\}$$

is a normed space, and even a Banach space (see for Example [Kle20, Theorem 7.18]). We will abbreviate $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P_L)$ *with the Lebesgue-measure* P_L *to just* $L^p(\mathbb{R}^d)$ *(as we already did in Section 2.1) and analogously for subsets of* $\mathbb{R}^d$*. For* $p = 2$ *the space is also a Hilbert space with norm as defined in* (2.2) *and scalar product*

$$\langle \phi, \psi\rangle = \int_{\mathbb{R}^d} \overline{\phi(x)}\psi(x)\mathrm{d}x.$$

We will also abbreviate $L^p(\mathbb{Z}^d, \mathcal{P}(\mathbb{Z}^d), P_c)$ *with the power set* $\mathcal{P}(\mathbb{Z}^d)$ *of* $\mathbb{Z}^d$ *and the counting measure* P_c *to* $\ell^p(\mathbb{Z}^d)$ *(as we already did in Section 2.3). This is also a Hilbert space in the case* $p = 2$ *with norm and scalar product as defined in* (2.3) *and* (2.4)*.*

Definition 2.4 (Cylinder sets)
*Let $(\Omega, \mathcal{A})$ be a measurable space and $\Omega^G = \{\omega = (\omega_i)_{i \in G} \mid \omega_i \in \Omega \ \forall i \in G\}$ for some set G a product space. A **(rectangular) cylinder set** is a set of the form*

$$A = \left\{\omega \in \Omega^G \mid \omega_i \in A_i \ \forall i \in \{k_1, ..., k_n\}\right\} = \bigtimes_{i=1}^{n} A_{k_i} \times \Omega^{G \setminus \{k_1,...,k_n\}},$$

*where $n \in \mathbb{N}$, $\{k_1, ..., k_n\}$ is a finite subset of G and $A_i \in \mathcal{A}$ for all $i \in \{k_1, ..., k_n\}$. The smallest σ-algebra that contains all rectangular cylinder sets is called the **product σ-algebra generated by cylinder sets**, written $\mathcal{A}^G$.*

Remark 2.5. The product σ-algebra $\mathcal{A}^G$ is equivalently the smallest σ-algebra $\mathcal{S}$ such that all projections

$$\Pi_i : \left(\Omega^G, \mathcal{S}\right) \to (\Omega, \mathcal{A}), \ \Pi_i(\omega) := \omega_i$$

are measurable.
Furthermore $\mathcal{A}^G$ is also the smallest σ-algebra $\mathcal{S}$ such that for every finite subset D of G the projection

$$\Pi_D : \left(\Omega^G, \mathcal{S}\right) \to \left(\Omega^D, \mathcal{A}^D\right), \ \left(\Pi_D(\omega)\right)_i := \omega_i \ \forall i \in D$$

is measurable.
Just as for σ-algebras there is also a product of topologies. If $(\Omega, \mathcal{T})$ is a topological space, then the product topology $\mathcal{T}^G$ is the smallest topology S on Ω^G such that all projections

$$\Pi_i : \left(\Omega^G, S\right) \to (\Omega, \mathcal{T}), \ \Pi_i(\omega) := \omega_i$$

are continuous. For topological spaces the product σ-algebra of Borel σ-algebras is the Borel σ-algebra of the product topology, and specifically for $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $G = \{1, ..., d\}$ we have $\mathcal{B}(\mathbb{R})^G = \mathcal{B}(\mathbb{R}^d)$, where $\mathcal{B}(\mathbb{R}^d)$ is the smallest σ-algebra that contains all open sets in $\mathbb{R}^d$. Since all projections have to be continuous, convergence in the product topology is point-wise convergence.

Definition 2.6 (Ergodicity and mixing)
*Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A measurable map $T \colon \Omega \to \Omega$ is called **measure preserving** if $\mathbb{P}_T = \mathbb{P}$. If T is measure preserving and every T-invariant event $A \in \mathcal{A}$, meaning $T^{-1}(A) = A$, has either measure 0 or 1, then T is called **ergodic**.*
*If $\{T_g \mid g \in G\}$ is a group of measure preserving maps for some set G such that every event that is invariant under every T_g has either measure 0 or 1, then $\{T_g \mid g \in G\}$ is called an **ergodic group** or a **group acting ergodocally on** Ω.*
If T is measure preserving and for all $A, B \in \mathcal{A}$ we have

$$\lim_{n \to \infty} \mathbb{P}\left(T^{-n}(A) \cap B\right) = \mathbb{P}(A)\mathbb{P}(B)$$

*then T is called **mixing**.*

Remark 2.7. Mixing implies ergodicity since it results in

$$\mathbb{P}(A) = \mathbb{P}(T^{-n}(A) \cap A) = \mathbb{P}(A)^2$$

for every T-invariant event A.

Definition 2.8 (Independence)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ is called ***independent*** *if*

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_{n_i}\right) = \prod_{i=1}^{k} \mathbb{P}\left(A_{n_i}\right)$$

for all $k \in \mathbb{N}$ and all indices $n_1 < n_2 < ... < n_k$.
A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables $X_n\colon (\Omega, \mathcal{A}, \mathbb{P}) \to (Y_n, \mathcal{B}_n)$ is called independent if the sequence $\left(X_n^{-1}(B_n)\right)_{n\in\mathbb{N}}$ is independent for all choices of $B_n \in \mathcal{B}_n$, $n \in \mathbb{N}$.
A sequence $(X_n)_{n\in\mathbb{N}}$ of independent random variables where the induced image measures P_{X_n} are all the same is called ***independent and identically distributed (i.i.d.)****.*

We will also need the following Borel-Cantelli-Lemma.

Lemma 2.9 ([Kle20, Theorem 2.7])
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(A_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ and A_∞ the ***Limes superior*** *of the sequence $(A_n)_{n\in\mathbb{N}}$, defined as*

$$A_\infty = \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m,$$

so that

$$\omega \in A_\infty \Leftrightarrow \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}.$$

Then:

1. *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_\infty) = 0$.*
2. *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $(A_n)_{n\in\mathbb{N}}$ is an independent sequence, then $\mathbb{P}(A_\infty) = 1$.*

2.5 The Anderson model

With the definitions from the previous section we can now define a random potential. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $v_{\cdot}(z) : \Omega \to \mathbb{R}$ a random variable for every $z \in \mathbb{Z}^d$. For a specific $\omega \in \Omega$ the potential is then $v_\omega(z)$ and we can define a multiplication operator V_ω via point-wise multiplication with v_ω. This model for the potential was developed by Philip W. Anderson [And58] in 1957 to investigate localization, and is thus called the **Anderson model**. The operator arising now is $H_\omega = -\Delta + V_\omega : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ with

$$H_\omega \Psi(z) := \sum_{y \in \mathbb{Z}^d:\ |y-z|=1} (\Psi(z) - \Psi(y)) + v_\omega(z)\Psi(z). \tag{2.6}$$

for each $\omega \in \Omega$. This operator is called the **random Schrödinger operator of the Anderson model**. For readability, we will refer to this operator as just the **Anderson operator** in this thesis. The goal in the following is to find “typical” properties for this operator.
Since the operator has random coefficients, every solution to the Schrödinger equation is also random and the distribution of solutions depends on the random variables $v_{\cdot}(z)$. The deterministic operator H can be recovered by choosing a set Ω that only contains a single point, but the Anderson model also allows -among other things- to describe a setting where at each site z the potential has randomly either the value one or two, each with probability 1/2, independently of all other sites.
This would be a toy model for a crystal that contains two kinds of nuclei, one with twice the charge of the other, but there is no regular ordering where the different nuclei are located.
In conclusion, the Anderson model is an approximation for the operator that describes the behaviour of electrons in crystals that have random irregularities or doping.
The main simplifications are

- the nuclei, and thus their potentials, do not change with time and only the wave function of electrons evolves in time,
- interaction between electrons can be ignored, so only a single electron is considered,
- the crystal is infinitely large, so no boundary effects are considered,
- all of this is done on a lattice instead of the continuous space $\mathbb{R}^d$,
- the missing knowledge of the irregularities is replaced by a random distribution of potentials.

In principle, the Schrödinger equation could now be solved for each ω to get the random solutions, but Ω is typically uncountable, so every computed solution could be “atypical”. Therefore, the next question is whether it is possible to show some properties of the operator despite the randomness. In the next chapter, will define the spectrum of an operator and show that for the Anderson operator the spectrum and some related objects are indeed almost surely non-random.

3 The spectrum and eigenvalue-counting functions

We start off with some mathematical context and definitions for the objects we will encounter in the following chapters. First we will define the spectrum of an operator and then revisit a well-known result for the spectrum of the Anderson operator. Even though the Anderson operator is random, we will show that its spectrum is almost surely a deterministic set. As part of the proof we will already encounter some concepts that also appear in later chapters. This part mostly serves as a demonstration and for familiarisation with the Anderson operator, so not every proof will be given in full generality. Furthermore references in this chapter are chosen more for nice formulations and do not necessarily point to the original source. After this we turn to the concept of spectral families and elements of spectral calculus, and building on it to the introduction of two functions that are key objects for the rest of this thesis: the eigenvalue-counting functions (evcf) and the integrated density of states (IDS). The last part of this chapter is a discussion of previous results for the convergence of the former to the latter. Here, we will show where this thesis connects to prior research, and identify a closed formula for the IDS that will appear again in Chapter 7.

3.1 Eigenvalues and operators

First we will consider an easier system: a matrix $A \in \mathbb{C}^{n \times n}$, vectors $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^n$ and the equation

$$Ax = y.$$

For a large matrix computing y may take a long time for general x, but there may be specific $x \neq 0$ that are only scaled by A, i.e. there is a $\lambda \in \mathbb{C}$ so that $Ax = \lambda x$. In that case λ is called an **eigenvalue** of A with **eigenvector** x. Eigenvalues can be found without knowing their associated eigenvectors since the existence of an eigenvector x means $(A - \lambda I)x = 0$ with the identity matrix I, which is equivalent to the matrix $A - \lambda I$ not being invertible. This itself is equivalent to $\det(A - \lambda I) = 0$. This determinant is a polynomial in λ whose zeroes are therefore the eigenvalues of A.
By the fundamental theorem of algebra every non-constant polynomial with complex coefficients has at least one zero, so at least one eigenvalue always exists. For Hermitian matrices even more is true: every eigenvalue is real and the eigenvectors form an

orthogonal basis of $\mathbb{C}^n$.
The matrix A being **Hermitian** means that

$$\forall v, w \in \mathbb{C}^n\colon\ \langle Av, w\rangle = \langle v, Aw\rangle, \tag{3.1}$$

where $\langle\cdot,\cdot\rangle$ is the scalar product. All real symmetric matrices are hermitian. As discussed before there is at least one eigenvalue λ_1 with an eigenvector e_1, so (3.1) results in

$$\lambda_1\langle e_1, e_1\rangle = \langle e_1, \lambda_1 e_1\rangle = \langle e_1, Ae_1\rangle = \langle Ae_1, e_1\rangle = \langle \lambda_1 e_1, e_1\rangle = \overline{\lambda_1}\langle e_1, e_1\rangle,$$

where $\overline{\lambda_1}$ is the complex conjugate of λ_1, and we can conclude that λ_1 is real. Now consider the space $K_1 = \{v \in \mathbb{C}^n \mid \langle v, e_1\rangle = 0\}$ of vectors orthogonal to e_1. Let $v \in K_1$, then $\langle Av, e_1\rangle = \langle v, Ae_1\rangle = \langle v, \lambda_1 e_1\rangle = 0$, so $Av \in K_1$ or in other words K_1 is invariant under A. We can now restrict A to K_1 to get a new operator A_1. Applying the same procedure as before but now for A_1 leads to a real eigenvalue λ_2 with eigenvector $e_2 \in K_1$. Repeating this procedure n times leads to the full set of real eigenvalues and the orthogonal set of eigenvectors. Note that a matrix can have the same eigenvalue multiple times with different eigenvectors.
Since the eigenvectors form a basis, every $x \in \mathbb{C}^n$ can be written as a sum of eigenvectors, and the eigenvalues give a quantification for how much A is going to change x in the direction of the corresponding eigenvector.

The space $\ell^2(\mathbb{Z}^d)$ that the Anderson operator lives on is not finite dimensional, so it is necessary to define something corresponding to the eigenvalues and eigenvectors for infinite dimensional spaces as well.

Definition 3.1 ((Bounded) operator)
A linear map $T\colon D(T) \subset X \to Y$ on vector spaces X and Y is called an ***operator*** *with* ***domain*** *$D(T)$. The domain is a linear subspace of X. If X and Y are Banach spaces, then T is called a* ***bounded operator*** *if there is an $M \geq 0$ such that*

$$\|Tx\|_Y \leq M\,\|x\|_X\ \ \forall x \in D(T),$$

where $\|\cdot\|_X$ and $\|\cdot\|_Y$ are the norms on X and Y, respectively. The smallest such M, i.e.

$$\inf\left\{M \geq 0 \mid \|Tx\|_Y \leq M\,\|x\|_X\ \ \forall x \in X\right\} =: \|T\|$$

is called the ***norm of*** *T.*
Two operators $T\colon D(T) \to Y$ and $S\colon D(S) \subset Y \to Z$ can be combined to

$$TS\colon D(TS) \to Z, \quad TSx = T(Sx)$$

with $D(TS) = S^{-1}(D(T))$, and two operators $T\colon D(T) \subset X \to Y$ and $S\colon D(S) \subset X \to Y$ can be combined to

$$T + S\colon D(T) \cap D(S) \to Y, \quad (T+S)x = Tx + Sx.$$

Remark 3.2 ([Wer07, Theorem II.1.2]). Bounded operators are by definition continuous, but the converse is also true. If $T : D(T) \to Y$ is a linear map between Banach spaces and T were unbounded, then for all $n \in \mathbb{N}$ there would have to be a $x_n \in X$ such that $\|Tx_n\|_Y > n\,\|x_n\|_X$. Then

$$\left\| T \frac{x_n}{n\,\|x_n\|_X} \right\|_Y > \frac{n\,\|x_n\|_X}{n\,\|x_n\|_X} = 1,$$

but $\left\| \frac{x_n}{n\|x_n\|_X} \right\|_X = \frac{1}{n} \to 0$ so T can not be continuous in 0.

This chapter will mostly be restricted to bounded operators, but we also have to consider unbounded operators in some instances.

If not noted otherwise $D(T) = X$ in this thesis. Unless noted otherwise all Hilbert spaces in this thesis are assumed to be complex and separable.

For operators between Hilbert spaces the additional scalar product means that for each operator we can find an associated adjoint operator.

Definition 3.3 (Adjoint operator)
*Let $T\colon D(T) \subset V \to W$ be an operator between Hilbert spaces V, W where $D(T)$ is dense in V. The **adjoint operator** $T^*\colon D(T^*) \subset W \to V$ is defined by the equation*

$$\langle Tx, y\rangle_W = \langle x, T^*y\rangle_V \quad \forall x \in D(T), y \in D(T^*)$$

and $D(T^*)$ is the set of all $y \in W$ where such a T^*y exists for all $x \in D(T)$. Since $D(T)$ is dense, T^*y is unique. If $D(T) = V$ and T is bounded, then $D(T^*) = W$ by the Fréchet-Riesz representation theorem, and in this case T^* is also bounded [Wei80, Theorem 4.14].

Of special interest are operators that are their own adjoint.

Definition 3.4 (Self-adjoint operator)
*Let $T\colon D(T) \subset V \to V$ be an operator on a Hilbert space V. Then T is **symmetric** if*

$$\langle Tx, y\rangle_V = \langle x, Ty\rangle_V \quad \forall x, y \in D(T).$$

*T is called **self-adjoint** if T is symmetric, $D(T)$ is dense in V and $D(T) = D(T^*)$, meaning that $T = T^*$. Thus, any symmetric bounded operator with domain V is self-adjoint.*

For unbounded operators it is sometimes hard to prove self-adjointness, but there is an intermediate step that is sometimes useful.

Definition 3.5 (Essentially self-adjoint operator)
Let $T\colon D(T) \subset V \to V$ be an operator on a Hilbert space V. Then

$$\Gamma(T)\colon = \{(x, Tx) \mid x \in D(T)\} \subset V \times V$$

is the graph of T. If $\Gamma(T)$ is closed, then T is called closed as well. If T_1 and T_2 are operators on V and $\Gamma(T_1) \subset \Gamma(T_2)$, then T_2 is called an extension of T_1. T is called closable if it has a closed extension, and the smallest closed extension is called the closure of T.
If T is closed and there is a set $D \subset D(T)$ such that T restricted to D is closable with closure T, then D is called a core of T.
If $D(T)$ is dense and T is symmetric then T is closable, and if the closure of T is self-adjoint then T is called essentially self-adjoint.
If merely $TT^* = T^*T$ and T is closed, then T is called ***normal****.*

The advantage here is that an essentially self-adjoint operator has a unique self-adjoint extension. Furthermore, all self-adjoint operators are closed, so self-adjoint operators restricted to a dense set are essentially self-adjoint. For details, see [RS80, section VIII.2].
For our setting the most important operators are the discrete Laplace operator and multiplication operators. Following [Kir07] in part, we get the following result.

Lemma 3.6
Let X be $\mathbb{Z}^d$ or $[0, 2\pi]^d$. If $v : X \to \mathbb{R}$ is a function such that there is a $M \geq 0$ with $|v(x)| \leq M \ \forall x \in X$, then the multiplication operator

$$V\colon L^2(X) \to L^2(X),\ (V\phi)(x) = v(x)\phi(x)$$

is bounded and self-adjoint. Here we use the notation $L^2(\mathbb{Z}^d) = \ell^2(\mathbb{Z}^d)$.
The negative discrete Laplace operator $-\Delta\colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ with

$$-\Delta\phi(x) = \sum_{y \in \mathbb{Z}^d:\ \|y-x\|_1 = 1} (\phi(x) - \phi(y)) = \sum_{y \in \mathbb{Z}^d:\ \|y\|_1 = 1} (\phi(x) - \phi(x+y))$$

is bounded and self-adjoint. Additionally

$$\left(\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}u\right)(k) = \left(2d - 2\sum_{n=1}^{d} \cos(k^{(n)})\right) u(k)$$

holds for $u \in L^2([0, 2\pi]^d)$, where $\tilde{\mathcal{F}}$ is the Fourier transform defined in (3.2).
Consequently with $X = \mathbb{Z}^d$ above the operator $H = -\Delta + V$ is also bounded and self-adjoint in this case.

Proof. V is bounded since for $\phi \in L^2(X)$,

$$\|V\phi\|_{L^2(X)} = \|v \cdot \phi\|_{L^2(X)} \leq M \|\phi\|_{L^2(X)} .$$

The self-adjointness of V follows for $X = [0, 2\pi]^d$ from

$$\langle V\phi, \psi\rangle = \int_{[0,2\pi]^d} \overline{v(x)\phi(x)}\psi(x)\mathrm{d}x = \int_{[0,2\pi]^d} \overline{\phi(x)}v(x)\psi(x)\mathrm{d}x = \langle \phi, V\psi\rangle$$

and for $X = \mathbb{Z}^d$ from

$$\langle V\phi, \psi\rangle = \sum_{x\in\mathbb{Z}^d} \overline{v(x)\phi(x)}\psi(x) = \sum_{x\in\mathbb{Z}^d} \overline{\phi(x)}v(x)\psi(x) = \langle \phi, V\psi\rangle$$

since $v(x)$ is real.
For $-\Delta$ we use the Fourier transform

$$\tilde{\mathcal{F}} \colon \ell^2(\mathbb{Z}^d) \to L^2([0, 2\pi]^d),\ \tilde{\mathcal{F}}\phi(k) = \sum_{x\in\mathbb{Z}^d} \phi(x)\mathrm{e}^{\mathrm{i}x\cdot k}. \tag{3.2}$$

With the scalar product

$$\langle u_1, u_2\rangle_{L^2([0,2\pi]^d)} = \frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} \overline{u_1(k)}u_2(k)\mathrm{d}k$$

and the induced norm

$$\|u\|_{L^2\left([0,2\pi]^d\right)} = \frac{1}{(2\pi)^{d/2}} \sqrt{\int_{[0,2\pi]^d} |u(k)|^2 \,\mathrm{d}k}$$

this transform maps each orthonormal basis vector δ_x of $\ell^2(\mathbb{Z}^d)$ to an orthonormal basis vector $\mathrm{e}^{\mathrm{i}x\cdot k}$ of $L^2([0, 2\pi]^d)$. Thus, the transform is unitary and therefore invertible, bounded, has a bounded inverse and preserves scalar products (and norms), i.e. $\langle \tilde{\mathcal{F}}\phi, \tilde{\mathcal{F}}\psi\rangle_{L^2([0,2\pi]^d)} = \langle \phi, \psi\rangle_{\ell^2(\mathbb{Z}^d)}$ for all $\phi, \psi \in \ell^2(\mathbb{Z}^d)$. At the same time we have $\langle \tilde{\mathcal{F}}^{-1}u, \tilde{\mathcal{F}}^{-1}v\rangle_{\ell^2(\mathbb{Z}^d)} = \langle u, v\rangle_{L^2([0,2\pi]^d)}$ for all $u, v \in L^2([0, 2\pi]^d)$.

Let $u \in L^2([0,2\pi]^d)$ and $\phi := \tilde{\mathcal{F}}^{-1}u$. Then

$$\begin{aligned}
\left(\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}u\right)(k) &= \left(\tilde{\mathcal{F}}(-\Delta)\phi\right)(k) = \sum_{x\in\mathbb{Z}^d} (-\Delta\phi)(x)\mathrm{e}^{\mathrm{i}x\cdot k} \\
&= \sum_{x\in\mathbb{Z}^d} \sum_{y\in\mathbb{Z}^d\colon \|y\|_1=1} (\phi(x) - \phi(x+y))\,\mathrm{e}^{\mathrm{i}x\cdot k} \\
&= \sum_{x\in\mathbb{Z}^d} 2d\ \phi(x)\mathrm{e}^{\mathrm{i}x\cdot k} - \sum_{y\in\mathbb{Z}^d\colon \|y\|_1=1} \sum_{x\in\mathbb{Z}^d} \phi(x+y)\mathrm{e}^{\mathrm{i}x\cdot k} \\
&= 2d\ \tilde{\mathcal{F}}\phi(k) - \sum_{y\in\mathbb{Z}^d\colon \|y\|_1=1} \sum_{x\in\mathbb{Z}^d} \phi(x)\mathrm{e}^{\mathrm{i}x\cdot k}\mathrm{e}^{-\mathrm{i}y\cdot k} \\
&= 2d\ u(k) - \sum_{y\in\mathbb{Z}^d\colon \|y\|_1=1} \mathrm{e}^{-\mathrm{i}y\cdot k} \sum_{x\in\mathbb{Z}^d} \phi(x)\mathrm{e}^{\mathrm{i}x\cdot k} \\
&= 2d\ u(k) - \sum_{y\in\mathbb{Z}^d\colon \|y\|_1=1} \mathrm{e}^{-\mathrm{i}y\cdot k}\ u(k) \\
&= 2d\ u(k) - \sum_{n=1}^{d} \left(\mathrm{e}^{\mathrm{i}k^{(n)}} + \mathrm{e}^{-\mathrm{i}k^{(n)}}\right) u(k) = \left(2d - 2\sum_{n=1}^{d}\cos(k^{(n)})\right) u(k)
\end{aligned}$$

where we used an index shift in line four and $k^{(n)} = e_n \cdot k$ is the n-th coordinate of k. This proves that $\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}$ is a multiplication operator on $L^2([0,2\pi]^d)$. Since

$$\left\{2d - 2\sum_{n=1}^{d}\cos(k^{(n)}) \mid k \in [0,2\pi]^d\right\} = [0,4d].$$

this multiplication operator is bounded. Since $\tilde{\mathcal{F}}$ preserves norms we have for all $\phi \in \ell^2(\mathbb{Z}^d)$

$$\|(-\Delta)\phi\|_{\ell^2(\mathbb{Z}^d)} = \left\|\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}u\right\|_{L^2} \leq 4d\,\|u\|_{L^2} = 4d\,\|\phi\|_{\ell^2}$$

where $u = \tilde{\mathcal{F}}\phi$. Thus, $(-\Delta)$ is bounded with $\|-\Delta\| \leq 4d$.
The bounded multiplication operator $\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}$ is symmetric and therefore self-adjoint, and since for all $\phi, \psi \in \ell^2(\mathbb{Z}^d)$ we have

$$\begin{aligned}
\langle(-\Delta)\phi,\psi\rangle_{\ell^2(\mathbb{Z}^d)} &= \langle\tilde{\mathcal{F}}(-\Delta)\phi, \tilde{\mathcal{F}}\psi\rangle_{L^2} = \langle\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}u, v\rangle_{L^2} \\
&= \langle u, \tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}v\rangle_{L^2} = \langle\tilde{\mathcal{F}}\phi, \tilde{\mathcal{F}}(-\Delta)\psi\rangle_{L^2} \\
&= \langle\phi, (-\Delta)\psi\rangle_{\ell^2(\mathbb{Z}^d)}
\end{aligned}$$

with $u = \tilde{\mathcal{F}}\phi$ and $v = \tilde{\mathcal{F}}\psi$, we get that $-\Delta$ is also self-adjoint. □

3.2 The spectrum

Now we want to define analogues of eigenvalues for infinite dimensional operators. We keep the idea of looking at whether $T - \lambda I$ is invertible, where I is now the identity operator on the Hilbert space H.

Definition 3.7 (Spectrum and resolvent set)
Let H be a Hilbert space and $T\colon D(T) \subset H \to H$ an operator. Then

$$\rho(T) = \big\{\lambda \in \mathbb{C} \mid (T - \lambda I) \text{ is bijective and } (T - \lambda I)^{-1} \text{ is bounded}\big\}$$

*is called the **resolvent set of** T. Its complement*

$$\sigma(T) = \mathbb{C} \setminus \rho(T)$$

*is called the **spectrum of** T.*

For finite dimensional operators the spectrum is exactly the set of eigenvalues. Additionally if λ belongs to the resolvent set then $(T-\lambda I)^{-1}$ is linear since $T-\lambda I$ is linear. One importance consequence of self-adjointness for the spectrum lies in the following Lemma.

Lemma 3.8 ([Wer07, Lemma VII.1.1 and Corollary VII.1.2])
If $T : H \to H$ is a bounded linear operator on a Hilbert space H, then

$$\sigma(T) \subset \overline{W(T)},$$

where

$$W(T) = \{\langle Tx, x\rangle \mid \|x\| = 1\}$$

*is the **numerical range** of T and $\overline{W(T)}$ is the closure of the set $W(T)$.*
Moreover, if T is self-adjoint then $\sigma(T) \subset \mathbb{R}$.

This supports why operators associated to observables should be self-adjoint. The spectrum can be said to contain the possible results of measurements of the application of an operator, and for self-adjoint operators this is thus ensured to be real. Since measurements of physical quantities in quantum theory can only yield real numbers this is a reasonable restriction.

Remark 3.9. If $B : H_1 \to H_2$ is a bounded linear operator between Hilbert spaces with bounded inverse and $T : H_1 \to H_1$ is a linear bounded operator, then $\sigma(T) = \sigma(BTB^{-1})$ since $BTB^{-1} - \lambda I = B(T - \lambda I)B^{-1}$ is invertible with bounded inverse if and only if this is true for $T - \lambda I$.

We want to calculate the spectrum of Anderson operators now, so we will start out by calculating the spectrum of multiplication operators and the discrete Laplacian separately, with some ideas taken from [Kir07].

Example 3.10 ([Sch12a, Example 3.8]/ [Kir07])
Let X be a measure space (see Section 2.4) with a σ-finite measure P, $v\colon X \to \mathbb{R}$ a measurable map and $V\colon D(V) \to L^2(X), V\varphi(x) = v(x)\varphi(x)$ with $D(V) = \{\varphi \in L^2(X) \mid v\varphi \in L^2(X)\}$. Then

$$\sigma(V) = \{\lambda \in \mathbb{C} \mid \forall \varepsilon > 0 : P(x \in X \mid |v(x) - \lambda| < \varepsilon) \neq 0\} =: A. \tag{3.3}$$

*The set A is also called the **essential range** of v with respect to P.*
From this follows that

1. *the spectrum of $V : D(V) \to \ell^2(\mathbb{Z}^d)$, $(V\phi)(x) = v(x)\phi(x)$ with a function $v : \mathbb{Z}^d \to \mathbb{R}$ is*
$$\sigma(V) = \overline{\operatorname{ran}\, v}$$
where $\operatorname{ran}\, v = \{y \in \mathbb{R} \mid \exists x \in \mathbb{Z}^d : v(x) = y\}$ *is the **range of** v,*
2. *the spectrum of the negative discrete Laplace operator $-\Delta : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ is*
$$\sigma(-\Delta) = [0, 4d].$$

Proof. First, if $\lambda \notin A$, then there is an $\varepsilon > 0$ such that $P(x \in X \mid |v(x) - \lambda| < \varepsilon) = 0$, meaning that $P(X \setminus B) = 0$ where

$$B := \left\{x \in X \mid \frac{1}{|v(x) - \lambda|} \leq \frac{1}{\varepsilon}\right\} \cap \{x \in X \mid |v(x) - \lambda| \neq 0\}.$$

The operator

$$W\colon L^2(X) \to L^2(X), W\varphi(x) := \begin{cases} \dfrac{1}{v(x) - \lambda}\varphi(x) & , x \in B \\ 0 & , x \notin B \end{cases}$$

is therefore well defined, bounded and satisfies

$$(V - \lambda I)W\varphi(x) = \begin{cases} \varphi(x) & , x \in B \\ 0 & , x \notin B \end{cases}$$

for all $\varphi \in L^2(X)$ and $x \in X$, and $W(V - \lambda I)\varphi = (V - \lambda I)W\varphi$ for all $\varphi \in D(V)$. As a consequence of $P(X \setminus B) = 0$ we have $\|(V - \lambda I)W\varphi - \varphi\|_{L^2} = 0$ and therefore $W = (V - \lambda I)^{-1}$, resulting in $\lambda \in \rho(V)$.

Now we show that if $\lambda \in A$, then $\lambda \in \sigma(V)$. If $\lambda \in A$, then for all $\varepsilon > 0$ since P is σ-finite there is an event B_ε with $\infty > P(B_\varepsilon) > 0$ and

$$B_\varepsilon \subset \{x \in X \mid |v(x) - \lambda| < \varepsilon\}.$$

As before, we have two cases now:
If also $P(B_0) > 0$ with $B_0 := \{x \in X \mid |v(x) - \lambda| = 0\}$ then we have $(V - \lambda I)\varphi_0 = 0$ for the function $\varphi_0 = \mathbb{1}_{B_0} \not\equiv 0$, so $V - \lambda I$ is not injective and $\lambda \in \sigma(V)$. On the other hand, if $P(B_0) = 0$ then we define

$$\varphi_\varepsilon := \frac{1}{\sqrt{P(B_\varepsilon)}} \mathbb{1}_{B_\varepsilon \cap B_0^C},$$

so that $\|\varphi_\varepsilon\|_{L^2} = 1$ and $(V - \lambda I)\varphi_\varepsilon(x) = (v(x) - \lambda)\varphi_\varepsilon(x) \neq 0$ for all $x \in B_\varepsilon \cap B_0^C$ and 0 otherwise. If a $W = (V - \lambda I)^{-1}$ exists, then $W\varphi_\varepsilon(x) = \frac{1}{v(x)-\lambda}\varphi_\varepsilon(x) \geq \frac{1}{\varepsilon}\varphi_\varepsilon(x)$ has to be true for all $x \in B_\varepsilon \cap B_0^C$. This would mean $\|W\| > \frac{1}{\varepsilon}$ for all $\varepsilon > 0$, so W would be unbounded and $\lambda \in \sigma(V)$. This proves (3.3).
We can now apply this result to the two special cases mentioned before.

1. Here, we choose $X = \mathbb{Z}^d$ with the counting measure, i.e. $L^2(X) = \ell^2(\mathbb{Z}^d)$. Then

$$P\left(x \in \mathbb{Z}^d \mid |v(x) - \lambda| < \varepsilon\right) \neq 0 \Leftrightarrow \exists x \in \mathbb{Z}^d : |v(x) - \lambda| < \varepsilon,$$

and therefore

$$\lambda \in \sigma(V) \Leftrightarrow \lambda \in \overline{\operatorname{ran} v}.$$

2. If $X = [0, 2\pi]^d$ and the measure on $[0, 2\pi]^d$ is the Lebesgue measure, then

$$P\left(x \in [0, 2\pi]^d \mid |v(x) - \lambda| < \varepsilon\right) \neq 0 \Leftrightarrow \exists x \in [0, 2\pi]^d : |v(x) - \lambda| < \varepsilon$$

is true if v is continuous, since open balls have positive Lebesgue-measure. As $[0, 2\pi]^d$ is compact we also have

$$\sigma(V) = \overline{\operatorname{ran} v} = \operatorname{ran} v$$

if v is continuous.
As already shown in Lemma 3.6 the negative Laplace operator can be transformed into a multiplication operator on $L^2([0, 2\pi]^d)$ (with the Lebesgue measure) via the Fourier transform, and

$$\left(\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}u\right)(k) = \left(2d - 2\sum_{n=1}^{d} \cos(k^{(n)})\right) u(k)$$

for all $u \in L^2([0,2\pi]^d)$. Since $\left(2d - 2\sum_{n=1}^{d} \cos(k^{(n)})\right)$ is continuous we have

$$\sigma(\tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}) = \left\{2d - 2\sum_{n=1}^{d} \cos(k^{(n)}) \mid k \in [0,2\pi]^d\right\} = [0,4d].$$

By using Remark 3.9 with the Fourier transform we arrive at the stated result.

□

3.3 Random potentials

The previous section covered the spectrum of deterministic multiplication operators, now we want to extend this to random multiplication operators as well. The Anderson model features a multiplication with a real random variable $v.(z)$ for every point $z \in \mathbb{Z}^d$. For now we will assume that $(v.(z))_{z\in\mathbb{Z}^d}$ is a sequence of i.i.d. random variables with shared image measure $\mathbb{P}_0$ on $\mathbb{R}$.
Since the random potentials used for the Anderson operator are all real, we can also define the support of the image measure. For later use we will note some properties of the support of a real image measure here.

Lemma 3.11
Let $\mathbb{P}_0$ be some measure on $\mathbb{R}$. Then for the ***support of the image measure****,*

$$\text{supp } \mathbb{P}_0 := \{x \in \mathbb{R} \mid \forall\varepsilon > 0\colon \mathbb{P}_0\left((x-\varepsilon, x+\varepsilon)\right) > 0\}$$

we have

- supp $\mathbb{P}_0$ *is closed*
- $\mathbb{P}_0\left(\text{supp } \mathbb{P}_0\right) = 1$
- supp $\mathbb{P}_0$ *contains (like any subset of $\mathbb{R}$) a countable subset D that is dense in* supp $\mathbb{P}_0$*, i.e. for all $\varepsilon > 0$ and $p \in$ supp $\mathbb{P}_0$ there exists a $p' \in D$ such that $|p - p'| < \varepsilon$.*

These properties hold in much greater generality than just for measures in $\mathbb{R}$, but we only need them for this special case.

Proof.
- If $x \in (\text{supp } \mathbb{P}_0)^C$ then there is an $\varepsilon > 0$ such that $\mathbb{P}_0\left((x-\varepsilon, x+\varepsilon)\right) = 0$. Now for every $q \in \mathbb{R}$ with $|x - q| < \varepsilon/2$ we have $(q - \varepsilon/2, q + \varepsilon/2) \subset (x-\varepsilon, x+\varepsilon)$, and thus $\mathbb{P}_0\left((q-\varepsilon/2, q+\varepsilon/2)\right) = 0$ as well. Therefore, $(\text{supp } \mathbb{P}_0)^C$ is open and supp $\mathbb{P}_0$ is closed.

- We take a look at the complement and see that

$$\begin{aligned}(\text{supp } \mathbb{P}_0)^C &= \{x \in \mathbb{R} \mid \exists \varepsilon > 0 : \mathbb{P}_0((x-\varepsilon, x+\varepsilon)) = 0\} \\ &= \{x \in \mathbb{R} \mid \exists a, b \in \mathbb{R} : x \in (a,b),\ \mathbb{P}_0((a,b)) = 0\} \\ &= \{x \in \mathbb{R} \mid \exists a, b \in \mathbb{Q} : x \in (a,b),\ \mathbb{P}_0((a,b)) = 0\} \\ &= \bigcup \{(a,b) \mid a, b \in \mathbb{Q},\ \mathbb{P}_0((a,b)) = 0\}\end{aligned}$$

  The last line is a countable union of sets of probability zero, and thus the complement of supp $\mathbb{P}_0$ has probability zero as well. This proves $\mathbb{P}_0(\text{supp } \mathbb{P}_0) = 1$.

- To see this consider for example all intervals $I_{r,q}$ with a rational midpoint r and rational length q. If $I_{r,q} \cap$ supp $\mathbb{P}_0 \neq \emptyset$ we can choose one point from the intersection and label it $x_{r,q}$. Now for a given $x \in$ supp $\mathbb{P}_0$ and $\varepsilon > 0$ there are $r, q \in \mathbb{Q}$ such that $I_{r,q}$ contains x and is also contained in an ε-interval around x. Thus, $|x - x_{r,q}| < \varepsilon$ and $D = \{x_{r,q} \mid r, q \in \mathbb{Q}\}$ is dense in supp $\mathbb{P}_0$ and countable.

□

To stay in the space of bounded operators we will assume in this chapter that the random variables are almost surely uniformly bounded, i.e. that supp $\mathbb{P}_0$ is bounded. We have infinitely many independent identically distributed random variables, so we can even conclude that if we choose a finite subset of $\mathbb{Z}^d$ and fix a given value in supp $\mathbb{P}_0$ for each point in the subset, then nearly exactly this configuration will almost surely appear somewhere, and it will even appear infinitely often. This is shown in the following Lemma.

Lemma 3.12 ([Kir07, Proposition 3.8])
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(v_.(z))_{z\in\mathbb{Z}^d}$ a sequence of i.i.d. real random variables with an image measure $\mathbb{P}_0$ with bounded support. Then there is a set $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that for every $\omega \in \Omega_0$, every finite $\Lambda \subset \mathbb{Z}^d$, every sequence $(q_i)_{i\in\Lambda}$ in supp $\mathbb{P}_0$ *and every $\varepsilon > 0$ there exists a sequence $(j_n)_{n\in\mathbb{N}}$ in $\mathbb{Z}^d$ such that $\|j_n\|_\infty = \max_{1\le i\le d} \left|j_n^{(i)}\right| \to \infty$ (where $j_n^{(i)}$ is the i-th coordinate of $j_n \in \mathbb{Z}^d$) and*

$$\sup_{i\in\Lambda} |q_i - v_\omega(i + j_n)| < \varepsilon \quad \forall n \in \mathbb{N}.$$

Proof. Let us fix Λ, $(q_i)_{i\in\Lambda}$ and ε as given in the statement. We know that the event

$$A := \left\{\omega \in \Omega \mid \sup_{i\in\Lambda} |q_i - v_\omega(i)| < \varepsilon\right\}$$

has probability $\mathbb{P}(A) > 0$, since $\mathbb{P}(|q_i - v_\omega(i)| < \varepsilon) > 0$ for all $i \in \Lambda$ by the definition of the support of the image measure and all random variables are independent. We pick a sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{Z}^d$ such that $\|k_i - k_j\|_\infty > 2 \operatorname{diam} \Lambda$ for all $i \neq j$, where $\operatorname{diam} \Lambda := \max_{x,y\in\Lambda} \|x-y\|_\infty$ is the diameter of Λ. This way we ensure that $\{x + k_i \mid x \in \Lambda\} \cap \{x + k_j \mid x \in \Lambda\} = \emptyset$ for all $i \neq j$. Then the sequence $(A_n)_{n\in\mathbb{N}}$ with

$$A_n := \left\{\omega \in \Omega \mid \sup_{i\in\Lambda} |q_i - v_\omega(i + k_n)| < \varepsilon\right\}$$

is independent and we have $\mathbb{P}(A_n) = \mathbb{P}(A) > 0$ for all $n \in \mathbb{N}$, leading to $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A) = \infty$. Now we use the second part of the Borel-Cantelli-Lemma 2.9 to conclude that the event

$$\Omega_{\Lambda,(q_i)_{i\in\Lambda},\varepsilon} := \{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}\} = \limsup_{n\to\infty} A_n$$

has probability one.
We have thus shown that for a specific choice of Λ, $(q_i)_{i\in\Lambda}$ and ε we almost surely find approximately the same configuration at infinitely many points. We want a single set Ω_0 that works for any choice, so we would hypothetically need to form the intersection over all $\Omega_{\Lambda,(q_i)_{i\in\Lambda},\varepsilon}$. But we can only be sure that the intersection of countably many sets of probability one has probability one itself, and there are uncountably many choices for both the q_i and ε. The set Υ of all finite subsets of $\mathbb{Z}^d$ is countable, so this is not a problem. By the third point of Lemma 3.11 the support of the image measure $\operatorname{supp} \mathbb{P}_0$ contains a countable set D dense in this subset, so we restrict $(q_i)_{i\in\Lambda}$ to this subset.
For the ε we do not actually need an intersection over all epsilon, since choosing a smaller ε is always possible. Thus, we just have to consider $\varepsilon_n = \frac{1}{n}$.
This way we form the set

$$\Omega_0 := \bigcap_{\Lambda\in\Upsilon,(d_i)\in D^\Lambda,n\in\mathbb{N}} \Omega_{\Lambda,(d_i)_{i\in\Lambda},1/n},$$

with $\mathbb{P}(\Omega_0) = 1$. By uniformly approximating any sequence $(q_i)_{i\in\Lambda}$ in $\operatorname{supp} \mathbb{P}_0$ by a sequence $(d_i)_{i\in\Lambda}$ in D we see that Ω_0 fulfils all criteria from the statement. $\square$

We can now use the preceding lemma to calculate the spectrum of the random multiplication operator V_ω, the potential of the Anderson operator.

Corollary 3.13 ([Kir07])
Let $(v_\cdot(z))_{z\in\mathbb{Z}^d}$ be a sequence of i.i.d. real random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with an image measure $\mathbb{P}_0$ with bounded support. Then the spectrum of the random multiplication operator $V_\omega \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d), V_\omega\phi(z) := v_\omega(z)\phi(z)$ is almost surely the support of $\mathbb{P}_0$, i.e. there is a set Ω_1 of measure one such that

$$\sigma(V_\omega) = \operatorname{supp} \mathbb{P}_0 \ \forall \omega \in \Omega_1.$$

Proof. First, if $x \in \text{supp } \mathbb{P}_0$ we can use Lemma 3.12 for each $n \in \mathbb{N}$ with $\Lambda = \{0\}$, $q_0 = x$ and $\varepsilon = 1/n$ to get a point $j_n \in \mathbb{Z}^d$ such that

$$|x - v_\omega(j_n)| < \frac{1}{n}$$

and all $\omega \in \Omega_0$, just by taking the first point in the sequences constructed in the lemma. The thus formed sequence $(j_n)_{n\in\mathbb{N}}$ satisfies $|x - v_\omega(j_n)| \xrightarrow[n\to\infty]{} 0$ and we have $x \in \overline{\text{ran } v_\omega} = \sigma(V_\omega)$ for all $\omega \in \Omega_0$ by Example 3.10.
To prove the other direction we will use that supp $\mathbb{P}_0$ is closed and $\mathbb{P}_0(\text{supp } \mathbb{P}_0) = 1$ by the first and second property of Lemma 3.11. By the definition of the image measure this means that for each $z \in \mathbb{Z}^d$ we have $\mathbb{P}(v_\omega(z) \in \text{supp } \mathbb{P}_0) = 1$ and as a countable intersection of sets of probability one also that

$$\Omega' := \bigcap_{z\in\mathbb{Z}^d} \{\omega \in \Omega \mid v_\omega(z) \in \text{supp } \mathbb{P}_0\} = \{\omega \in \Omega \mid \text{ran } v_\omega \subset \text{supp } \mathbb{P}_0\}$$

has probability one. Since supp $\mathbb{P}_0$ is closed, we get $\sigma(V_\omega) = \overline{\text{ran } v_\omega} \subset \text{supp } \mathbb{P}_0$ for all $\omega \in \Omega'$ by Example 3.10. With $\Omega_1 = \Omega_0 \cap \Omega'$ follows the claimed result. □

We need one last ingredient before we can calculate the spectrum of the Anderson operator, and that is the so called Weyl criterion, which shows that even though not every point in the spectrum is an eigenvalue with an associated eigenfunction, we can nevertheless find a sequence of functions that behaves approximately like eigenfunctions, and the converse is also true.

Theorem 3.14 (Weyl criterion, Weidmann [Wei80, Theorem 7.22])
Let $T : D(T) \subset H \to H$ be a self-adjoint operator on a Hilbert space H. Then

$$\lambda \in \sigma(T) \Leftrightarrow \exists \varphi_n \in D(T): \ \|\varphi_n\|_H = 1, \ \|(T - \lambda I)\varphi_n\|_H \to 0, \tag{3.4}$$

Such a sequence is called a ***Weyl sequence****.*
If $H' \subset D(T)$ is dense then the sequence $(\varphi_n)_{n\in\mathbb{N}}$ can also be chosen to lie in H'.

Finally we can show what the spectrum of the Anderson operator is:

Theorem 3.15 ([Kir07, Theorem 3.9])
Let $(v_\cdot(z))_{z\in\mathbb{Z}^d}$ be a sequence of i.i.d. real random variables with an image measure $\mathbb{P}_0$ with bounded support. Then the spectrum of the ***Anderson operator***
$H_\omega = -\Delta + V_\omega \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ with

$$H_\omega \Psi(z) = -\Delta\Psi(z) + v_\omega(z)\Psi(z),$$

where Δ is the discrete Laplacian, is almost surely

$$\sigma(H_\omega) = \text{supp } \mathbb{P}_0 + [0, 4d] = \{x + y \ \in \mathbb{R} \mid x \in \text{supp } \mathbb{P}_0, y \in [0, 4d]\}.$$

Proof. As shown in Example 3.10 the negative discrete Laplace operator is transformed into a multiplication operator with range $[0,4d]$ by the Fourier transform. For each $\phi \in \ell^2(\mathbb{Z}^d)$ we have thus

$$\langle -\Delta\phi, \phi\rangle = \langle \tilde{\mathcal{F}}(-\Delta)\tilde{\mathcal{F}}^{-1}\tilde{\mathcal{F}}\phi, \tilde{\mathcal{F}}\phi\rangle \in [0, 4d\,\|\phi\|_{\ell^2}^2],$$

since the Fourier transform preserves norms and scalar products. An operator A on a Hilbert space H such that $\langle A\phi, \phi\rangle \geq 0$ for all $\phi \in D(A)$ is called **non-negative** and by the preceding argument we see that $-\Delta$ is indeed non-negative. If A is a bounded non-negative self-adjoint operator and B is a self-adjoint operator on the same Hilbert space H then as shown by Seelmann in Corollary 2.2 of [See19] the spectrum of $A+B$ is linked to the spectrum of B and $\|A\|$ via

$$\sigma(A+B) \subset \sigma(B) + [0, \|A\|]. \tag{3.5}$$

We will apply this result with $A = -\Delta$ and $B = V_\omega$. We already know from Corollary 3.13 that the spectrum of V_ω is almost surely supp $\mathbb{P}_0$, from Example 3.10 that the spectrum of $-\Delta$ is $[0,4d]$ and from Lemma 3.6 that $\|-\Delta\| \leq 4d$. From this, the non-negativity of $-\Delta$ and Lemma 3.8 follows

$$[0,4d] = \sigma(-\Delta) \subset \overline{W(-\Delta)} \subset [0,4d]$$

where $W(-\Delta)$ is the numerical range of $-\Delta$. Thus, there has to be a sequence $(\phi_n)_{n\in\mathbb{N}}$ in $\ell^2(\mathbb{Z}^d)$ with $\|\phi_n\|_{\ell^2} = 1$ for all $n \in \mathbb{N}$ such that

$$\langle -\Delta\phi_n, \phi_n\rangle \to 4d,$$

which proves that the norm of $-\Delta$ is indeed $4d$.
By (3.5) we get the inclusion

$$\sigma(H_\omega) \subset \text{supp}\ \mathbb{P}_0 + [0,4d] \tag{3.6}$$

for almost all $\omega \in \Omega$.
Note that (3.6) follows directly from Lemma 3.8 if supp $\mathbb{P}_0$ is an interval. In this case $\overline{W(V_\omega)} = \text{supp}\ \mathbb{P}_0$ holds almost surely, since every point in

$$W(V_\omega) = \left\{ \sum_{z\in\mathbb{Z}^d} v_\omega(z)\phi^2(z) \mid \phi \in \ell^2(\mathbb{Z}^d) \right\}$$

is a convex combination of all points in $\overline{\text{ran}\ v_\omega} = \text{supp}\ \mathbb{P}_0$, which is already convex if the support is an interval. Then (3.6) follows from

$$\sigma(H_\omega) \subset \overline{W(H_\omega)} \subset \overline{W(-\Delta) + W(V_\omega)} = [0,4d] + \text{supp}\ \mathbb{P}_0.$$

For the converse inclusion let $\lambda \in \operatorname{supp} \mathbb{P}_0 + [0,4d] = \operatorname{supp} \mathbb{P}_0 + \sigma(-\Delta)$. Then there are $\lambda_1 \in \operatorname{supp} \mathbb{P}_0$ and $\lambda_2 \in \sigma(-\Delta)$ such that $\lambda = \lambda_1 + \lambda_2$. The space

$$\ell_0^2(\mathbb{Z}^d) = \{\varphi \in \ell^2(\mathbb{Z}^d) \mid \varphi(x) \neq 0 \text{ for only finitely many } x \in \mathbb{Z}^d\} \tag{3.7}$$

is dense in $\ell^2(\mathbb{Z}^d)$, so we can use Weyls criterion from Theorem 3.14 for λ_2 to get a sequence $(\varphi_n)_{n\in\mathbb{N}}$ in $\ell_0^2(\mathbb{Z}^d)$ such that $\|\varphi_n\|_{\ell^2(\mathbb{Z}^d)} = 1$ and

$$\|(-\Delta - \lambda_2 I)\varphi_n\|_{\ell^2(\mathbb{Z}^d)} \to 0.$$

For λ_1 we use Lemma 3.12 with $\Lambda_n = \{x \in \mathbb{Z}^d \mid \varphi_n(x) \neq 0\}$ and $q_i = \lambda_1$ for all $i \in \Lambda_n$. By the statement of the lemma there is an event Ω_0 such that for every $n \in \mathbb{N}$ there is a j_n in $\mathbb{Z}^d$ with

$$\sup_{i\in\Lambda_n} |\lambda_1 - v_\omega(i + j_n)| < \frac{1}{n}$$

for all $\omega \in \Omega_0$.
The discrete Laplace operator is translation invariant, so for every $\psi \in \ell^2(\mathbb{Z}^d)$ and every **translation**

$$\tau_y \colon \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d), \tau_y\psi(x) := \psi(x-y) \tag{3.8}$$

by $y \in \mathbb{Z}^d$ we have

$$-\Delta(\tau_y(\psi)) = \tau_y(-\Delta\psi) \tag{3.9}$$

which can be checked by just inserting the relevant definitions. Furthermore $\|\varphi\|_{\ell^2(\mathbb{Z}^d)} = \|\tau_y\varphi\|_{\ell^2(\mathbb{Z}^d)}$ for all $\varphi \in \ell^2(\mathbb{Z}^d)$ and $y \in \mathbb{Z}^d$.
We set $\psi_n := \tau_{j_n}\varphi_n$. By the translation invariance of the norm we get $\|\psi_n\|_{\ell^2(\mathbb{Z}^d)} = 1$ as well as

$$\begin{aligned}
\|(-\Delta - \lambda_2 I)\psi_n\|_{\ell^2(\mathbb{Z}^d)} &= \|(-\Delta - \lambda_2 I)\tau_{j_n}\varphi_n\|_{\ell^2(\mathbb{Z}^d)} \\
&= \|\tau_{j_n}(-\Delta\varphi_n) - \tau_{j_n}(\lambda_2\varphi_n)\|_{\ell^2(\mathbb{Z}^d)} \\
&= \|(-\Delta - \lambda_2)\varphi_n\|_{\ell^2(\mathbb{Z}^d)} \to 0
\end{aligned}$$

by the linearity of the translation.
By the definition of ψ_n we also see that

$$\begin{aligned}
\|(V_\omega - \lambda_1)\psi_n\|^2_{\ell^2(\mathbb{Z}^d)} &= \sum_{x\in\mathbb{Z}^d} |(V_\omega(x) - \lambda_1)(\tau_{j_n}\varphi_n)(x)|^2 \\
&= \sum_{x\in\mathbb{Z}^d} |(V_\omega(x) - \lambda_1)\varphi_n(x - j_n)|^2 \\
&= \sum_{x\in\Lambda_n} |(V_\omega(x + j_n) - \lambda_1)\varphi_n(x)|^2 \\
&< \frac{1}{n^2} \|\varphi_n\|^2_{\ell^2(\mathbb{Z}^d)} = \frac{1}{n^2},
\end{aligned}$$

for all $\omega \in \Omega_0$. Combining the last two results leads to

$$\|(-\Delta + V_\omega - \lambda I)\psi_n\|_{\ell^2(\mathbb{Z}^d)} \leq \|(-\Delta - \lambda_2 I)\psi_n\|_{\ell^2(\mathbb{Z}^d)} + \|(V_\omega - \lambda_1 I)\psi_n\|_{\ell^2(\mathbb{Z}^d)} \to 0,$$

and thus $(\psi_n)_{n\in\mathbb{N}}$ is a Weyl sequence for $\lambda = \lambda_1 + \lambda_2$. As a consequence of Theorem 3.14 we have $\lambda \in \sigma(-\Delta + V_\omega)$ for all $\omega \in \Omega_0$. Since both inclusions hold almost surely, the proof is finished. □

This result is interesting, since we showed that the spectrum is actually almost surely not random, despite the large amount of randomness that can enter the potentials! The randomness "self-averages" by the large number of random variables. This phenomenon also appears for the integrated density of states, but to define it we first need to introduce the spectral theorem.

3.4 The spectral theorem

One important use of the spectrum of self-adjoint operators is that it can be used to decompose the space on which the operator acts in the same way that the normed eigenvectors of a Hermitian matrix form an orthogonal basis of the (finite dimensional) space on which the matrix acts. But since the space is no longer required to be finite dimensional this is accomplished by spectral families, which are an extension of the spectral decomposition for matrices, instead of projections onto eigenvectors.

Definition 3.16 (Spectral family, [Wei80, (7.11)])
A ***spectral family*** *(also called a* ***resolution of the identity****) is a map E from the real numbers to the bounded operators on some Hilbert space H which satisfies the following conditions:*

1. *For each $t \in \mathbb{R}$, $E(t)$ is an orthogonal projection, meaning that*

$$E(t)^2 = E(t) = E(t)^*$$

2. *Whenever $s \leq t$ we have $\langle x, E(s)x\rangle \leq \langle x, E(t)x\rangle$ for all $x \in H$*
3. *For all $x \in H$ and all $t \in \mathbb{R}$ we have $E(t+\delta)x \to E(t)x$ as $\delta \searrow 0$*
4. *For all $x \in H$ we have $E(t)x \to 0$ as $t \to -\infty$*
5. *For all $x \in H$ we have $E(t)x \to x$ as $t \to \infty$*

If E is a spectral family, then it is also possible to define $E(t-)$ for all $t \in \mathbb{R}$ via $E(t-)x := \lim_{\delta \searrow 0} E(t-\delta)x$.
Furthermore since all $E(t)$ are orthogonal projections, the second condition is equivalent to

$$E(t)E(s) = E(\min(t,s)).$$

Based on this we have

$$(E(t_2) - E(t_1))\,(E(s_2) - E(s_1)) = 0$$

if $[t_1, t_2]$ and $[s_1, s_2]$ are disjoint and we further define $E((a,b]) = E(b) - E(a)$, $E((a,b)) = E(b-) - E(a)$, $E([a,b]) = E(b) - E(a-)$ and $E([a,b)) = E(b-) - E(a-)$ with $a < b \in \mathbb{R}$. This allows us to define a projection-valued measure.
For every spectral family E and every $x \in H$ the function

$$r_x \colon \mathbb{R} \to [0, \|x\|], \quad r_x(t) := \langle x, E(t)x \rangle = \|E(t)x\|^2$$

is non-decreasing, right-continuous and bounded and therefore the cumulative distribution function of a measure ρ_x.

Definition 3.17 (E-measurable functions, spectral integral)
*Given a spectral family E we call a function $u \colon \mathbb{R} \to \mathbb{C}$ **E-measurable** if for each $x \in H$ there is a sequence $(u_n)_{n\in\mathbb{N}}$ of step functions such that $u_n(t) \to u(t)$ for ρ_x-almost all t. A step function s (see also Definition 2.3) is a function of the type*

$$s = \sum_{i=1}^{k} c_i \mathbb{1}_{A_i}$$

where $k \in \mathbb{N}$, $c_i \in \mathbb{C}\ \forall 1 \leq i \leq k$ and each $A_i \subseteq \mathbb{R}$ is either an open, closed or half open interval, a single point or empty.

Continuous functions, step functions and Borel-measurable functions are E-measurable for all spectral families E. We can use this to define an integral of such a u with regard to the spectral family E to gain an operator on H.
To do this, we first define for a step function $s = \sum_{k=1}^{n} c_k \mathbb{1}_{I_k}$ a linear integral

$$\int s(t) \mathrm{d}E(t) := \sum_{k=1}^{n} c_k E(I_k).$$

This is now a linear operator on H and

$$\left\| \int s(t) \mathrm{d}E(t)x \right\|^2 = \sum_{k=1}^{n} |c_k|^2 \|E(I_k)x\|^2 = \sum_{k=1}^{n} |c_k|^2 r_x(I_k) = \int |s(t)|^2 \mathrm{d}\rho_x(t).$$

This integral can then be extended to $u \in L^2(\rho_x)$ by the same approximation as in Definition 2.3. More information on the construction and existence of this integral can be found in Chapter 7.2 of [Wei80]. It is also possible to start the construction of the integral with spectral measures instead and derive spectral families from them. This can for example be found in Chapter 4.2 of [Sch12a].
Since this integral is constructed from a spectral family, it is itself an operator on H.

Theorem 3.18 ([Wei80, Theorem 7.14])
Let E be a spectral family on H and $u\colon \mathbb{R} \to \mathbb{C}$ be E-measurable. Then

$$\hat{E}(u)x = \int u(t)\mathrm{d}E(t)x$$

for x in

$$D(\hat{E}(u)) = \{x \in H \mid u \in L^2(\rho_x)\}$$

defines a normal operator on H.
This operator is also written as

$$\hat{E}(u) = \int u(t)\mathrm{d}E(t)$$

If u is real-valued, then $\hat{E}(u)$ is self-adjoint.

Thus, a spectral family can be used to construct operators. However, in some cases the reverse is also true! For example if H is finite dimensional and M a self-adjoint matrix, then the operator $T\colon H \to H$, $Tx := Mx$ can be diagonalized to

$$T = \sum_{k=1}^{\dim H} \lambda_k P_{\lambda_k}$$

where λ_k is the k-th eigenvalue of M and P_{λ_k} the projection onto the normalized eigenvector associated to λ_k, if the eigenvalues are counted with multiplicity and the eigenvectors are orthogonalized. We may then define a spectral family E_T via

$$E_T(t) = \sum_{k:\lambda_k \leq t} P_{\lambda_k}$$

which satisfies

$$T = \int t \; \mathrm{d}E_T(t).$$

The extension of this result to self-adjoint operators is the following.

Theorem 3.19 (Spectral theorem for self-adjoint operators, [Wei80, Theorem 7.17])
Let $T\colon D(T) \subset H \to H$ be a self-adjoint operator on the Hilbert space H. Then there exists exactly one spectral family E_T such that $T = \hat{E_T}(id)$, also written as

$$T = \int t \, \mathrm{d}E_T(t).$$

An explanation for the name "spectral" family can be found in the following theorem that links the spectrum of T and the spectral family E_T.

Theorem 3.20 ([Wei80, Theorem 7.22])
Let T be a self-adjoint operator on the Hilbert space H with spectral family E_T. Then

$$\lambda \in \sigma(T) \Leftrightarrow E_T(\lambda + \varepsilon) - E_T(\lambda - \varepsilon) \neq 0 \ \forall \varepsilon > 0$$

One use of spectral families for self-adjoint operators and the integrals with respect to spectral families lies in their relation to one-parameter unitary groups.

Theorem 3.21 ([Wei80, Theorem 7.37])
Let $T\colon D(T) \subset H \to H$ be a self adjoint operator on a complex Hilbert space H with spectral family E_T, $t \in \mathbb{R}$ and let

$$U(t) = \mathrm{e}^{\mathrm{i}tT} := \int \mathrm{e}^{\mathrm{i}tr} \mathrm{d}E_T(r).$$

Then $(U(t))_{t\in\mathbb{R}}$ is a ***one-parameter group****. This means that $U(t)$ is a bounded operator on H for all $t \in \mathbb{R}$ with $U(0) = I$ (where I is the identity operator) and for all $s, t \in \mathbb{R}$ the formula $U(s)U(t) = U(s+t)$ holds. The group is also* ***strongly continuous****, meaning that $U(\cdot)x : \mathbb{R} \to H$ is continuous for all $x \in H$. For every $t \in \mathbb{R}$ the operator $U(t)$ is furthermore unitary and $U(t)x \in D(T)$ for all $x \in D(T)$.*

Conversely, every strongly continuous one-parameter unitary group has a representation as an integral with regard to a spectral family.

Theorem 3.22 (Stone's Theorem, [Wei80, Theorem 7.38])
Let $(U(t))_{t\in\mathbb{R}}$ be a strongly continuous one-parameter unitary group on a complex Hilbert space H. Then there exists a uniquely determined self-adjoint operator T on H such that

$$U(t) = \mathrm{e}^{\mathrm{i}tT} \ \forall t \in \mathbb{R}$$

In this case $\mathrm{i}T$ is called the ***infinitesimal generator*** *of $(U(t))_{t\in\mathbb{R}}$,*

$$D(\mathrm{i}T) = \left\{ x \in H \mid \lim_{t\to 0} \frac{1}{t} \left(U(t) - I\right) x \ \textit{exists} \right\},$$
$$(\mathrm{i}T)x = \lim_{t\to 0} \frac{1}{t} \left(U(t) - I\right) x \quad \forall x \in D(\mathrm{i}T).$$

As already noted in the introduction, the time evolution of a wave function needs to be a unitary one-parameter group. So if the group is also strongly continuous, then there is a self-adjoint operator that generates the group. We can even show that this generator would be involved in an initial value problem.

Corollary 3.23 ([Wei80, Corollary after Theorem 7.38])
If $(U(t))_{t\in\mathbb{R}}$ is a strongly continuous one-parameter unitary group with infinitesimal generator $\mathrm{i}T$, then the initial value problem

$$\frac{1}{\mathrm{i}}\frac{\mathrm{d}}{\mathrm{d}t}u(t) = Tu(t), \quad u(0) = x$$

is uniquely solvable for every $x \in D(T)$ and the solution is $u(t) = U(t)x$.

Since the time-dependent Schrödinger equation is exactly of this type, we see that the time evolution is explicitly linked to the Hamilton operator and its spectral family, and via Theorem 3.20 it is thus useful to study the spectrum to find information about the solutions of the Schrödinger equation.

3.5 The integrated density of states

As already shown, the spectrum of the Anderson operator is almost surely a deterministic set and every self-adjoint operator has an associated spectral family. Thus, the next question is whether the non-randomness of the spectrum also extends to the spectral families. Theorem 3.20 already gives us a hint that this is at least partially true.
The non-randomness of the spectrum depends mainly on the fact that the Anderson operator involves an infinite amount of identically distributed random variables. On a finite sub-graph the randomness can not be averaged out completely, but eigenvalues and -vectors of an operator on a finite graph are easier to calculate than the spectrum of an operator on an infinite graph. It is therefore interesting to investigate if there is some convergence if the Anderson operator is restricted to larger and larger finite subgraphs, and if this convergence approximates results for the operator on the whole graph. The restriction we want to use here is the following.

Definition 3.24 (The restricted Anderson operator)
For a finite subset $\Lambda \subset \mathbb{Z}^d$ the ***restricted Anderson operator*** *is*

$$H_\omega^\Lambda : \ell^2(\Lambda) \to \ell^2(\Lambda),\ H_\omega^\Lambda = p_\Lambda H_\omega i_\Lambda.$$

Here, H_ω is an Anderson operator $H_\omega = -\Delta + V_\omega$ consisting of the negative discrete Laplace operator $-\Delta$ and a random multiplication operator V_ω on $\ell^2(\mathbb{Z}^d)$,
$i_\Lambda : \ell^2(\Lambda) \to \ell^2(\mathbb{Z}^d)$ with

$$i_\Lambda \phi(z) = \begin{cases} \phi(z) & z \in \Lambda \\ 0 & z \notin \Lambda \end{cases}$$

is the ***embedding*** *of $\ell^2(\Lambda)$ into $\ell^2(\mathbb{Z}^d)$ and $p_\Lambda : \ell^2(\mathbb{Z}^d) \to \ell^2(\Lambda)$ with*

$$p_\Lambda \varphi(z) = \varphi(z)\ \forall z \in \Lambda$$

is the ***projection*** *from $\ell^2(\mathbb{Z}^d)$ to $\ell^2(\Lambda)$.*

Since $\ell^2(\Lambda)$ has dimension $|\Lambda|$, which is the number of points in Λ, the restricted Anderson operator is now an operator on a finite dimensional space and can be represented as a matrix with regard to some orthonormal basis. Here, we choose the basis

$$\left\{\delta_z \mid z \in \mathbb{Z}^d\right\}$$

with the **delta-functions**

$$\delta_z(y) = \begin{cases} 1 & y = z \\ 0 & y \neq z \end{cases}$$

for $y \in \mathbb{Z}^d$. Since H_ω is symmetric, $\delta_z \in D(H_\omega)$ for all $z \in \mathbb{Z}^d$ and $\langle \delta_y, H_\omega \delta_z \rangle = \langle \delta_y, H_\omega^\Lambda \delta_z \rangle$ as long as both z and y are in Λ, the corresponding matrix-representation of H_ω^Λ will be Hermitian. As mentioned at the beginning of Section 3 this means that H_ω^Λ has only real eigenvalues and the eigenvectors form an orthogonal basis. The number of eigenvalues is $|\Lambda|$ if we take into account multiplicities, so it is now possible to calculate the fraction of all eigenvalues that lie under a given real number x. In a sense this measures how much of $\ell^2(\Lambda)$ is covered by eigenvectors corresponding to the eigenvalues up to this number. The function to do this is the following.

Definition 3.25 (The normalized eigenvalue-counting function)
The ***eigenvalue counting function (evcf)*** *for a restricted Anderson operator H_ω^Λ is*

$$\tilde{N}_\omega^\Lambda \colon \mathbb{R} \to [0, |\Lambda|],\ \tilde{N}_\omega^\Lambda(x) := \#\left\{\text{eigenvalues of } H_\omega^\Lambda \leq x\right\}.$$

The normalized eigenvalue counting function (nevcf) for a restricted Anderson operator H_ω^Λ is

$$N_\omega^\Lambda \colon \mathbb{R} \to [0,1],\ N_\omega^\Lambda(x) := \frac{\#\left\{\text{eigenvalues of } H_\omega^\Lambda \leq x\right\}}{|\Lambda|} = \frac{\tilde{N}_\omega^\Lambda(x)}{|\Lambda|}.$$

All eigenvalues are always counted with their multiplicity.

3.5.1 Point-wise convergence

With these definitions we can give one concrete form to the question posed in the beginning of the section: Does the nevcf converge to something if Λ is chosen larger and larger. The existence of such a limit was shown in papers of Pastur in 1973 [Pas73] and Shubin in 1979 [Shu79], mostly formulated for Schrödinger operators on $\mathbb{R}^d$ and operators with almost periodic coefficients, which behave somewhat similar to random operators. On $\mathbb{Z}^d$ it takes the form of the following result:

Theorem 3.26 (Pastur-Shubin formula, [Pas80, Theorem 12])
Let H_ω be the Anderson operator as defined in (2.6) *where the random potentials are defined via $v_\omega(x) = \omega_x$ with bounded ω_x and under assumption of ergodicity and some regularity assumptions, and for any finite set $\Lambda \subset \mathbb{Z}^d$ let H_ω^Λ be the restricted Anderson operator and N_ω^Λ the normalized eigenvalue counting function as defined in Definitions 3.24 and 3.25. Let further $Q_n = [-n, n]^d \cap \mathbb{Z}^d$ and $N \colon \mathbb{R} \to [0, 1]$ be the non-random function given by*

$$N(x) := \mathbb{E}\langle \delta_0, E_\omega(x)\delta_0 \rangle,$$

where E_ω is the (measurable) spectral family of H_ω as defined in Theorem 3.19. Then for almost all $\omega \in \Omega$ and all x where N is continuous we have

$$N_\omega^{Q_n}(x) \xrightarrow[n \to \infty]{} N(x). \tag{3.10}$$

*The function N is called the **integrated density of states (IDS)**.*

Remark 3.27.
1. For this theorem it is important that the spectral family E_ω of H_ω is (weakly) measurable, which is not shown in detail in [Pas80]. A paper by Kirsch and Martinelli [MK82] covers this and other problems of measurability and showed that the IDS is well defined for our use-cases.

2. An alternative to the nevcf is the function

   $$N'^{\Lambda_n}_\omega(x) = \frac{\mathrm{tr}\,(E_\omega(x)\mathbb{1}_{\Lambda_n})}{|\Lambda_n|}.$$

   Here, the operator itself is not restricted, instead a restriction of the trace is used to insert a finite sub-graph. This function generally also converges almost surely to the IDS with growing n, see for example [Kir07, Theorem 5.22] for the case of i.i.d. potentials.

3. The properties of the spectral family in Definition 3.16 and the continuity of the scalar product imply that $x \mapsto \langle \delta_0, E_\omega(x)\delta_0 \rangle$ is monotone increasing and right-continuous for every $\omega \in \Omega$, and satisfies $\lim_{x\to\infty} \langle \delta_0, E_\omega(x)\delta_0 \rangle = 0$ and $\lim_{x\to\infty} \langle \delta_0, E_\omega(x)\delta_0 \rangle = 1$. By the dominated convergence theorem these properties also carry over to N.

As already mentioned above, some form of the Pastur-Shubin formula is valid in more general settings. One of these is the following result, which relaxes the assumption that the potentials $(v.(x))_{x\in\mathbb{Z}^d}$ have to be bounded. Here we use the notation of Section 2.4 and especially Remark 2.5.

Theorem 3.28
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with $\Omega = \mathbb{R}^{\mathbb{Z}^d}$, $\mathcal{A}$ the Borel σ-algebra generated by cylinder sets and $\mathbb{P}$ a probability measure that satisfies

- *γ_z is measurable and $\mathbb{P} = \mathbb{P} \circ \gamma_z$ for all $z \in \mathbb{Z}^d$, where*

$$\gamma_z : \Omega \to \Omega, (\gamma_z\omega)_y := \omega_{y+z}. \tag{3.11}$$

- *There is an $r \geq 0$ such that for all $n \in \mathbb{N}$ and finite sets $\Lambda_1, ..., \Lambda_n$ with $|\Lambda_i| \neq 0$ for all $1 \leq i \leq n$ and $\min\{\mathrm{d}(\Lambda_i, \Lambda_j) \mid i \neq j\} > r$ we have that the projections $(\Pi_{\Lambda_i})_{1\leq i\leq n}$ are independent.*
- *$\mathbb{E}\,|\omega_0|^2 < \infty$.*

Let H_ω be the Anderson operator as defined in (2.6) where the random potentials are defined via $v_\omega(x) = \omega_x$, and let H_ω^Λ, N_ω^Λ and N be defined as in Theorem 3.26. Then for almost all $\omega \in \Omega$ and all x where N is continuous the limit

$$N_\omega^{Q_n}(x) \xrightarrow[n\to\infty]{} N(x) \tag{3.12}$$

holds, where $Q_n = [-n, n]^d \cap \mathbb{Z}^d$.

Sketch of the proof. The proof follows from a combination of Theorems found in [PF91]. For a function ϕ in the space ℓ_0^2 of functions on $\mathbb{Z}^d$ that are only non-zero on finitely many points (see (3.7)) we can write an Anderson operator as given in the theorem as a matrix operator

$$H_\omega\phi(z) = \sum_{y\in\mathbb{Z}^d} a(z, y, \omega)\phi(y)$$

with the matrix

$$a(z, y, \omega) = \langle \delta_z, H_\omega \delta_y \rangle = \begin{cases} 2d + \omega_z & z = y \\ -1 & \|z - y\|_1 = 1 \\ 0 & \text{else} \end{cases}.$$

By [PF91, Example 1.4c] this is a symmetric random operator (as defined there) since ω_z is real-valued. This matrix operator is metrically transitive (also called ergodic) as defined in [PF91, Section 1.D] with the shifts $T_z = \gamma_z$ and $U_{\gamma_z}\phi(y) = \phi(y+z)$. To show this we first note that the equation

$$H_{T_z\omega} = U_{T_z} H_\omega U_{T_z}^{-1}$$

holds as required because of

$$a(z, y, \gamma_k\omega) = a(z+k, y+k, \omega)$$

for all $k, y, z \in \mathbb{Z}^d$ and all $\omega \in \Omega$. The shifts γ_z are automorphisms of Ω, so next we need to prove the ergodicity (or metric transitivity) of the group $\{\gamma_z \mid z \in \mathbb{Z}^d\}$. To do that we will show that the group is mixing, and we first observe that if

$$A = \{\omega \in \Omega \mid \omega_i \in A_i \ \forall i \in \{k_1, ..., k_n\}\}$$

is a rectangular cylinder set then

$$A = \Pi^{-1}_{\{k_1,...,k_n\}} \left(\bigtimes_{j=1}^{n} A_{k_j} \right)$$

and

$$\gamma_z A = \{\omega \in \Omega \mid \omega_i \in A_i \ \forall i \in \{k_1 - z, ..., k_n - z\}\} \tag{3.13}$$

is a rectangular cylinder set as well. If

$$B = \{\omega \in \Omega \mid \omega_i \in B_i \ \forall i \in \{l_1, ..., l_m\}\}$$

is also a cylinder set the second condition on the probability measure in the statement of the theorem requires that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

if $\min\{\|i-j\|_1 \mid i \in \{k_1, ..., k_n\}, j \in \{l_1, ..., l_m\} > d$. We can deduce from (3.13) that for all rectangular cylinder sets A, B we have

$$\mathbb{P}\left((\tau_z A) \cap B\right) = \mathbb{P}(A)\mathbb{P}(B)$$

for large enough z. Thus,

$$\lim_{n\to\infty} \mathbb{P}(((\tau_z)^n A) \cap B) = \lim_{n\to\infty} \mathbb{P}((\tau_{nz} A) \cap B) = \mathbb{P}(A)\mathbb{P}(B) \ \forall z \neq 0 \tag{3.14}$$

for all rectangular cylinder sets A, B. We need to extend this to all $A, B \in \mathcal{A}$. To this end we use a standard strategy via λ-systems. First we note that the rectangular cylinder sets generate $\mathcal{A}$, meaning that $\mathcal{A}$ is the smallest σ-algebra that contains every rectangular cylinder set. Then let

$$\mathcal{D} = \left\{ B \in \mathcal{A} \mid \lim_{n\to\infty} \mathbb{P}(((\tau_z)^n A) \cap B) = \mathbb{P}(A)\mathbb{P}(B) \ \forall \ z \neq 0, \text{ rect. cylinder sets } A \right\}.$$

If $B_1, ..., B_k \in \mathcal{D}$ are disjoint then $\cup_{i=1}^k B_i \in \mathcal{D}$, which follows from $\mathbb{P}(C \cap (\cup_{i=1}^k B_i)) = \sum_{i=1}^k \mathbb{P}(C \cap B_i)$ for all sets C. This can even be extended to confirm that countable disjoint unions of sets in $\mathcal{D}$ are themself elements of $\mathcal{D}$. If $B_1 \subset B_2$ and both sets are elements of $\mathcal{D}$ then so is $B_2 \setminus B_1$ by the same argument. From this and the fact that $B = \Omega \in \mathcal{D}$ follows that the class $\mathcal{D}$ is a λ-system. If $\mathcal{E}$ is a class of sets that is closed under intersections (such as the class of rectangular cylinder sets) and contained in a λ-system $\mathcal{D}$, then the σ-algebra generated by $\mathcal{E}$ (in this case $\mathcal{A}$) is also contained in $\mathcal{D}$. For a proof see [Kle20, Theorem 1.19] or [Kal21, Theorem 1.1]. Thus, all $B \in \mathcal{A}$ satisfy (3.14) for all rectangular cylinder sets A. Next we can use the invariance of $\mathbb{P}$ under shifts for

$$\mathbb{P}\left(((\tau_z)^n A) \cap B\right) = \mathbb{P}\left((\tau_{-z})^n \left(((\tau_z)^n A) \cap B\right)\right) = \mathbb{P}\left(A \cap \left((\tau_{-z})^n B\right)\right)$$

and reuse the arguments from above to show that (3.14) holds for all $A, B \in \mathcal{A}$. This means that all γ_z, $z \neq 0$ are mixing and therefore ergodic, making the group $\{\gamma_z \mid z \in \mathbb{Z}^d\}$ ergodic as required and H_ω a metrically transitive matrix operator.
Because of the third requirement on $\mathbb{P}$ and the Cauchy-Schwarz inequality we also have

$$\mathbb{E}a(0,0,\omega) < \infty, \ \mathbb{E}\left(a(0,0,\omega)\right)^2 < \infty,$$

and thus

$$\mathbb{E}\left(\sum_{z\in\mathbb{Z}^d} a(z,0,\omega)\right)^2 < \infty.$$

Now we can use [PF91, Theorem 4.8] to conclude that the measure arising from $N_\omega^{Q_n}$ converges weakly to the measure arising from $\mathbb{E}\langle\delta_0, E_\omega(x)\delta_0\rangle$ which is equivalent to the claimed statement (see Definition 13.21 and Theorem 13.23 in [Kle20]).

In some cases including the Anderson model, there is another feature of the IDS: As shown in [CS83] the IDS is log Hölder continuous in some cases. A very short proof of continuity can be found in [DS84]. From this follows that the convergence of the evcfs is point-wise for all $x \in \mathbb{R}$.

3.5.2 Uniform convergence

In view of these results the question arises whether the convergence could even be extended to be uniform in the “energy variable” $x \in \mathbb{R}$. A longer discussion of the history of this line of research can be found in Section 1 of [LSV10], but we will give a short overview here.
This direction of inquiry was developed by Lenz [Len02] and Stollmann [LS05] who proved uniform convergence for operators on Delone sets by proving a Banach-space

valued ergodic theorem that also applied for more general functions.
A parallel development was the establishment of uniform convergence specifically for the IDS in a series of papers by Elek ([Ele06a], [Ele06b]) and Lenz and Veselić [LV09] which could be used to prove a uniform Pastur-Shubin formula in a wide variety of cases.
These results generally did not give explicit quantifications for the uniform and almost sure convergence. That was achieved (again by finding a Banach-space valued ergodic theorem) among others in the case of the Anderson operator on $\mathbb{Z}^d$ in [LMV08] if the image measure of the potential has finite support, that is if the random variables at each lattice point can only take on values from a fixed finite set. Since every point of the graph can in this case be thought of as "colored" by one of these finite values, one realization of the random variables is then called a "coloring". This result was later generalized by Lenz, Schwarzenberger and Veselić [LSV10] to allow not only the lattice $\mathbb{Z}^d$ but also some graphs that are Cayley graphs of finitely generated amenable groups. It was later again extended to all Cayley graphs of finitely generated amenable groups by Pogorzelski and Schwarzenberger [PS16] through the use of ε-quasitilings. See [SSV20] for an overview of these results for potentials with finite support. However, since all these results relied on counting the frequencies at which finite "patterns" appeared in the coloring, this approach did not work if the random potentials could take on an uncountable number of colors.
This was partially achieved later by Schumacher, Schwarzenberger and Veselić in [SSV17] for operators on $\mathbb{Z}^d$ and later extended to all Cayley graphs of finitely generated amenable groups in [SSV18], but here only a quantification of the speed of the almost-sure convergence was achieved, without explicit constants. The probabilistic part of those proofs relied on a multi-dimensional Glivenko-Cantelli theorem due to DeHardt and Wright (Theorem 1 and 2 of [Wri81]). The relevant result for the Anderson operators with potentials satisfying the conditions of Theorem 3.28 from these papers is the following

Theorem 3.29 ([SSV17, Theorem 7.2])
Let $N_\omega^{\Lambda_n}$ be the normalized eigenvalue-counting function as in Definition 3.25 for an Anderson operator as in (2.6) *with random potentials satisfying the requirements of Theorem 3.28. Let N be the integrated density of states from Theorem 3.26 and $\Lambda_n = [0,n)^d \cap \mathbb{Z}^d$. For all $\kappa > 0$ and natural numbers $n > 2m > 0$ there are constants $a(\kappa, m)$ and $b(\kappa, m)$ such that there is an event $\Omega(\kappa, m, n)$ such that*

$$\mathbb{P}\left(\Omega(\kappa, n)\right) \geq 1 - b(\kappa, m) \exp\left(-a(\kappa, m) \lfloor n/m \rfloor^d\right)$$

and

$$\left\| N_\omega^{\Lambda_n} - N \right\|_\infty \leq 2^{2d+1} \left(\frac{26m^d + 8}{n - 2m} + \frac{24}{m} \right) + \frac{\kappa}{m^d}$$

for all $\omega \in \Omega(\varepsilon, n)$.

The next step is thus to find explicit bounds on the numbers $a(\varepsilon, m)$ and $b(\varepsilon, m)$, which will be the goal of the following chapters. First we will explain the approximation procedure used in the papers by Schumacher, Schwarzenberger and Veselić in more detail.

4 Framework

This chapter will introduce the most important requirements and definitions for probability spaces and functions used in the following parts of the thesis, as well as give a broad outline of the proof of explicit quantification of the almost-sure uniform convergence that we seek in Section 4.2. The primary object we will define are admissible functions, which formalize all needed properties of the eigenvalue-counting functions of the Anderson model. Both this definition and the strategy of the proof are based on [SSV17].

4.1 Geometric and probabilistic framework in $\mathbb{Z}^d$

First we want to fix some notation following [SSV17] and [SSV18].
The graph we are interested in for now is $(\mathbb{Z}^d, \mathcal{E})$, where $\mathbb{Z}^d$ is the set of vertices and $\mathcal{E} \subset \mathbb{Z}^d \times \mathbb{Z}^d$ is the set of edges between vertices. Some examples of other vertex or edge sets are considered in Chapter 8. Specifically we equip the graph with edges between closest neighbouring points, so for $\{x, y\} \in \mathbb{Z}^d \times \mathbb{Z}^d$ we have

$$\{x, y\} \in \mathcal{E} \Leftrightarrow \|x - y\|_1 = 1.$$

Here, $\|x\|_1 = \sum_{i=1}^{d} |x_i|$ denotes the ℓ^1-norm in $\mathbb{Z}^d$.
We also use the ℓ^1-norm to define a metric on $\mathbb{Z}^d$ via

$$\mathrm{d}_{\mathbb{Z}^d} : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{N}_0,\ \mathrm{d}_{\mathbb{Z}^d}(x, y) = \|x - y\|_1 .$$

For $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^d$ we define the distance between Λ_1 and Λ_2 as

$$\mathrm{d}_{\mathsf{set}}(\Lambda_1, \Lambda_2) := \min \{\mathrm{d}_{\mathbb{Z}^d}(x, y) \mid x \in \Lambda_1, y \in \Lambda_2\} .$$

If $\Lambda_1 = \{x\}$ for some $x \in \mathbb{Z}^d$ we write $\mathrm{d}_{\mathsf{set}}(x, \Lambda_2)$ instead of $\mathrm{d}_{\mathsf{set}}(\{x\}, \Lambda_2)$.
We write $\mathcal{C}$ for the set of all finite subsets of $\mathbb{Z}^d$ and for $\Lambda \in \mathcal{C}$ we write $|\Lambda|$ for the number of elements in Λ.
We also extend the addition on $\mathbb{Z}^d$ to translations of sets by writing

$$\Lambda + z := \{x + z \mid x \in \Lambda\}$$

for any set $\Lambda \subseteq \mathbb{Z}^d$ and $z \in \mathbb{Z}^d$.

We call any translation of $\{x \in \mathbb{Z}^d : \ 0 \leq x_i < n \ \forall 1 \leq i \leq d\}$ a **cube of side length** n, which always contains n^d elements.
We further extend these translations to sums of sets via

$$\Lambda_1 + \Lambda_2 := \bigcup_{z \in \Lambda_2} (\Lambda_1 + z) = \{x + z \mid x \in \Lambda_1, z \in \Lambda_2\}$$

for $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^d$.
We call

$$\partial^r(\Lambda) := \{x \in \Lambda : \mathrm{d_{set}}(x, \mathbb{Z}^d \setminus \Lambda) \leq r\} \cup \{x \in \mathbb{Z}^d \setminus \Lambda : \mathrm{d_{set}}(x, \Lambda) \leq r\}$$

the r**-boundary** of a set $\Lambda \subset \mathbb{Z}^d$ and define

$$\Lambda^r := \Lambda \setminus \partial^r(\Lambda) = \{x \in \Lambda \mid \mathrm{d_{set}}(x, \mathbb{Z}^d \setminus \Lambda) > r\}.$$

For indexed sets like Λ_i we use the shorthand $\Lambda_i^r = (\Lambda_i)^r$.
We also assign each point in $\mathbb{Z}^d$ a value, which is often called the **color** of the point or the **potential** at the point. For this we choose a set $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ to be the set of possible values and define for each point $z \in \mathbb{Z}^d$ a **color/potential** ω_z that takes values in $\mathcal{A}$. Together, these form our sample set

$$\Omega := \mathcal{A}^{\mathbb{Z}^d} = \{\omega = (\omega_z)_{z \in \mathbb{Z}^d} \mid \omega_z \in \mathcal{A} \ \forall z \in \mathbb{Z}^d\} \subseteq \mathbb{R}^{\mathbb{Z}^d}.$$

On this sample set we define for each $z \in \mathbb{Z}^d$ the **translation**

$$\gamma_z : \Omega \to \Omega, (\gamma_z \omega)_y := \omega_{y+z}. \tag{4.1}$$

We also define the restriction of Ω to $\Lambda \in \mathcal{C}$ by

$$\Omega_\Lambda := \mathcal{A}^\Lambda = \{\omega = (\omega_z)_{z \in \Lambda} \mid \omega_z \in \mathcal{A} \ \forall z \in \Lambda\}$$

and define the projection

$$\Pi_\Lambda : \Omega \to \Omega_\Lambda, \ \Pi_\Lambda(\omega) := \omega_\Lambda := (\omega_z)_{z \in \Lambda}.$$

Let $\mathcal{B}(\mathcal{A})$ be the Borel σ-algebra of $\mathcal{A}$ that arises as a sub-σ-algebra of the usual Borel σ-algebra of $\mathbb{R}$. Then let $\mathcal{B}(\Omega) = \mathcal{B}(\mathcal{A})^{\mathbb{Z}^d}$ be the product σ-algebra on Ω generated by cylinder sets, so that $(\Omega, \mathcal{B}(\Omega))$ forms a measurable space. Furthermore let $\mathcal{B}(\Omega_\Lambda)$ be the product σ-algebra on Ω_Λ generated by cylinder sets. See Section 2.4 for the relevant definitions.
Lastly we need some assumptions on the probability measures we consider, which were already part of the requirements in Theorem 3.28.

Definition 4.1
We assume that there is a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{B}(\Omega))$ that fulfils the following assumptions:

(M1) ***Translation invariance****: γ_z is measurable and $\mathbb{P} \circ \gamma_z = \mathbb{P}$ for each $z \in \mathbb{Z}^d$.*

(M2) ***Independence at a distance****: There is an $r \geq 0$ such that for all $n \in \mathbb{N}$ if*

- $\Lambda_1, ..., \Lambda_n \in \mathcal{C}$,
- $|\Lambda_i| \neq 0$ *for all* $1 \leq i \leq n$ *and*
- $\min\{\mathrm{d}_{set}(\Lambda_i, \Lambda_j) \mid i \neq j\} > r$

then the projections $(\Pi_{\Lambda_i})_{1 \leq i \leq n}$ are independent.

Remark 4.2. Here, the r in (M2) can be thought of as a correlation length. If $r = 0$, then the values at each vertex are chosen independently. This would also mean that $\mathbb{P}$ is a product measure, which fulfils (M1).
As a consequence of (M2) if $\Lambda \in \mathcal{C}$ and $z \in \mathbb{Z}^d$ are such that $\mathrm{d}_{\mathrm{set}}(\Lambda, \Lambda + z) > r$, then Π_Λ and $\Pi_{\Lambda+z}$ are independent. This is also true for $(\Pi_{\Lambda+z_i})_{1 \leq i \leq n}$ if $\mathrm{d}_{\mathbb{Z}^d}(\Lambda + z_i, \Lambda + z_j) > r$ for all $i \neq j$. Note that since

$$\Pi_{\Lambda+z}(\omega) = (\omega_y)_{y \in \Lambda+z}$$

and

$$\Pi_\Lambda(\gamma_z \omega) = (\omega_{y+z})_{y \in \Lambda}$$

both $\Pi_{\Lambda+z}(\omega)$ and $\Pi_\Lambda(\gamma_z\omega)$ contain the same values, just differently indexed. Therefore, $(\Pi_\Lambda \circ \gamma_{z_i})_{1 \leq i \leq n}$ are also independent in the above case.

Now that we have defined the graph and the coloring/potentials on it, we can turn to the functions on it we want to consider. Our specific goal are the eigenvalue-counting functions, but we will generalize from there to all functions that possess the properties of the eigenvalue-counting function that we will need.
These functions will take values in the Banach space

$$\mathbb{B} := \{F : \mathbb{R} \to \mathbb{R} : F \text{ right-continuous and bounded}\}$$

equipped with the supremum norm $\|\cdot\|_\infty$.
The class of $\mathbb{B}$-valued functions we want to consider will be called *admissible* functions.

Definition 4.3
*A function $a : \mathcal{C} \times \Omega \to \mathbb{B}$ is called **admissible** if the following conditions are satisfied*

(A1) ***translation invariance**: For $\Lambda \in \mathcal{C},\ z \in \mathbb{Z}^d$ and $\omega \in \Omega$ we have*

$$a(\Lambda + z, \omega) = a(\Lambda, \gamma_z \omega).$$

(A2) ***locality**: For all $\Lambda \in \mathcal{C}$ and $\omega, \omega' \in \Omega$ with $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ we have*

$$a(\Lambda, \omega) = a(\Lambda, \omega').$$

(A3) ***almost additivity**: There exists a function $b \colon \mathcal{C} \to [0, \infty)$ such that for each $\omega \in \Omega$, pairwise disjoint $\Lambda_1, ..., \Lambda_n \in \mathcal{C}$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i$ we have*

$$\left\| a(\Lambda, \omega) - \sum_{i=1}^n a(\Lambda_i, \omega) \right\|_\infty \leq \sum_{i=1}^n b(\Lambda_i),$$

and b satisfies

- *for all $\Lambda \in \mathcal{C}$ and $z \in \mathbb{Z}^d$ we have $b(\Lambda) = b(\Lambda + z)$*
- *there is a $D > 0$ such that $b(\Lambda) \leq D\,|\Lambda|$ for all $\Lambda \in \mathcal{C}$*
- *if $(\Lambda_n)_{n \in \mathbb{N}}$ is a sequence of cubes with strictly increasing side length, then*

$$\lim_{n \to \infty} \frac{b(\Lambda_n)}{|\Lambda_n|} = 0$$

(A4) ***boundedness**: There is an $E < \infty$ such that*

$$\sup_{\omega \in \Omega} \|a(\{0\}, \omega)\|_\infty \leq E$$

(A5) ***monotonicity**: The function $a(\Lambda, \omega)$ is monotone increasing at all points, i.e.*

$$\forall \Lambda \in \mathcal{C}, \omega \in \Omega :\ \ x < y : a(\Lambda, \omega)(x) \leq a(\Lambda, \omega)(y).$$

(A6) ***point-wise measurability**: The function $a(\Lambda, \cdot)(x) : (\Omega, \mathcal{B}(\Omega)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable for all $x \in \mathbb{R}$ and $\Lambda \in \mathcal{C}$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of $\mathbb{R}$.*

Remark 4.4. Some remarks on this definition:

- Admissible functions were already defined in [SSV17], but with a slight difference: there (A5) and (A6) were not a requirement, instead there was an additional requirement that an admissible function should be separately monotone in each ω_x.

- Property (A2) ensures that the **restricted admissible function** $a_\Lambda(\cdot) : (\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda)) \to \mathbb{B}$ defined via

$$a_\Lambda(\omega_\Lambda) := a(\Lambda, \hat{i}_\Lambda(\omega_\Lambda)), \tag{4.2}$$

where $\hat{i}_\Lambda : (\Omega_\Lambda, \mathcal{B}(\Omega_\Lambda)) \to (\Omega, \mathcal{B}(\Omega))$ is

$$(\hat{i}_\Lambda(\omega_\Lambda))_z = \begin{cases} (\omega_\Lambda)_z & z \in \Lambda \\ c & z \notin \Lambda \end{cases},$$

the embedding of Ω_Λ into Ω with a fixed $c \in \mathcal{A}$, is well defined. The embedding is continuous with regard to the product topology and therefore measurable (see Section 2.4), and since $a(\Lambda, \cdot)(x)$ is measurable for all $x \in \mathbb{R}$ the same is true for $a_\Lambda(x)$. We also have $a_\Lambda(\Pi_\Lambda(\omega)) = a(\Lambda, \omega)$ and by (A1) also

$$a_{\Lambda+z}(\Pi_{\Lambda+z}(\omega)) = a(\Lambda + z, \omega) = a(\Lambda, \gamma_z(\omega)) = a_\Lambda(\Pi_\Lambda(\gamma_z(\omega)))$$

for all $z \in \mathbb{Z}^d$ and $\omega \in \Omega$.

- From property (A3) follows that

$$\begin{aligned} \|a(\Lambda, \omega)\|_\infty &\leq \left\| a(\Lambda, \omega) - \sum_{z\in\Lambda} a(\{z\}, \omega) \right\|_\infty + \left\| \sum_{z\in\Lambda} a(\{z\}, \omega) \right\|_\infty \\ &\leq \sum_{z\in\Lambda} b(\{z\}) + \sum_{z\in\Lambda} \|a(\{z\}, \omega)\|_\infty \\ &\leq D\,|\Lambda| + \sum_{z\in\Lambda} \|a(\{z\}, \omega)\|_\infty . \end{aligned}$$

By properties (A1) and (A4) we further get

$$\begin{aligned} \|a(\Lambda, \omega)\|_\infty &\leq D\,|\Lambda| + \sum_{z\in\Lambda} \|a(\{0\}, \gamma_z\omega)\|_\infty \\ &\leq D\,|\Lambda| + \sum_{z\in\Lambda} E = (D + E)\,|\Lambda| \end{aligned} \tag{4.3}$$

for all $\omega \in \Lambda$.

- The properties (A4) and (A6) ensure that $\mathbb{E}a(\Lambda, \omega)(x)$ exists for all $x \in \mathbb{R}$.

The functions we are most interested in are naturally the eigenvalue-counting functions of the Anderson operator from Definition 3.25. Because of this the parameter x in $a(\Lambda, \omega)(x)$ is often called "energy". Most of the properties required of an admissible function were already proved for the eigenvalue-counting functions of the Anderson operator in [SSV17] and [LSV10].

Lemma 4.5
The function

$$\tilde{N}: \mathcal{C} \times \Omega \to \mathbb{B},\ \tilde{N}(\Lambda, \omega) = \tilde{N}^\Lambda_\omega$$

with the eigenvalue-counting function $\tilde{N}^\Lambda_\omega$ *as defined in Definition 3.25 for an Anderson operator* $H_\omega = -\Delta + V_\omega : \ell^2(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ *with*

$$H_\omega \Psi(z) = -\Delta\Psi(z) + \omega_z(z)\Psi(z),$$

with the discrete Laplacian Δ *and* ω *from a probability space satisfying the properties of Definition 4.1 is an admissible function.*

Proof. Every evcf $\tilde{N}^\Lambda_\omega$ is bounded by $|\Lambda|$ and right-continuous (because of the $\leq$ in the definition), so the image of $\tilde{N}$ is indeed in $\mathbb{B}$. We will now show all needed properties in the order they appear in in Definition 4.3.

To (A1) We need to show that $\tilde{N}(\Lambda + z, \omega) = \tilde{N}(\Lambda, \gamma_z\omega)$ for all $\Lambda \in \mathcal{C}$ and all $z \in \mathbb{Z}^d$.
Let $z \in \mathbb{Z}^d$ and let H^Λ_ω be the restricted Anderson operator as in Definition 3.24. We will show that $H^{\Lambda+z}_\omega$ and $H^\Lambda_{\gamma_z\omega}$ have the same spectrum, thus their evcfs are also the same.
Let λ be an eigenvalue of $H^{\Lambda+z}_\omega$ with corresponding eigenvector $v_\lambda \in \ell^2(\Lambda + z)$, i.e. $H^{\Lambda+z}_\omega v_\lambda = \lambda v_\lambda$. Then for all $y \in \Lambda$ the equality

$$\lambda v_\lambda(y+z) = \left(H^{\Lambda+z}_\omega v_\lambda\right)(y+z) = -(\Delta v_\lambda)(y+z) + \omega_{y+z} v_\lambda(y+z)$$

follows from the definition of $H^{\Lambda+z}_\omega$.
Based on this eigenfunction we define another function $v'_\lambda \in \ell^2(\Lambda)$ via
$v'_\lambda(y) := v_\lambda(y+z)$ for $y \in \Lambda$. Now we apply $H^\Lambda_{\gamma_z\omega}$ to this function and get

$$\begin{aligned}
\left(H^\Lambda_{\gamma_z\omega} v'_\lambda\right)(y) &= -(\Delta v'_\lambda)(y) + (V_{\gamma_z\omega})v'_\lambda(y) \\
&= (-\Delta v_\lambda)(y+z) + (\gamma_z\omega)_y v_\lambda(y+z) \\
&= (-\Delta v_\lambda)(y+z) + \omega_{y+z} v_\lambda(y+z) \\
&= \lambda v_\lambda(y+z) = \lambda v'_\lambda(x)
\end{aligned}$$

for $y \in \Lambda$, where we used the translation invariance (3.9) of the Laplace operator at the second equality. Therefore, λ is also an eigenvalue of $H^\Lambda_{\gamma_z\omega}$. This is true for all $\Lambda \in \mathcal{C}$, $\omega \in \Omega$ and $z \in \mathbb{Z}^d$. By using $\Lambda' = \Lambda + z$, $\omega' = \gamma_z\omega$ and $z' = -z$ we also get the reverse implication from $H^{\Lambda'+z'}_{\omega'} = H^\Lambda_{\gamma_z\omega}$ and $H^{\Lambda'}_{\gamma_{z'}\omega'} = H^{\Lambda+z}_\omega$. This proves that $H^{\Lambda+z}_\omega$ and $H^\Lambda_{\gamma_z\omega}$ have the same eigenvalues. Using the above scheme for a whole set of independent eigenvectors in $\ell^2(\Lambda + z)$ leads to a new set of eigenvectors to the same eigenvalue, which are again independent (since the eigenvectors are only shifted by z). This means that the multiplicity of the eigenvalues is also the same for $H^{\Lambda+z}_\omega$ and $H^\Lambda_{\gamma_z\omega}$ and therefore
$\tilde{N}(\Lambda + z, \omega) = \tilde{N}(\Lambda, \gamma_z\omega)$ as required by (M1).

To (A2) $\Pi_\Lambda(\omega) = \Pi_\Lambda(\omega')$ implies $\omega_y = \omega'_y$ for all $y \in \Lambda$, so $p_\Lambda V_\omega i_\Lambda = p_\Lambda V_{\omega'} i_\Lambda$ and thus $H^\Lambda_\omega = H^\Lambda_{\omega'}$ and $\tilde{N}(\Lambda,\omega) = \tilde{N}(\Lambda,\omega')$.

To (A3) Here we use two results from [LSV10], the first is Proposition 7.1. As proved there, for two self-adjoint operators A and C on a finite dimensional Hilbert space the inequality

$$|n(A)(x) - n(A+C)(x)| \leq \operatorname{rank}(C)$$

holds for the eigenvalue-counting function n and all $x \in \mathbb{R}$. This follows from the minmax principle of Courant and Fischer.

This result is then used for [LSV10, Proposition 7.2]. Here it was shown that if V is a finite dimensional Hilbert space with a subspace U and the inclusion $i\colon U \to V$ and orthogonal projection $p : V \to U$, then the inequality

$$|n(A)(x) - n(pAi)(x)| \leq 4\operatorname{rank}(\mathrm{id}_V - i \circ p)$$

holds for all self-adjoint operators A and $x \in \mathbb{R}$. We use this result for $\Lambda' \subseteq \Lambda \in \mathcal{C}$ with $A = H^\Lambda_\omega$, $V = \ell^2(\Lambda)$ and $U = \ell^2(\Lambda')$. Then $pAi = H^{\Lambda'}_\omega$ and

$$(\mathrm{id}_V - i \circ p)\phi(x) = \begin{cases} \phi(x) & x \in \Lambda \setminus \Lambda' \\ 0 & x \notin \Lambda \setminus \Lambda' \end{cases},$$

leading to $\operatorname{rank}(\mathrm{id}_V - i \circ p) = |\Lambda \setminus \Lambda'|$ and the equation

$$\left\| \tilde{N}(\Lambda,\omega) - \tilde{N}(\Lambda',\omega) \right\|_\infty \leq 4 \left| \Lambda \setminus \Lambda' \right| \tag{4.4}$$

for all $\omega \in \Omega$. To prove almost additivity, let $\Lambda_i \in \mathcal{C}$, $i = 1, ..., n$ be disjoint sets and $\Lambda = \cup_{i=1}^n \Lambda_i$. Now we use (4.4) and the triangle inequality to show

$$\begin{aligned} &\left\| \tilde{N}(\Lambda,\omega) - \sum_{i=1}^n \tilde{N}(\Lambda_i,\omega) \right\|_\infty \qquad (4.5) \\ &\leq \left\| \tilde{N}(\Lambda,\omega) - \tilde{N}\left(\bigcup_{i=1}^n \Lambda_i^1, \omega \right) \right\|_\infty + \left\| \tilde{N}\left(\bigcup_{i=1}^n \Lambda_i^1, \omega \right) - \sum_{i=1}^n \tilde{N}(\Lambda_i,\omega) \right\|_\infty \\ &\leq 4 \sum_{i=1}^n \left| \Lambda_i \setminus \Lambda_i^1 \right| + \left\| \tilde{N}\left(\bigcup_{i=1}^n \Lambda_i^1, \omega \right) - \sum_{i=1}^n \tilde{N}(\Lambda_i^1,\omega) \right\|_\infty \\ &\quad + \sum_{i=1}^n \left\| \tilde{N}(\Lambda_i^1,\omega) - \tilde{N}(\Lambda_i,\omega) \right\|_\infty \\ &\leq 8 \sum_{i=1}^n \left| \Lambda_i \setminus \Lambda_i^1 \right| + \left\| \tilde{N}\left(\bigcup_{i=1}^n \Lambda_i^1, \omega \right) - \sum_{i=1}^n \tilde{N}(\Lambda_i^1,\omega) \right\|_\infty . \qquad (4.6) \end{aligned}$$

Since the Λ_i are disjoint, the distance between them is at least one. We also exclude the 1-boundary of all Λ_i, so we get

$$\mathrm{d_{set}}(\Lambda_i^1, \Lambda_j^1) \geq 3 \ \forall i \neq j.$$

This means that for $i \neq j$, each $\phi \in \ell^2(\Lambda_i^1 \cup \Lambda_j^1)$ has a unique representation as

$$\phi(z) = \begin{cases} \phi_1(z), & \text{if } \ z \in \Lambda_i^1 \\ \phi_2(z), & \text{if } z \in \Lambda_j^1 \end{cases}$$

with $\phi_1 \in \ell^2(\Lambda_i^1)$ and $\phi_2 \in \ell^2(\Lambda_j^1)$.
The operator H_ω has hopping range 1, i.e. $H_\omega \phi(z)$ only depends on $\phi(y)$ for y with $|z-y| \leq 1$. With the representation of $\phi \in \ell^2(\Lambda_i^1 \cup \Lambda_j^1)$ from above, we get

$$H_\omega^{\Lambda_i^1 \cup \Lambda_j^1} \phi(y) = \begin{cases} H_\omega^{\Lambda_i^1} \phi_1(y), & \text{if } y \in \Lambda_i^1 \\ H_\omega^{\Lambda_j^1} \phi_2(y), & \text{if } y \in \Lambda_j^1 \end{cases}$$

i.e. $H_\omega^{\Lambda_i^1 \cup \Lambda_j^1} = H_\omega^{\Lambda_i^1} \oplus H_\omega^{\Lambda_j^1}$.
The eigenvalues of $H_\omega^{\Lambda_i^1 \cup \Lambda_j^1}$ are thus the union of the eigenvalues of $H_\omega^{\Lambda_i^1}$ with those of $H_\omega^{\Lambda_j^1}$ and the multiplicity of an eigenvalue of $H_\omega^{\Lambda_i^1 \cup \Lambda_j^1}$ is exactly the sum of the multiplicity of that eigenvalue for $H_\omega^{\Lambda_i^1}$ and the multiplicity for $H_\omega^{\Lambda_j^1}$. This summing of eigenvalues then also carries over to the evcfs, so $\tilde{N}\left(\Lambda_i^1 \cup \Lambda_j^1, \omega\right) = \tilde{N}(\Lambda_i^1, \omega) + \tilde{N}(\Lambda_j^1, \omega)$ and by induction

$$\tilde{N}\left(\bigcup_{i=1}^n \Lambda_i^1, \omega\right) = \sum_{i=1}^n \tilde{N}(\Lambda_i^1, \omega).$$

The second term on the right side of (4.5) therefore vanishes and we arrive at

$$\left\| \tilde{N}(\Lambda, \omega) - \sum_{i=1}^n \tilde{N}(\Lambda_i, \omega) \right\|_\infty \leq 8 \sum_{i=1}^n \left| \Lambda_i \setminus \Lambda_i^1 \right|,$$

and $b : \mathcal{C} \to [0, \infty), b(\Lambda) = 8 \left| \Lambda \setminus \Lambda^1 \right|$ is a natural definition for a boundary term. We now only need to check if this b fulfills all the requirements from (A3).
First we see that for all $\Lambda \in \mathcal{C}$ and $z \in \mathbb{Z}^d$ we have $b(\Lambda + z) = b(\Lambda)$ and $b(\Lambda) \leq 8 |\Lambda|$, so the first two requirements on b from (A3) are fulfilled and $D = 8$. For the last point we need to confirm that if Λ_n is a series of cubes of growing side length, then $\frac{b(\Lambda_n)}{|\Lambda_n|} \xrightarrow{n \to \infty} 0$. We can already establish some useful formulas for later use here, so we will prove a bit more than necessary. From $\Lambda_n \setminus \Lambda_n^1 \subset \partial^1(\Lambda_n) \subset \partial^r(\Lambda_n)$ for any $r \geq 1$ we see that

$$\frac{b(\Lambda_n)}{|\Lambda_n|} \leq \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}.$$

We will now consider a cube Λ_m with side length exactly $m > 2r$ and take a look at $|\partial^r(\Lambda_m)|$. Since Λ_m is a cube, $\partial^r(\Lambda_m) \cup \Lambda_m$ is a cube with side length $m + 2r$ and $\Lambda_m \setminus \partial^r(\Lambda_m)$ is a cube with side length $m - 2r$. Since $\partial^r(\Lambda_m) = (\partial^r(\Lambda_m) \cup \Lambda_m) \setminus (\Lambda_m \setminus \partial^r(\Lambda_m))$ we have

$$|\partial^r(\Lambda_m)| = |\Lambda_{m+2r}| - |\Lambda_{m-2r}| = (m + 2r)^d - (m - 2r)^d \sim m^{d-1},$$

and thus

$$\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|} \xrightarrow{m\to\infty} 0 \tag{4.7}$$

for all $r \in \mathbb{N}$ since $|\Lambda_m| = m^d$. The same idea results in

$$\frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \xrightarrow{n\to\infty} 0$$

for all $m \in \mathbb{N}$ where $n > 2m$. As a consequence $\frac{b(\Lambda_n)}{|\Lambda_n|} \xrightarrow{n\to\infty} 0$ holds just as required and we have confirmed that the evcfs are almost additive.

To (A4) The operator $H_\omega^{\{0\}}$ only has one eigenvalue $(2d + \omega_{\{0\}})$, so

$$\left\| \tilde{N}(\{0\}, \omega) \right\|_\infty = 1 \ \forall \omega \in \Omega$$

which means that (A4) is fulfilled with $E = 1$.

To (A5) The evcf count the number of eigenvalues below a given number x, so the counted number of eigenvalues can only increase if x is increased (while keeping everything else fixed). Therefore, the evcf are monotone increasing for every fixed $\omega \in \Omega$ and satisfy (A5).

To (A6) Let $\lambda_j^\downarrow(H_\omega^\Lambda)$ be the j-th largest eigenvalue of H_ω^Λ, counted with multiplicities. We use Corollary III.2.6 of [Bha13], which shows that for all $\Lambda \in \mathcal{C}$ and $\omega \neq \omega' \in \Omega$ we have

$$\max_{j=1,\dots,|\Lambda|} \left| \lambda_j^\downarrow(H_\omega^\Lambda) - \lambda_j^\downarrow(H_{\omega'}^\Lambda) \right| \leq \left\| H_\omega^\Lambda - H_{\omega'}^\Lambda \right\|$$

where the last norm is the operator norm. From the definition of the restricted Anderson operator follows

$$\left\| H_\omega^\Lambda - H_{\omega'}^\Lambda \right\| = \| p_\Lambda (V_\omega - V_{\omega'}) i_\Lambda \| = \| p_\Lambda (V_{\omega - \omega'}) i_\Lambda \| = \left\| \omega_\Lambda - \omega'_\Lambda \right\|_\infty .$$

Thus, $\lambda_j^\downarrow(H_\cdot^\Lambda) : \Omega^\Lambda \to \mathbb{R}$ is continuous and therefore measurable for all j. This in turn means that $\{\omega \in \Omega : \lambda_j^\downarrow(H_\omega^\Lambda) \leq x\}$ is a measurable set for all $x \in \mathbb{R}$

and all $1 \le j \le |\Lambda|$, and consequently the functions $\mathbb{1}\left\{\{\lambda_j^{\downarrow}(H_\omega^\Lambda) \le x\}\right\}$ are also measurable. Finally

$$\tilde{N}(\Lambda,\omega)(x) = \sum_{j=1}^{|\Lambda|} \mathbb{1}\left\{\{\lambda_j^{\downarrow}(H_\omega^\Lambda) \le x\}\right\},$$

which shows that $\tilde{N}(\Lambda,\omega)(x)$ is measurable for all $x \in \mathbb{R}$ and $\Lambda \in \mathcal{C}$.

□

4.2 The strategy towards quantitative approximation

The general strategy we will use follows [SSV18], but variations of it were already used earlier, e.g. in [LMV08] and [LSV10].
We want to establish an almost-sure uniform convergence of $\frac{a(\Lambda,\omega)}{|\Lambda|}$ to a limit function a^* along larger and larger sets Λ, with quantification in every aspect of the convergence. Inspired by the Glivenko-Cantelli theorem, we want to approximate $\frac{a(\Lambda,\omega)}{|\Lambda|}$ with an averaged sum over a large number of independent, identically distributed random variables. Then we need a probabilistic argument to show that this averaged sum converges almost surely uniformly to an expected value and that also gives explicit error bounds in the form of a concentration inequality. After that we no longer have to deal with randomness, but we still need to show that the expected value we got from the last step converges to some limit. Symbolically, the strategy can be summarised as follows:

$$\frac{a(\Lambda_n,\omega)}{|\Lambda_n|} \approx \frac{1}{k(n)} \sum_{i=1}^{k(n)} \frac{a(\Lambda_{m,i},\omega)}{|\Lambda_{m,i}|} \xrightarrow[a.s.]{n\to\infty} \mathbb{E}\left(\frac{a(\Lambda_m,\omega)}{|\Lambda_m|}\right) \xrightarrow{m\to\infty} a^*. \tag{4.8}$$

Here, Λ_n is a cube with side length n, $k(n)$ is a number that grows monotonously with n, $(\Lambda_{m,i})_{1\le i\le k(n)}$ are a series of cubes with side length $m < 2n$ that do not intersect and a^* is some limit to be determined.
The first step of this approximation has to hold for all $\omega \in \Omega$ and primarily uses the almost-additivity as well as some of the other properties of admissible functions. The last step also does not depend on anything random, and it turns out that we can reuse the first step for this part as well. These steps do not depend on ω and we want uniform bounds here, so what we are looking for is something like explicit numbers $c_1(n,m)$ and $c_2(m)$ with $\lim\limits_{n\to\infty,m\to\infty} c_1(n,m) + c_2(m) = 0$ such that

$$\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{k(n)} \sum_{i=1}^{k(n)} \frac{a(\Lambda_{m,i},\omega)}{|\Lambda_{m,i}|} \right\|_\infty \le c_1(n,m)$$

and

$$\left\| \mathbb{E}\left(\frac{a(\Lambda_m, \omega)}{|\Lambda_m|} \right) - a^* \right\|_\infty \leq c_2(m).$$

Both of these steps will be carried out in Chapter 5. It is also possible to generalize the procedure somewhat to other shapes than just cubes, which will also be carried out in that chapter.

The second step on the other hand does not really rely on the considered geometry and is nearly entirely probabilistic. This is where we will get the new explicit quantification of the almost sure convergence from, and based on the results already achieved in [SSV17] and [SSV18] we anticipate that we should get explicit functions $c_3(\kappa, n, m)$ and $c_4(\kappa, n, m)$ with a concentration inequality of the sort

$$\left\| \frac{1}{k(n)} \sum_{i=1}^{k(n)} \frac{a(\Lambda_{m,i}, \omega)}{|\Lambda_{m,i}|} - \mathbb{E}\left(\frac{a(\Lambda_m, \omega)}{|\Lambda_m|} \right) \right\|_\infty < \kappa$$

for ω in a set $\Omega(\kappa, n, m)$ with

$$\mathbb{P}\left(\Omega(\kappa, n, m) \right) \geq 1 - c_3(\kappa, n, m) \exp\left(-c_4(\kappa, n, m) \right).$$

This step will be shown in Chapter 6 and is based on empirical process theory.

Finally we will combine all of the results in Chapter 7.

5 Geometric partition arguments

As outlined in Section 4.2, we first need to establish uniform bounds on the difference between an admissible function on a large set and the sum over that admissible function on subsets that form a disjoint partition of the large set. This bound further needs to be independent of any random influence. Such replacements of a larger system by a sum of smaller ones are a general feature of thermodynamic arguments, but we follow the specific procedure as outlined in [SSV17].

Furthermore we will see that the same bound also allows us to establish that the expected values of normalized admissible functions on growing sets form a Cauchy sequence, and since they are elements of the Banach space $\mathbb{B}$ there is a limit a^*.

Here, we will first demonstrate the scheme for the explicit example of cubes. This both makes the used definitions and geometric bounds very easy to visualize and also allows us to get very explicit bounds that would not be available for general sets. A more general case of sets with small boundary compared to their volume will be treated after that, but the techniques are nearly identical.

5.1 For cubes

We start with the cube $\Lambda_n := ([0, n) \cap \mathbb{Z})^d$ with side length $n \in \mathbb{N}$. We "tile" this cube by smaller cubes with side length m, such that $2m < n$ and define the **tiling set**

$$T_{m,n} := \{t \in T_m : \Lambda_m + t \subset \Lambda_n\}$$

where $T_m := m\mathbb{Z}^d = \{(z^{(i)})_{1 \leq i \leq d} \in \mathbb{Z}^d : z^{(i)} \bmod m = 0 \ \forall 1 \leq i \leq d\}$. The size of this tiling set is $|T_{m,n}| = \lfloor n/m \rfloor^d$ and all of the sets $\Lambda_m + t$ with $t \in T_m$ are pairwise disjoint. Together, they form

$$\Lambda_{m,n} := \bigcup_{t \in T_{m,n}} (\Lambda_m + t) = \Lambda_m + T_{m,n}.$$

Note that $\Lambda_{m,n} = \Lambda_{\lfloor n/m \rfloor m}$. We further define

$$\hat{\Lambda}_{m,n} := \Lambda_n \setminus \Lambda_{m,n}.$$

See Figure 5.1 for an illustration.

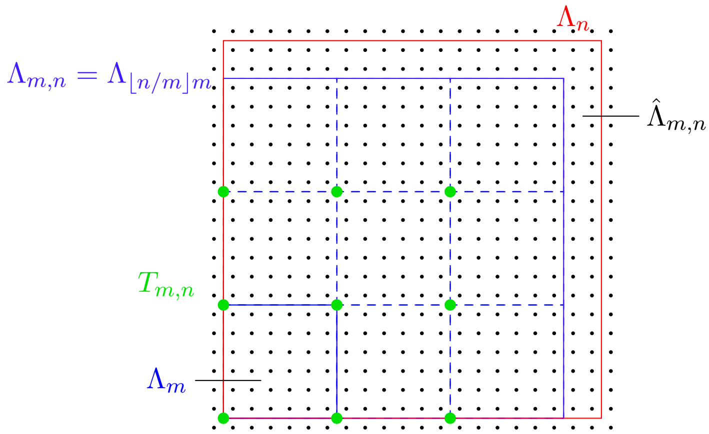

Figure 5.1: Illustration of the different definitions employed in $d = 2$.

The result we aim for is the first step of (4.8). We want a uniform bound on the difference between a normalized admissible function on Λ_n and the averaged sum over the same normalized admissible function on translates of Λ_m, while also cutting of r-boundaries of the tiles $\Lambda_m + t$, $t \in T_{m,n}$. Since we required some independence at a distance for our probability space, this will allow us to use the normalized admissible function on the translates as independent random variables. The proof mostly works by invoking the almost-additivity (A3) from Definition 4.3 and using convenient features of cubes and cube-tilings.

Lemma 5.1
Let $\Lambda_n := ([0, n) \cap \mathbb{Z})^d$. For any admissible function $a : \mathcal{C} \times \Omega \to \mathbb{B}$ we have for all $\omega \in \Omega$, and all $n, m, r \in \mathbb{N}$ with $n > 2m > 4r$

$$\begin{aligned} &\left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} \right\|_\infty \\ &\quad \leq \frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} + \frac{(3D + 2E)\left(|\Lambda_n| - |\Lambda_n^m|\right)}{|\Lambda_n^m|} \\ &\quad\quad + \frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D + E)\left(|\Lambda_m| - |\Lambda_m^r|\right)}{|\Lambda_m|} \end{aligned} \tag{5.1}$$

This results in

$$\lim_{m\to\infty} \lim_{n\to\infty} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} \right\|_\infty = 0 \tag{5.2}$$

Proof. To get (5.1) we will use the triangle inequality with three intermediate steps and estimate

$$\begin{aligned}
&\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty \qquad (5.3)\\
&\quad \le \left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|} \right\|_\infty + \left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} \right\|_\infty \\
&\quad + \left\| \frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&\quad + \left\| \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty
\end{aligned}$$

We will bound these four terms on the right separately:
For the first we note that

$$\hat{\Lambda}_{m,n} = \Lambda_n \setminus \Lambda_{m,n} \subseteq \partial^m(\Lambda_n)$$

and therefore

$$\Lambda_n^m \subseteq \Lambda_{m,n} \tag{5.4}$$

and

$$\left|\hat{\Lambda}_{m,n}\right| \le |\Lambda_n| - |\Lambda_n^m|\,, \tag{5.5}$$

see also Figure 5.1. This also means that Λ_n^m does not intersect $\hat{\Lambda}_{m,n}$.
Since $\Lambda_n^m \cup \partial^m(\Lambda_n^m) = \Lambda_n$we also see that

$$|\Lambda_n| - |\Lambda_n^m| \le |\partial^m(\Lambda_n^m)|\,. \tag{5.6}$$

From (5.4) we get

$$0 \le \frac{1}{|\Lambda_{m,n}|} - \frac{1}{|\Lambda_n|} \le \frac{1}{|\Lambda_n^m|} - \frac{1}{|\Lambda_n|} = \frac{|\Lambda_n| - |\Lambda_n^m|}{|\Lambda_n|\,|\Lambda_n^m|}$$

and with (4.3) follows

$$\begin{aligned}
\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|} \right\|_\infty &\le \frac{|\Lambda_n| - |\Lambda_n^m|}{|\Lambda_n|\,|\Lambda_n^m|} \left\| a(\Lambda_n,\omega) \right\|_\infty \\
&\le (D+E)\frac{|\Lambda_n| - |\Lambda_n^m|}{|\Lambda_n^m|} \qquad (5.7)
\end{aligned}$$

We use the almost additivity (A3) and the triangle inequality for the second term on the right side of (5.3) to get

$$\begin{aligned}\|a(\Lambda_n,\omega)-a(\Lambda_{m,n},\omega)\|_\infty &\le \left\|a(\Lambda_n,\omega)-a(\Lambda_{m,n},\omega)-a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty \\ &\quad+\left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty \\ &\le b(\Lambda_{m,n})+b(\hat{\Lambda}_{m,n})+\left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty \\ &\le b(\Lambda_{m,n})+D\left|\hat{\Lambda}_{m,n}\right|+\left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty\end{aligned}$$

and then (4.3), (5.4) and (5.5)to obtain

$$\begin{aligned}\left\|\frac{a(\Lambda_n,\omega)-a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|}\right\|_\infty &\le \frac{b(\Lambda_{m,n})+D\left|\hat{\Lambda}_{m,n}\right|+\left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty}{|\Lambda_{m,n}|} \\ &\le \frac{b(\Lambda_{m,n})+(2D+E)\left|\hat{\Lambda}_{m,n}\right|}{|\Lambda_{m,n}|} \\ &\le \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|}+\frac{(2D+E)\left(|\Lambda_n|-|\Lambda_n^m|\right)}{|\Lambda_n^m|}.\end{aligned} \tag{5.8}$$

For the third term on the right of (5.3) we use

$$|\Lambda_{m,n}|=|T_{m,n}|\,|\Lambda_m|$$

and apply (A3) again for

$$\begin{aligned}&\left\|\frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|}-\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}\frac{a(\Lambda_m+t,\omega)}{|\Lambda_m|}\right\|_\infty \\ &=\frac{1}{|T_{m,n}|\,|\Lambda_m|}\left\|a(\Lambda_{m,n},\omega)-\sum_{t\in T_{m,n}}a(\Lambda_m+t,\omega)\right\|_\infty \\ &\le \frac{1}{|T_{m,n}|\,|\Lambda_m|}\sum_{t\in T_{m,n}}b(\Lambda_m+t)=\frac{b(\Lambda_m)}{|\Lambda_m|}\end{aligned} \tag{5.9}$$

because of the properties of b from (A3).
For the fourth term of (5.3) we use the triangle inequality, then (A3) again as well as

(4.3) (similar to the first and second inequality in (5.8)) to get

$$\begin{aligned}
&\left\| \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} a(\Lambda_m + t, \omega) - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} a(\Lambda_m^r + t, \omega) \right\|_\infty \\
&\quad \leq \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \|a(\Lambda_m + t, \omega) - a(\Lambda_m^r + t, \omega)\|_\infty \\
&\quad \leq \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \left(b(\Lambda_m^r + t) + b(\Lambda_m \setminus \Lambda_m^r + t) + \|a((\Lambda_m \setminus \Lambda_m^r + t), \omega)\|_\infty\right) \\
&\quad \leq b(\Lambda_m^r) + b(\Lambda_m \setminus \Lambda_m^r) + (D + E) \left|\Lambda_m \setminus \Lambda_m^r\right| . \qquad (5.10)
\end{aligned}$$

Now we note that

$$|\Lambda_m \setminus \Lambda_m^r| = |\Lambda_m| - |\Lambda_m^r| \qquad (5.11)$$

and

$$\Lambda_m \setminus \Lambda_m^r = \Lambda_m \cap \partial^r(\Lambda_m) \subseteq \partial^r(\Lambda_m). \qquad (5.12)$$

We use (5.11) in combination with (5.10) and the properties of b from (A3) for

$$\begin{aligned}
&\left\| \frac{1}{|T_{m,n}|\,|\Lambda_m|} \sum_{t \in T_{m,n}} a(\Lambda_m + t, \omega) - \frac{1}{|T_{m,n}|\,|\Lambda_m|} \sum_{t \in T_{m,n}} a(\Lambda_m^r + t, \omega) \right\|_\infty \\
&\qquad \leq \frac{b(\Lambda_m^r) + (2D + E)\left(|\Lambda_m| - |\Lambda_m^r|\right)}{|\Lambda_m|}. \qquad (5.13)
\end{aligned}$$

Now we note again that $\Lambda_{m,n} = \Lambda_{\lfloor n/m \rfloor m}$ and finish the proof of (5.1) by applying (5.7), (5.8), (5.9) and (5.13) to the inequality (5.3).
For the proof of the convergence (5.2) we first note that from the properties of b from (A3) we immediately see that

$$\frac{b(\Lambda_{\lfloor n/m \rfloor m})}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} \xrightarrow{n \to \infty} 0 \text{ and } \frac{b(\Lambda_m)}{|\Lambda_m|} \xrightarrow{m \to \infty} 0.$$

Since Λ_m^r is a cube with side length $m - 2r$ and $|\Lambda_m| \geq |\Lambda_m^r|$ we also get that $\frac{b(\Lambda_m^r)}{|\Lambda_m|} \xrightarrow{m \to \infty} 0$. By (5.6) and (5.12) we have

$$\begin{aligned}
&\left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} \right\|_\infty \\
&\quad \leq \frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} + \frac{(3D + 2E)\,|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} + \frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D + E)\,|\partial^r(\Lambda_m)|}{|\Lambda_m|}.
\end{aligned}$$

From (4.7) we gain

$$\forall r \in \mathbb{N} : \frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|} \xrightarrow{m\to\infty} 0 \quad \text{and} \quad \forall m \in \mathbb{N} : \frac{|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} \xrightarrow{n\to\infty} 0.$$

Thus,

$$\begin{aligned}&\frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} + \frac{(3D+2E)\,|\partial^m(\Lambda_n^m)|}{|\Lambda_n^m|} + \frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D+E)\,|\partial^r(\Lambda_m)|}{|\Lambda_m|}\\ &\xrightarrow{n\to\infty} \frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D+E)\,|\partial^r(\Lambda_m)|}{|\Lambda_m|} \xrightarrow{m\to\infty} 0\end{aligned}$$

and we can deduce (5.2) from (5.1). □

Now we are able to use the previous result to establish the third step of (4.8), a convergence for the expected values of normalized admissible functions on cubes. Here, we first use the translation invariance of the admissible function and the measure to establish that the expected value of an admissible function on a cube is the same as the expected value of an averaged sum over the admissible function on disjoint translates of the same cube. Thus, we can use the tiling argument from before again. For two cubes Λ_n and Λ_m we can just find another very large cube Λ_k such that the averaged sum over the admissible functions of tilings of Λ_k by either Λ_n or Λ_m is very close to the admissible function on Λ_k. Since our bound from before does not depend on ω we can then use triangle inequalities to get the bounds we need for the proof of the Cauchy property.

Lemma 5.2
Let $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be an admissible function, $r, n \in \mathbb{N}$ such that $2r < n$ and $\Lambda_n := ([0,n) \cap \mathbb{Z})^d$. Then

$$n \mapsto \left(x \mapsto \frac{\mathbb{E}a(\Lambda_n^r, \cdot)(x)}{|\Lambda_n|} \right)$$

forms a Cauchy sequence in $\mathbb{B}$ and there exists a limit function $a^* \in \mathbb{B}$ that is monotonically increasing in x. Furthermore

$$\left\| \frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|} - a^* \right\|_\infty \le \frac{b(\Lambda_n) + b(\Lambda_n^r) + (2D+E)\,(|\Lambda_n| - |\Lambda_n^r|)}{|\Lambda_n|}. \tag{5.14}$$

Proof. First we note that because of (A1) of Definition 4.3 and the assumed property (M1) of Definition 4.1 the equation

$$\mathbb{E}a(\Lambda_n^r + t, \cdot) = \mathbb{E}a(\Lambda_n^r, \cdot) \circ \gamma_t = \mathbb{E}a(\Lambda_n^r, \cdot)$$

is true for all $t \in \mathbb{Z}^d$ and $n, r \in \mathbb{N}$ with $n > 2r$. Thus,

$$\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{a(\Lambda_n^r + t, \cdot)}{|\Lambda_n|} = \frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{\mathbb{E}a(\Lambda_n^r + t, \cdot)}{|\Lambda_n|} = \frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|}. \tag{5.15}$$

holds for $k > n$.
For the next step, let $n < m \in \mathbb{N}$ and $\delta > 0$. Then by (5.6) there exists a $k > 2m$ that is divisible by n and m (resulting in $\lfloor k/n \rfloor n = \lfloor k/m \rfloor m = k$) such that

$$\max\left\{\frac{b(\Lambda_k)}{|\Lambda_k|}, \frac{(|\Lambda_k| - |\Lambda_k^n|)}{|\Lambda_k^n|}, \frac{(|\Lambda_k| - |\Lambda_k^m|)}{|\Lambda_k^m|}\right\} < \delta$$

because of (A3) and (4.7). We use the triangle inequality and (5.15) to get

$$\begin{aligned} \left\|\frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|}\right\|_\infty &\leq \left\|\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{a(\Lambda_n^r + t, \cdot)}{|\Lambda_n|} - \frac{\mathbb{E}a(\Lambda_k, \cdot)}{|\Lambda_k|}\right\|_\infty \\ &+ \left\|\frac{\mathbb{E}a(\Lambda_k, \cdot)}{|\Lambda_k|} - \mathbb{E}\frac{1}{|T_{m,k}|}\sum_{t \in T_{m,k}} \frac{a(\Lambda_m^r + t, \cdot)}{|\Lambda_m|}\right\|_\infty. \end{aligned} \tag{5.16}$$

Both terms on the right side can be treated the same way by first considering a specific $x \in \mathbb{R}$ and using Lemma 5.1 for

$$\begin{aligned} &\left|\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{a(\Lambda_n^r + t, \cdot)(x)}{|\Lambda_n|} - \frac{\mathbb{E}a(\Lambda_k, \cdot)(x)}{|\Lambda_k|}\right| \\ &\leq \mathbb{E}\left|\frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{a(\Lambda_n^r + t, \cdot)(x)}{|\Lambda_n|} - \frac{a(\Lambda_k, \cdot)(x)}{|\Lambda_k|}\right| \\ &\leq \mathbb{E}\left(\frac{b\left(\Lambda_k\right)}{|\Lambda_k|} + \frac{(3D + 2E)\left(|\Lambda_k| - |\Lambda_k^n|\right)}{|\Lambda_k^n|} + \frac{b(\Lambda_n) + b(\Lambda_n^r) + (2D + E)\left(|\Lambda_n| - |\Lambda_n^r|\right)}{|\Lambda_n|}\right) \\ &\leq (3D + 2E + 1)\delta + \frac{b(\Lambda_n) + b(\Lambda_n^r) + (2D + E)\left(|\Lambda_n| - |\Lambda_n^r|\right)}{|\Lambda_n|}. \end{aligned}$$

Since x was arbitrary, this results in

$$\begin{aligned} &\left\|\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t \in T_{n,k}} \frac{a(\Lambda_n^r + t, \cdot)}{|\Lambda_n|} - \frac{\mathbb{E}a(\Lambda_k, \cdot)}{|\Lambda_k|}\right\|_\infty \\ &\quad \leq (3D + 2E + 1)\delta + \frac{b(\Lambda_n) + b(\Lambda_n^r) + (2D + E)\left(|\Lambda_n| - |\Lambda_n^r|\right)}{|\Lambda_n|} \end{aligned} \tag{5.17}$$

where the second term on the right side converges to 0 for rising n as established in Lemma 5.1. Therefore, for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$\frac{b(\Lambda_n) + b(\Lambda_n^r) + (2D + E)\left(|\Lambda_n| - |\Lambda_n^r|\right)}{|\Lambda_n|} < \varepsilon \ \forall n \geq N. \tag{5.18}$$

The bound (5.17) is also valid if n is exchanged for m. Thus, if $n, m \geq N$ then

$$\left\| \frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty \leq 2(3D + 2E + 1)\delta + 2\varepsilon$$

which follows from (5.16), (5.17) and (5.18). This proves that $\frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|}$ is a Cauchy sequence and has a uniform limit a^*. Since $\frac{a(\Lambda_n^r, \omega)}{|\Lambda_n|}$ is right-continuous, monotone increasing and bounded by 1 for every n and every ω this is also true for its expected value by the bounded convergence theorem, see for example [Kle20, Corollary 6.26]. Therefore, every $\mathbb{E}\frac{a(\Lambda_n^r, \cdot)}{|\Lambda_n|}$ lies in $\mathbb{B}$ and is monotone increasing, and as a limit this is true for a^* as well. By taking the limit $m \to \infty$ we can also gain the bound (5.14) for the distance of $\frac{\mathbb{E}a(\Lambda_n^r, \cdot)}{|\Lambda_n|}$ to a^* since δ was arbitrary. $\square$

With these bounds we have achieved everything we wanted from the geometric part for cubes, the only step of (4.8) left is the almost sure uniform convergence of the averaged sums. This will be the content of Chapter 6.
But first we want to investigate which properties of cubes were actually needed, and to what other sets we could extend these geometric bounds.

5.2 For monotiles

The main properties of cubes we actually needed in the previous section were on one hand the tilings we used to get to averaged sums of admissible functions, and on the other hand the fact that the surface of cubes grows slower than their volume with growing side-length. This is also true for a lot of other sets, for example rectangular cuboids. Thus, we can generalize our results by just using these two properties.

Definition 5.3 (Monotiling Følner sequence)
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a sequence of finite subsets of $\mathbb{Z}^d$ such that

(MF1) $(\Lambda_n)_{n\in\mathbb{N}}$ *has the Følner property, that is for all $r \in \mathbb{N}$ the sequence $\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}$ converges to* 0.

(MF2) $(\Lambda_n)_{n\in\mathbb{N}}$ *is monotiling, that is for all $n \in \mathbb{N}$ there is a set $T_n \subset \mathbb{Z}^d$ such that*

$$\mathbb{Z}^d = \bigcup_{t\in T_n} (\Lambda_n + t) = \Lambda_n + T_n$$

and $(\Lambda_n + t_1)$ *is disjoint from* $(\Lambda_n + t_2)$ *for all* $t_1 \neq t_2 \in T_n$ *and* T_n *is symmetric, meaning that*

$$t \in T_n \Leftrightarrow -t \in T_n.$$

Then $(\Lambda_n)_{n\in\mathbb{N}}$ *is called a* ***monotiling Følner sequence*** *of* $\mathbb{Z}^d$*.*

Remark 5.4. Condition (MF1) implies that for every monotiling Følner sequence there is a function $W\colon \mathbb{N} \to \mathbb{N}$ such that

$$\forall n > W(r): \ \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} < 1.$$

We need to show that the properties (MF1) and (MF2) are enough to establish analogous bounds and convergences to those we used for cubes in Lemma 5.1 and Lemma 5.2, namely the following four results.

Lemma 5.5
Let $(\Lambda_n)_{n\in\mathbb{N}}$ *be a monotiling Følner sequence of* $\mathbb{Z}^d$*. Then*

(a)

$$\Lambda_n \subseteq \partial^{\rho(m)}(\Lambda_n) \cup \Lambda_{m,n} \ \forall n, m \in \mathbb{N}$$

where $\rho(m) = \operatorname{diam}(\Lambda_m) = \max_{x,y\in\Lambda_m} \mathrm{d}_{\mathbb{Z}^d}(x, y)$ *is the* ***diameter*** *of* Λ_m *and*

$$\Lambda_{m,n} := \bigcup_{t\in T_{m,n}} (\Lambda_m + t)$$

with $T_{m,n} := \{t \in T_m : \Lambda_m + t \subseteq \Lambda_n\}$ *as in Lemma 5.1.*

(b) *For all* $m \in \mathbb{N}$*,* $|T_{m,n}| \geq 1$ *is true for all* $n > W(\rho(m))$ *and*

$$\lim_{n\to\infty} \frac{|\Lambda_n|}{|\Lambda_{m,n}|} = 1 \ \forall m \in \mathbb{N}$$

Furthermore for all $m \in \mathbb{N}$ *we have* $|T_{m,n}| \xrightarrow[n\to\infty]{} \infty$*.*

(c)

$$\lim_{n\to\infty} \frac{|\partial^r(\Lambda_{m,n})|}{|\Lambda_{m,n}|} = 0 \ \forall m, r \in \mathbb{N}$$

(d)

$$\lim_{n\to\infty} \frac{\left|\partial^{r'}(\Lambda_n \setminus \partial^r(\Lambda_n))\right|}{|\Lambda_n \setminus \partial^r(\Lambda_n)|} = 0 \ \forall r, r' \in \mathbb{N}$$

Proof. We will check each claim separately.

To (a) By definition we have $\Lambda_{m,n} \subseteq \Lambda_n$. Thus, we only need to check that if $x \in \Lambda_n \setminus \Lambda_{m,n}$, then $\mathrm{d}_{\mathrm{set}}(x, \Lambda_n^c) \leq \rho(m)$. We show this by contradiction and assume that there is an $x \in \Lambda_n \setminus \Lambda_{m,n}$ such that $\mathrm{d}_{\mathrm{set}}(x, \Lambda_n^c) > \rho(m)$. Since $\mathbb{Z}^d = \bigcup_{t \in T_n} (\Lambda_n + t)$ by definition, there exists a $t \in T_n$ such that $x \in \Lambda_m + t$ and $t \notin T_{m,n}$. But since $\mathrm{d}_{\mathrm{set}}(x, \Lambda_n^c) > \rho(m)$ every point in $\Lambda_m + t$ must also lie in Λ_n, which is a contradiction to $t \notin T_{m,n}$, proving (a) and following from this

$$\Lambda_n \setminus \Lambda_{m,n} \subseteq \partial^{\rho(m)}(\Lambda_n).$$

To (b) We now use (a) for the next part of the proof. Since $\Lambda_n \subseteq \partial^{\rho(m)}(\Lambda_n) \cup \Lambda_{m,n}$ we also have $|\Lambda_n| \leq \left|\partial^{\rho(m)}(\Lambda_n)\right| + |\Lambda_{m,n}|$ and

$$|\Lambda_n| - \left|\partial^{\rho(m)}(\Lambda_n)\right| \leq |\Lambda_{m,n}| \leq |\Lambda_n|,$$

leading to

$$1 - \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_n|} \leq \frac{|\Lambda_{m,n}|}{|\Lambda_n|} \leq 1.$$

With the function W from Remark 5.4 we see that for all $n > W(\rho(m))$ we have $|\Lambda_{m,n}| = |T_{m,n}|\,|\Lambda_m| > 0$. This is only possible if $|T_{m,n}| \geq 1$, leading to the first part of (b). The Følner property (MF1) means that

$$\lim_{n\to\infty} \frac{|\Lambda_{m,n}|}{|\Lambda_n|} = \lim_{n\to\infty} \frac{|T_{m,n}|\,|\Lambda_m|}{|\Lambda_n|} = 1 \tag{5.19}$$

which implies the second part of (b) via the reciprocal value. For the last part we just need to show $|\partial^r(\Lambda_n)| > 1$ for any $r > 1$ and any $n \in \mathbb{N}$. Every Λ_n contains at least one point x and since Λ_n is finite, there has to be a point y that is not in Λ_n. $\mathbb{Z}^d$ with next-neighbour-connections is a connected graph, so there has to be a $k > 0$ and a finite sequence $(x_j)_{0 \leq j \leq k}$ with $\mathrm{d}_{\mathbb{Z}^d}(x_j, x_{j+1}) = 1$ for all $0 \leq j \leq k-1$ and $x_0 = x$, $x_k = y$. Somewhere along this line there has to be an x_l such that $x_l \in \Lambda_n$ and $x_{l+1} \notin \Lambda_n$. Then $x_l \in \partial^r(\Lambda_n)$ and thus $|\partial^r(\Lambda_n)| > 1$. We can use this for

$$\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} \geq \frac{1}{|\Lambda_n|}$$

and since the left side converges to zero we gain $|\Lambda_n| \to \infty$. In combination with (5.19) this ensures $|T_{m,n}| \xrightarrow[n\to\infty]{} \infty$.

To (c) For (c) and (d) we first show a general result: For any two finite sets Λ and Λ' such that $\Lambda' \subset \Lambda$ and $\Lambda \setminus \Lambda' \subset \partial^r(\Lambda)$ for some $r > 0$ we have $\partial^{r'}(\Lambda') \subset \partial^{r'+r}(\Lambda)$ for all $r' > 0$.

Let $r' \in \mathbb{N}$, then for any $z \in \partial^{r'}(\Lambda')$ there are by definition two options

- $z \in \Lambda'$ and $\mathrm{d_{set}}(z, \Lambda'^C) \leq r'$
  We can further distinguish two cases here:
  - There is a $y \in \Lambda \setminus \Lambda'$ such that $\mathrm{d}_{\mathbb{Z}^d}(z, y) \leq r'$. Since $\Lambda \setminus \Lambda' \subseteq \partial^r(\Lambda)$ we have $\mathrm{d_{set}}(y, \Lambda^C) \leq r$. Then
  $$\mathrm{d_{set}}(z, \Lambda^C) \leq \mathrm{d}_{\mathbb{Z}^d}(z, y) + \mathrm{d_{set}}(y, \Lambda^C) \leq r' + r$$
  and consequently $z \in \partial^{r+r'}(\Lambda)$.
  - There is a $y \in \Lambda^C$ such that $\mathrm{d}_{\mathbb{Z}^d}(z, y) \leq r'$. In this case $z \in \partial^{r'}(\Lambda)$.
- $z \in \Lambda'^C$ and $\mathrm{d_{set}}(z, \Lambda') \leq r'$
  There are also two cases here:
  - If $z \in \Lambda \setminus \Lambda'$, then we have $z \in \partial^r(\Lambda)$.
  - If $z \in \Lambda^C$, then $\mathrm{d_{set}}(z, \Lambda) \leq r'$ follows from $\Lambda' \subseteq \Lambda$ and results in $z \in \partial^{r'}(\Lambda)$.

In all cases we have shown $z \in \partial^{r'+r}(\Lambda)$.
We first apply this result to $\Lambda_{m,n}$. We know from (a) that $\Lambda_n \setminus \Lambda_{m,n} \subset \partial^{\rho(m)}(\Lambda_n)$ and thus $\partial^r(\Lambda_{m,n}) \subset \partial^{r+\rho(m)}(\Lambda_n)$. Therefore,

$$\frac{|\partial^r(\Lambda_{m,n})|}{|\Lambda_{m,n}|} \leq \underbrace{\frac{\left|\partial^{r+\rho(m)}(\Lambda_n)\right|}{|\Lambda_n|}}_{\xrightarrow[n\to\infty]{(MF1)}0} \underbrace{\frac{|\Lambda_n|}{|\Lambda_{m,n}|}}_{\xrightarrow[n\to\infty]{(b)}1} \to 0$$

is true, which proves (c).

To (d) We can also apply the previous result to $\Lambda_n \setminus \partial^r(\Lambda_n)$ for some $r > 0$, since this means $\Lambda_n \setminus (\Lambda_n \setminus \partial^r(\Lambda_n)) \subset \partial^r(\Lambda_n)$. For $r' \in \mathbb{N}$ this leads to

$$\partial^{r'}(\Lambda_n \setminus \partial^r(\Lambda_n)) \subseteq \partial^{r'+r}(\Lambda_n).$$

With $W(r)$ as defined in Remark 5.4 we know that

$$1 - \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} = \frac{|\Lambda_n| - |\partial^r(\Lambda_n)|}{|\Lambda_n|} > 0 \ \forall n > W(r)$$

and thus

$$|\Lambda_n \setminus \partial^r(\Lambda_n)| \geq |\Lambda_n| - |\partial^r(\Lambda_n)| > 0$$

for $n > W(r)$. Then

$$\frac{\left|\partial^{r'}\left(\Lambda_n \setminus \partial^r(\Lambda_n)\right)\right|}{|\Lambda_n \setminus \partial^r(\Lambda_n)|} \leq \frac{\left|\partial^{r'+r}(\Lambda_n)\right|}{|\Lambda_n|} \frac{|\Lambda_n|}{|\Lambda_n \setminus \partial^r(\Lambda_n)|} \leq \frac{\left|\partial^{r'+r}(\Lambda_n)\right|}{|\Lambda_n|} \frac{|\Lambda_n|}{|\Lambda_n| - |\partial^r(\Lambda_n)|}$$
$$= \underbrace{\frac{\left|\partial^{r'+r}(\Lambda_n)\right|}{|\Lambda_n|}}_{\to 0} \underbrace{\frac{1}{1 - \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}}}_{\to 1} \xrightarrow[n\to\infty]{} 0$$

as claimed in (d)

□

These properties replace the ones for cubes that were illustrated by Figure 5.1. With these replacements we can follow the steps laid out in Lemma 5.1 to get an equivalent result for monotiling Følner sequences, we just need to require that $b(\Lambda_n)$ from (A3) grows slower than $|\Lambda_n|$ for all monotiling Følner sequences, and not just for cubes as before. To this end we will define admissible functions for more general geometries.

Definition 5.6 (Generalized admissible function)
*A function $a\colon \mathcal{C} \times \Omega \to \mathbb{B}$ is called **generalized admissible** if it satisfies*

- *conditions (A1), (A2), (A4), (A5) and (A6) of Definition 4.3.*
- *There exists a function $b : \mathcal{C} \to [0, \infty)$ such that for each $\omega \in \Omega$, pairwise disjoint $\Lambda_1, ..., \Lambda_n \in \mathcal{C}$ and $\Lambda := \bigcup_{i=1}^n \Lambda_i$ we have*

$$\left\| a(\Lambda, \omega) - \sum_{i=1}^{n} a(\Lambda_i, \omega) \right\|_\infty \leq \sum_{i=1}^{n} b(\Lambda_i),$$

and b satisfies

  - *for all $\Lambda \in \mathcal{C}$ and $z \in \mathbb{Z}^d$ we have $b(\Lambda) = b(\Lambda + z)$*
  - *there is a $D > 0$ such that $b(\Lambda) \leq D\,|\Lambda|$ for all $\Lambda \in \mathcal{C}$*
  - *if $(\Lambda_n)_{n\in\mathbb{N}}$ is a Følner sequence, i.e.*

$$\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} \xrightarrow[n\to\infty]{} 0,$$

then

$$\lim_{n\to\infty} \frac{b(\Lambda_n)}{|\Lambda_n|} = 0.$$

We further call $a\colon \mathcal{C} \times \Omega \to \mathbb{B}$ a ***generalized admissible function with a sub-additive boundary function*** *if the function b above additionally satisfies $b(\Lambda_1 \cup \Lambda_2) \le b(\Lambda_1) + b(\Lambda_2)$, $b(\Lambda_1 \cap \Lambda_2) \le b(\Lambda_1) + b(\Lambda_2)$ and $b(\Lambda_1 \setminus \Lambda_2) \le b(\Lambda_1) + b(\Lambda_2)$ for all $\Lambda_1, \Lambda_2 \in \mathcal{C}$.*

With these generalized admissible functions we can formulate a lemma corresponding to Lemma 5.1. The necessary steps to achieve this result were already laid out in [SSV18], but it was not formulated explicitly before.

Lemma 5.7
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence and let $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be a generalized admissible function as in Definition 5.6. Then for all $\omega \in \Omega$, $r > 0$, $m > W(r)$ and $n > W(\rho(m))$ the inequality

$$\begin{aligned} &\left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} \right\|_\infty \\ \le \quad & (3D + 2E) \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} + \frac{b(\Lambda_m)}{|\Lambda_m|} + \frac{b(\Lambda_m^r)}{|\Lambda_m|} + (2D + E) \frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|} \end{aligned}$$

holds, where $\rho(m) = \max_{x,y \in \Lambda_m} \mathrm{d}_{\mathbb{Z}^d}(x, y)$ and $W(r)$ is the function defined in Remark 5.4. Furthermore

$$\lim_{m\to\infty} \lim_{n\to\infty} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} \right\|_\infty = 0.$$

This result is actually valid for more graphs than just $\mathbb{Z}^d$ with connections to next neighbours. This proof can also be applied to the case of Cayley graphs of finitely generated amenable groups (see Section 8.1 for details) if they have a monotiling Følner sequence.

Proof. We follow the proof of Lemma 5.1 to get the claimed bound. We use the triangle inequality and again bound the four terms on the right separately as

$$\begin{aligned}
&\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&\le \left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|} \right\|_\infty + \left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} \right\|_\infty \\
&+ \left\| \frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&+ \left\| \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty .
\end{aligned}$$

For the first term we use (a) in Lemma 5.5 for

$$0 \le \frac{1}{|\Lambda_{m,n}|} - \frac{1}{|\Lambda_n|} = \frac{|\Lambda_n| - |\Lambda_{m,n}|}{|\Lambda_{m,n}|\,|\Lambda_n|} \le \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|\,|\Lambda_n|}$$

and (4.3) to get

$$\begin{aligned}
\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|} \right\|_\infty &\le \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|\,|\Lambda_n|} \|a(\Lambda_n,\omega)\|_\infty \\
&\le (D+E)\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_n|}\frac{|\Lambda_n|}{|\Lambda_{m,n}|}. \qquad (5.20)
\end{aligned}$$

For the second term we follow (5.8) up to the second inequality and then use $\hat{\Lambda}_{m,n} = \Lambda_n \setminus \Lambda_{m,n} \subseteq \partial^{\rho(m)}(\Lambda_n)$ from (a) again to get

$$\begin{aligned}
\left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} \right\|_\infty &\le \frac{b(\Lambda_{m,n}) + D\left|\hat{\Lambda}_{m,n}\right| + \left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty}{|\Lambda_{m,n}|} \\
&\le \frac{b(\Lambda_{m,n}) + (2D+E)\left|\hat{\Lambda}_{m,n}\right|}{|\Lambda_{m,n}|} \\
&\le \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} + \frac{(2D+E)\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} \\
&\le \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} + (2D+E)\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_n|}\frac{|\Lambda_n|}{|\Lambda_{m,n}|}. \qquad (5.21)
\end{aligned}$$

For the third term we reuse (5.9) for

$$\left\| \frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} \right\|_\infty \le \frac{b(\Lambda_m)}{|\Lambda_m|} \qquad (5.22)$$

For the fourth term we use (5.10) and then $\Lambda_m \setminus \Lambda_m^r \subseteq \partial^r(\Lambda_m)$ to get

$$\begin{aligned}
&\left\| \frac{1}{|T_{m,n}|\,|\Lambda_m|} \sum_{t\in T_{m,n}} a(\Lambda_m + t,\omega) - \frac{1}{|T_{m,n}|\,|\Lambda_m|} \sum_{t\in T_{m,n}} a(\Lambda_m^r + t,\omega) \right\|_\infty \\
&\quad \le \frac{b(\Lambda_m^r) + b(\Lambda_m \setminus \Lambda_m^r) + (D+E)\,|\Lambda_m \setminus \Lambda_m^r|}{|\Lambda_m|} \\
&\quad \le \frac{b(\Lambda_m^r)}{|\Lambda_m|} + (2D+E)\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|}.
\end{aligned} \tag{5.23}$$

Combining (5.20), (5.21), (5.22) and (5.23) leads to

$$\begin{aligned}
&\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&\quad \le (3D+2E)\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_n|}\frac{|\Lambda_n|}{|\Lambda_{m,n}|} + \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} + \frac{b(\Lambda_m)}{|\Lambda_m|} + \frac{b(\Lambda_m^r)}{|\Lambda_m|} + (2D+E)\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|}
\end{aligned}$$

as claimed. We showed in (c) and (d) of Lemma 5.5 that the sequences $(\Lambda_{m,n})_{n\in\mathbb{N}}$ and $(\Lambda_m^r)_{m\in\mathbb{N}}$ fulfil condition (MF1) for all $m > W(r) \in \mathbb{N}$. Since we required that $\frac{b(\Lambda_n)}{|\Lambda_n|} \xrightarrow[n\to\infty]{} 0$ for all such sequences, we can conclude with (b) and

$$\frac{b(\Lambda_m^r)}{|\Lambda_m|} \le \frac{b(\Lambda_m^r)}{|\Lambda_m^r|} \text{ if } m > W(r)$$

that

$$\begin{aligned}
&\lim_{m\to\infty}\lim_{n\to\infty} \left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&\overset{m>W(r)}{\le} \lim_{m\to\infty}\left(\frac{b(\Lambda_m)}{|\Lambda_m|} + \frac{b(\Lambda_m^r)}{|\Lambda_m^r|} + (2D+E)\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|} \right) \\
&= 0
\end{aligned}$$

□

We can also extend Lemma 5.2 to this setting and get

Lemma 5.8
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence and let $a : \mathcal{C}\times\Omega \to \mathbb{B}$ be a generalized admissible function as defined in Definition 5.6. Then $\frac{\mathbb{E}a(\Lambda_n^r,\cdot)}{|\Lambda_n|}$ (defined pointwise) forms a $\|\cdot\|_\infty$-Cauchy sequence and there exists a limit function $a^* \in \mathbb{B}$ that is also monotone increasing. Furthermore for $n > W(r)$ we have

$$\left\| \frac{\mathbb{E}a(\Lambda_n^r,\cdot)}{|\Lambda_n|} - a^* \right\|_\infty \le \frac{b(\Lambda_n)}{|\Lambda_n|} + \frac{b(\Lambda_n^r)}{|\Lambda_n|} + (2D+E)\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}. \tag{5.24}$$

Proof. We use (5.15) from Lemma 5.7 again and this time choose for $n, m > W(r)$ and $\delta > 0$ a $k > \max(W(\rho(n)), W(\rho(m)))$ such that

$$\max\left\{\frac{\left|\partial^{\rho(n)}(\Lambda_k)\right|}{|\Lambda_{n,k}|}, \frac{\left|\partial^{\rho(m)}(\Lambda_k)\right|}{|\Lambda_{m,k}|}, \frac{b(\Lambda_{n,k})}{|\Lambda_{n,k}|}, \frac{b(\Lambda_{m,k})}{|\Lambda_{m,k}|}\right\} < \delta.$$

This is possible since the terms in brackets converge to 0 for growing k as shown in Lemma 5.7.
After this we use Lemma 5.7 for

$$\begin{aligned}
&\left|\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t\in T_{n,k}}\frac{a(\Lambda_n^r+t,\cdot)(x)}{|\Lambda_n|}-\frac{\mathbb{E}a(\Lambda_k,\cdot)(x)}{|\Lambda_k|}\right| \\
&\quad\le \mathbb{E}\left|\frac{1}{|T_{n,k}|}\sum_{t\in T_{n,k}}\frac{a(\Lambda_n^r+t,\cdot)(x)}{|\Lambda_n|}-\frac{a(\Lambda_k,\cdot)(x)}{|\Lambda_k|}\right| \\
&\quad\le \mathbb{E}\left((3D+2E)\frac{\left|\partial^{\rho(n)}(\Lambda_k)\right|}{|\Lambda_{n,k}|}+\frac{b(\Lambda_{n,k})}{|\Lambda_{n,k}|}+\frac{b(\Lambda_n)}{|\Lambda_n|}+\frac{b(\Lambda_n^r)}{|\Lambda_n|}+(2D+E)\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}\right) \\
&\quad\le (3D+2E+1)\delta+\frac{b(\Lambda_n)+b(\Lambda_n^r)+(2D+E)\,|\partial^r(\Lambda_n)|}{|\Lambda_n|}.
\end{aligned}$$

This results in

$$\begin{aligned}
&\left\|\mathbb{E}\frac{1}{|T_{n,k}|}\sum_{t\in T_{n,k}}\frac{a(\Lambda_n^r+t,\cdot)}{|\Lambda_n|}-\frac{\mathbb{E}a(\Lambda_k,\cdot)}{|\Lambda_k|}\right\|_\infty \\
&\quad\le (3D+2E+1)\delta+\frac{b(\Lambda_n)+b(\Lambda_n^r)+(2D+E)\,|\partial^r(\Lambda_n)|}{|\Lambda_n|}
\end{aligned}$$

just like in Lemma 5.2. We know from Lemma 5.7 that the second term on the right side of this equation converges to 0 with rising n. From this point on we can follow the proof of Lemma 5.2. □

Now that we have found general results for the geometric steps of the proof, we can move on to the probabilistic parts.

6 Analytic and probabilistic methods for empirical processes

In the previous chapter we established uniform bounds of the type

$$\left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a' \right\|_\infty \leq c_1(n,m) + c_2(m) + \sup_{x\in\mathbb{R}} \left| \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} - \mathbb{E}\frac{a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right|$$

with $\lim_{n\to\infty, m\to\infty} c_1(n,m) + c_2(m) = 0$ like we planned to in Section 4.2. We ensured that the distances between the translates $\Lambda_m^r + t$ are large enough for the eigenvalue-counting functions on them to be independent (see Chapter 7 for the exact arguments), and with that we hope to use something like a law of large numbers for the last missing step. For any fixed energy $x \in \mathbb{R}$ this would be enough, but the probabilistic challenge we have to overcome is the fact that we want a concentration equality for a supremum over an infinite set. Schumacher, Schwarzenberger and Veselić solved this in [SSV17] with a Glivenko-Cantelli-type argument that relied on the fact that the eigenvalue-counting functions are monotone in all ω_z, but it was not possible to gain explicit constants. The approach to find those constants now will use a different probabilistic argument, but the general idea is similar: due to the boundedness and monotonicity of the evcfs (but this time not in ω but in the argument x) there is a kind of “typical” behaviour of evcfs with close arguments.
The general idea now is to shift our point of view a bit. We will use

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} - \mathbb{E}\frac{a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right| = \sup_{f\in\mathcal{F}'} \left| \frac{1}{s} \left(\sum_{t=1}^{s} f(X_t) - \mathbb{E}f(X_t) \right) \right|$$

where $\mathcal{F}' = \left\{ \frac{a_{\Lambda_m^r}(x)}{|\Lambda_m|} : x \in \mathbb{R} \right\}$, $s = |T_{m,n}|$ and $X_t = (\gamma_t\omega)_{\Lambda_m^r}$. Instead of a supremum over the real numbers we will instead treat the eigenvalue-counting functions for each x as a separate function and then take the supremum over all of these functions. Because the evcfs are monotone increasing in x these functions will not differ much, which is what we will use.
Our goal is to prove a theorem of the type

Metatheorem
Let $\mathcal{F}$ be a countable set of measurable functions $f : Y \to \mathbb{R}$ for a measurable space Y with $0 \leq f \leq 1$ and let $X_1, X_2, \ldots$ be i.i.d. random variables taking values in Y. Then there are $c_3(\kappa, s)$ and $c_4(\kappa, s)$ such that

$$\forall \kappa > 0 : \ \mathbb{P}\left(\sup_{f\in\mathcal{F}} \left| \frac{1}{s}\sum_{i=1}^{s} f(X_i) - \mathbb{E}f(X_1)\right| \geq \kappa\right) \leq c_3(\kappa, s) \cdot \exp\left(c_4(\kappa, s)\right)$$

with as much knowledge about c_3 and c_4 as possible.
An intuition why a roughly exponential bound may exist can for example be derived from Bernstein's inequality (see Theorem 6.6 for details) that states that

$$\mathbb{P}\left(\frac{1}{s}\sum_{i=1}^{s}(X_i - \mathbb{E}X_i) \geq x\right) \leq \mathrm{e}^{-\frac{1}{2}\frac{s^2x^2}{d+esx}}$$

for i.i.d. bounded random variables $X_i \colon \Omega \to \mathbb{R}$ and constants d, e.
We will present two different results of this type whose proof follows the proof of Theorems 2.5.6 and 2.14.2 in [vW96] for the L^1 norm of a supremum over a countable set of functions of an empirical process, but with explicit computation of the involved constants, and results from [Pol90] are used to replace the L^1-norm with the Orlicz norm. The foundation here are results due to van der Vaart and Wellner, and part of a longer series of similar results due to among others Dudley, Ossiander, Arcones, Giné, Ledoux, Talagrand and Pollard. See for example [Dud78], [Oss87], [AG93], [LT91], [Tal94] and [Pol90] as well as the references therein and those in [vW96].
Section 6.1 will give the definition of the Orlicz norm and show some of its relevant properties. After that in Section 6.2 we will define the empirical process and first prove a bound on the Orlicz norm of a supremum over a finite number of functions of an empirical process. In Section 6.3 we will then prove Theorem 6.9, a bound for the Orlicz norm of a supremum over a countable set of functions of an empirical process, which is our main probabilistic ingredient. We will use this Theorem in Section 6.4 to derive two concentration inequalities, and Section 6.5 contains the necessary additional calculations to use these inequalities to quantify uniform convergence of monotone increasing, right-continuous random functions in $[0, 1]$, which allows us to get results for admissible functions in the next chapter.
The constants we arrive at will most likely be far from optimal, but they will give the first explicit quantitative results for the uniform convergence of the admissible functions defined in Section 4.1. It might be possible to improve the found constants further with other results from empirical process theory or uniform laws of large numbers, since there is already a large body of work in this field (see for example [vW96] and [Tal94]), albeit often without explicit constants. In this case Sections 6.4 and 6.5 should allow for easy integration of any other results into the surrounding setting.
First, we will prove some preliminaries that will be used later, mostly centred around the Orlicz norm and a supremum over a finite set of functions.

6.1 The Orlicz norm

This section will introduce the Orlicz norm, and some useful properties needed in the following sections.

Definition 6.1 (Orlicz norm)
The ***Orlicz norm*** *of a real random variable X associated to a monotone increasing, non-constant convex function $\Psi : [0,\infty) \to [0,\infty)$ with $0 \le \Psi(0) < 1$ is*

$$\|X\|_\Psi := \inf\left\{C > 0 : \mathbb{E}\Psi\left(\frac{|X|}{C}\right) \le 1\right\}$$

where $\inf(\emptyset) := \infty$.
The Orlicz norm for $\Psi(x) = x^p$ is the usual L^p-norm. If $|X| \le |Y|$ almost surely, then $\|X\|_\Psi \le \|Y\|_\Psi$. Of special interest are the Orlicz norms associated to the functions

$$\psi_{p,M}(x) := \frac{1}{M}\mathrm{e}^{(x^p)}$$

for $M \ge 2$ and $p \ge 1$ and especially $p = 1$ and $p = 2$. Notable is further the case $M = 2$, since the resulting Orlicz norm is equivalent to the one for

$$\Psi_p(x) := \mathrm{e}^{(x^p)} - 1.$$

In the rest of this proof we will refer to these specific functions with $\psi_{p,M}$ and Ψ_p and use Φ for general monotone increasing, non-constant, convex functions with $0 \le \Phi(0) < 1$.

We will first check that this does indeed define a norm:
$\|tX\|_\Phi = |t| \|X\|_\Phi$ immediately follows from the definition of the norm.
Furthermore $\|X\|_\Phi = 0$ implies $\mathbb{P}(|X| > 0) = 0$ and therefore $X = 0$ almost surely. This follows from the properties of Φ, since if there are $0 \le a < b \in \mathbb{R}$ with $\Phi(a) < \Phi(b)$, then for all $x > b$ there is $\Theta = \frac{x-b}{x-a}$ which fulfils $b = \Theta a + (1-\Theta)x$. The convexity of Φ then results in

$$\Phi(b) \le \Theta\Phi(a) + (1-\Theta)\Phi(x)$$

and therefore

$$\frac{\Phi(b) - \frac{x-b}{x-a}\Phi(a)}{\frac{b-a}{x-a}} = \frac{\Phi(b) - \Theta\Phi(a)}{1-\Theta} \le \Phi(x).$$

Rearranging leads to

$$\Phi(a) + (x-a)\frac{\Phi(b) - \Phi(a)}{b-a} \le \Phi(x)$$

and thus

$$\lim_{x\to\infty} \Phi(x) = \infty.$$

Now if $\mathbb{P}(|X| > 0) \neq 0$ there is an $n \in \mathbb{N}$ such that $\mathbb{P}(|X| > 1/n) > 0$, and for all $C > 0$ we have

$$\mathbb{E}\Phi\left(\frac{|X|}{C}\right) \geq \mathbb{P}(|X| > 1/n)\Phi\left(\frac{1}{nC}\right).$$

Now since $\lim_{C \searrow 0} \Phi\left(\frac{1}{nC}\right) = \infty$ there is a $C' > 0$ such that $\Phi\left(\frac{1}{nC}\right) > \frac{1}{\mathbb{P}(|X|>1/n)}$ for all $C \leq C'$. Therefore, $\mathbb{E}\phi\left(\frac{|X|}{C}\right) > 1 \ \forall C \leq C'$, leading to $\|X\|_\Phi \geq C' > 0$. As a consequence $\|X\|_\Phi = 0$ implies $\mathbb{P}(|X| > 0) = 0$.

Now we just have to prove the triangle inequality. For this we first show

$$\mathbb{E}\Phi\left(\frac{|X|}{\|X\|_\Phi}\right) \leq 1 \tag{6.1}$$

for random variables with non-zero Orlicz norm. If the Orlicz norm is infinite this follows from $\Phi(0) < 1$. For any random variable with finite positive Φ-Orlicz norm and any sequence $k_n \searrow \|X\|_\Phi$ we have

$$\mathbb{E}\Phi\left(\frac{|X|}{k_n}\right) \leq 1.$$

By monotone convergence (and continuity of Φ on $(0, \infty)$ due to convexity, see [Kle20, Theorem 7.7]) follows (6.1).

We can now show $\|X + Y\|_\Phi \leq \|X\|_\Phi + \|Y\|_\Phi$ by showing

$$\mathbb{E}\Phi\left(\frac{|X+Y|}{\|X\|_\Phi + \|Y\|_\Phi}\right) \leq \mathbb{E}\Phi\left(\frac{|X| + |Y|}{\|X\|_\Phi + \|Y\|_\Phi}\right) \overset{!}{\leq} 1$$

since Φ is monotone increasing. The first inequality follows from the monotonicity of Φ and for the second inequality we use

$$\frac{|X| + |Y|}{\|X\|_\Phi + \|Y\|_\Phi} = \frac{\|X\|_\Phi}{\|X\|_\Phi + \|Y\|_\Phi}\frac{|X|}{\|X\|_\Phi} + \left(1 - \frac{\|X\|_\Phi}{\|X\|_\Phi + \|Y\|_\Phi}\right)\frac{|Y|}{\|Y\|_\Phi}.$$

The convexity of Φ and (6.1) lead to

$$\mathbb{E}\Phi\left(\frac{|X| + |Y|}{\|X\|_\Phi + \|Y\|_\Phi}\right) \leq \frac{\|X\|_\Phi}{\|X\|_\Phi + \|Y\|_\Phi} + \left(1 - \frac{\|X\|_\Phi}{\|X\|_\Phi + \|Y\|_\Phi}\right) = 1$$

as claimed, and with this we have shown all needed properties of a norm.

The Orlicz norms, especially for $\psi_{p,M}$ and Ψ_p, have some some helpful properties that tie them to random variables with exponential tails that we will use later. These results and some of their proofs are collected from Section 2.2 and Lemma 2.2.1 in [vW96] and Section 1 of [Pol90].

Lemma 6.2
Some important properties of the Orlicz norm are:

(O1) *If* $\|X\|_\Phi \leq D < \infty$ *then* $\mathbb{E}\Phi\left(\frac{|X|}{\|X\|_\Phi}\right) \leq 1$ *and* $\mathbb{P}(|X| \geq y) \leq \frac{1}{\Phi(\frac{y}{D})}$

(O2) *If* X *is a real random variable with* $\mathbb{P}(|X| > x) \leq B\mathrm{e}^{-Cx^p}$ $\forall x > 0$ *with constants* $B, C > 0$ *and* $p \geq 1$ *then*

$$\|X\|_{\psi_{p,M}} \leq \left(\frac{B+M-1}{(M-1)C}\right)^{1/p}$$

(O3)

$$\|X\|_{L^p} \leq \|X\|_{\psi_{p,2}},$$

and thus

$$\mathbb{E}\,|X| = \|X\|_{L^1} \leq \|X\|_{\psi_{1,2}}.$$

(O4) *If* $M \geq 2$ *and* $X = c$ *almost surely, then*

$$\|X\|_{\psi_{p,M}} \leq \frac{|c|}{(\log(M))^{1/p}} \leq \frac{|c|}{(\log(2))^{1/p}}$$

(O5) *Let* X *and* Y *be real random variables, then*

$$\|XY\|_{\psi_{1,M}} \leq \|X\|_{\psi_{2,M}} \|Y\|_{\psi_{2,M}}.$$

As a consequence we have

$$\|X\|_{\psi_{1,M}} \leq \frac{1}{\sqrt{\log(M)}} \|X\|_{\psi_{2,M}}$$

for $M \geq 2$.

Especially the first two properties illustrate the connection between concentration inequalities and Orlicz norms, echoing the broader connection between the tails and moment estimates for random variables, which can be seen for example in [BLM13].

Proof. Since $\|X\|_\Phi = 0$ implies $\mathbb{P}(|X| > 0) = 0$, we do not have to consider this special case here.
Now to the proofs of the properties of the norm:

To (O1) We already proved the first part when we showed that the Orlicz norm is indeed a norm. Since Φ is convex and increasing, we can apply Markov's inequality for convex functions

$$\mathbb{P}(|X| \geq y)\Phi\left(\frac{y}{\|X\|_\Phi}\right) \leq \mathbb{E}\Phi\left(\frac{|X|}{\|X\|_\Phi}\right)$$

to get

$$\mathbb{P}(|X| \geq y) \leq \frac{\mathbb{E}\Phi\left(\frac{|X|}{\|X\|_\Phi}\right)}{\Phi\left(\frac{y}{\|X\|_\Phi}\right)} \leq \frac{1}{\Phi\left(\frac{y}{\|X\|_\Phi}\right)} \leq \frac{1}{\Phi\left(\frac{y}{D}\right)}.$$

To (O2) For this part, we calculate for $D < C$

$$\mathbb{E}\left(\mathrm{e}^{D|X|^p} - 1\right) = \mathbb{E}\int_0^{|X|^p} D\mathrm{e}^{Ds}\mathrm{d}s = \mathbb{E}\int_0^\infty \mathbb{1}_{\{s<|X|^p\}} D\mathrm{e}^{Ds}\mathrm{d}s.$$

Continuous functions, projections on coordinates and sums and products of measurable functions are measurable, so $\mathbb{1}_{\{0<|X|^p-s\}}D\mathrm{e}^{Ds}$ is $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$-measurable. It follows from Fubinis theorem [Kle20, Theorem 14.19] and the assumption on the tails of the distribution of $|X|$ that

$$\mathbb{E}\left(\mathrm{e}^{D|X|^p} - 1\right) = \int_0^\infty \mathbb{P}(|X| > s^{1/p})D\mathrm{e}^{Ds}\mathrm{d}s \leq \int_0^\infty BD\mathrm{e}^{(D-C)s}\mathrm{d}s = \frac{BD}{C-D}.$$

Now

$$\mathbb{E}\frac{1}{M}\mathrm{e}^{D|X|^p} = \mathbb{E}\frac{1}{M}\left(\mathrm{e}^{D|X|^p} - 1 + 1\right) \leq \frac{1}{M}\left(\frac{BD}{C-D} + 1\right)$$

This is less than or equal to 1 if $D \leq \frac{(M-1)C}{B+M-1}$ or equivalently if $D^{-1/p} \geq \left(\frac{B+M-1}{(M-1)C}\right)^{1/p}$.

Comparison with the definition of the Orlicz norm leads to $\|X\|_{\psi_{p,M}} \leq \left(\frac{B+M-1}{(M-1)C}\right)^{1/p}$.

To (O3) This follows from $x^p \leq \mathrm{e}^{(x^p)} - 1$ and the fact that

$$\mathbb{E}\exp\left(\left(\frac{|X|}{C}\right)^p\right) - 1 \leq 1 \Leftrightarrow \mathbb{E}\frac{\exp\left(\left(\frac{|X|}{C}\right)^p\right)}{2} \leq 1.$$

To (O4) If $d = \frac{|c|}{(\log(M))^{1/p}}$, then

$$\mathbb{E}\psi_{p,M}\left(\frac{|X|}{d}\right) = \frac{\exp\left(\left(\frac{|c|}{d}\right)^p\right)}{M} = 1$$

To (O5) The first inequality follows from Young's inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ for $a, b \geq 0$. We apply this formula twice and use (O1) for

$$\mathbb{E}\exp\left(\frac{|XY|}{\|X\|_{\psi_{2,M}}\|Y\|_{\psi_{2,M}}}\right) \leq \mathbb{E}\exp\left(\frac{1}{2}\left(\frac{|X|}{\|X\|_{\psi_{2,M}}}\right)^2\right)\exp\left(\frac{1}{2}\left(\frac{|Y|}{\|Y\|_{\psi_{2,M}}}\right)^2\right)$$
$$\leq \frac{1}{2}\mathbb{E}\exp\left(\left(\frac{|X|}{\|X\|_{\psi_{2,M}}}\right)^2\right) + \frac{1}{2}\mathbb{E}\exp\left(\frac{1}{2}\left(\frac{|Y|}{\|Y\|_{\psi_{2,M}}}\right)^2\right) \leq 1$$

which proves the inequality by the definition of the norm. The second inequality is a consequence of the first together with (O4).

□

6.2 The empirical process and suprema over finite sets

First we start with the object we consider for the next steps, the empirical process.

Definition 6.3 (Empirical process ([vW96, Chapter 2.1]))
Let $s \in \mathbb{N}$ and let $X_1, X_2, ..., X_s$ be i.i.d. random variables taking values in a measurable space Y with image measure P. Then the empirical process applied to a P-integrable function $f : Y \to \mathbb{R}$ is

$$\mathbb{G}_s(f) := \frac{1}{\sqrt{s}}\left(\sum_{i=1}^{s}\left(f(X_i) - \mathbb{E}f(X_i)\right)\right). \tag{6.2}$$

Let $\mathcal{F}$ be a set of P-integrable functions $f : Y \to \mathbb{R}$ then the map $\mathcal{F} \ni f \mapsto \mathbb{G}_s(f)$ is called the $\mathcal{F}$-indexed empirical process, which is linear in f.

The aim of this chapter will be to give the following bound on the Orlicz norm of the maximum of an empirical process indexed by a finite set of bounded functions based on Lemma 2.2.10 in [vW96], that will be used to gain a bound for a countable maximum in the next section.

Lemma 6.4
In addition to the notation of Definition 6.3, let $\mathcal{F}$ be a finite set of measurable bounded functions. Then for $M \geq 2$

$$\left\|\max_{f\in\mathcal{F}}|\mathbb{G}_s(f)|\right\|_{\psi_{1,M}} \leq K'_{\psi,M}\left(\max_{f\in\mathcal{F}}\frac{\|f\|_\infty}{\sqrt{s}}(2+\log(|\mathcal{F}|)) + \max_{f\in\mathcal{F}}\|f\|_{L^2(P)}\sqrt{2+\log(|\mathcal{F}|)}\right) \tag{6.3}$$

where $K'_{\psi,M} = \frac{4(M+1)}{\log(3/2)(M-1)}$.

The first prerequisite for the proof of Lemma 6.4 is the following result, which is Lemma 3.2 from [Pol90].

Lemma 6.5
For $p = 1, 2$ and real random variables $X_1, ..., X_m$,

$$\left\| \max_{1 \le i \le m} X_i \right\|_{\psi_{p,M}} \le \frac{(2 + \log(m))^{1/p}}{\log(3/2)} \max_{1 \le i \le m} \|X_i\|_{\psi_{p,M}}$$

if $M \ge 2$.

Proof. Since the lemma is trivial if $\max_{1 \le i \le m} \|X_i\|_{\psi_{p,M}} = \infty$, we only have to consider the case of a finite maximum. The function $\psi_{p,M}(x) = \frac{1}{M} \mathrm{e}^{x^p}$ is monotone increasing, so for all $C, D, a > 0$

$$\begin{aligned} \psi_{p,M}\left(\frac{\left| \max_{1 \le i \le m} X_i \right|}{C} \right) &\le \psi_{p,M}\left(\frac{\max_{1 \le i \le m} |X_i|}{C} \right) \\ &\le \psi_{p,M}(a) + \int_a^\infty \mathbb{1}\left\{ D \frac{\max_{1 \le i \le m} |X_i|}{C} > Dx \right\} \psi'_{p,M}(x) \mathrm{d}x. \end{aligned}$$

Now

$$D \frac{\max_{1 \le i \le m} |X_i|}{C} \mathbb{1}\left\{ D \frac{\max_{1 \le i \le m} |X_i|}{C} > Dx \right\} \ge Dx \mathbb{1}\left\{ D \frac{\max_{1 \le i \le m} |X_i|}{C} > Dx \right\},$$

and therefore

$$\psi_{p,M}\left(D \frac{\max_{1 \le i \le m} |X_i|}{C} \right) \mathbb{1}\left\{ D \frac{\max_{1 \le i \le m} |X_i|}{C} > Dx \right\} \ge \psi_{p,M}(Dx) \mathbb{1}\left\{ D \frac{\max_{1 \le i \le m} |X_i|}{C} > Dx \right\},$$

resulting in

$$\begin{aligned} \psi_{p,M}\left(\frac{\max_{1 \le i \le m} |X_i|}{C} \right) &\le \psi_{p,M}(a) + \int_a^\infty \frac{\psi_{p,M}\left(D \frac{\max_{1 \le i \le m} |X_i|}{C} \right)}{\psi_{p,M}(Dx)} \psi'_{p,M}(x) \mathrm{d}x \\ &\le \psi_{p,M}(a) + \sum_{i=1}^m \int_a^\infty \frac{\psi_{p,M}\left(D \frac{|X_i|}{C} \right)}{\psi_{p,M}(Dx)} \psi'_{p,M}(x) \mathrm{d}x. \end{aligned}$$

To estimate the Orlicz norm we are interested in the expected value of this term. Setting $C := D \max_{1 \le i \le m} \|X_i\|_{\psi_{p,M}}$ and using property (O1) in Lemma 6.2 results in

$$\begin{aligned}\mathbb{E}\psi_{p,M}\left(\frac{\max_{1\le i\le m}|X_i|}{D\max_{1\le i\le m}\|X_i\|_{\psi_{p,M}}}\right) &\le \psi_{p,M}(a) + \sum_{i=1}^{m}\int_a^\infty \frac{\mathbb{E}\psi_{p,M}\left(D\frac{|X_i|}{C}\right)}{\psi_{p,M}(Dx)}\psi'_{p,M}(x)\mathrm{d}x \\ &\le \psi_{p,M}(a) + m\int_a^\infty \frac{1}{\psi_{p,M}(Dx)}\psi'_{p,M}(x)\mathrm{d}x\end{aligned}$$

since all appearing terms are non negative, so all sums, integrals and expected values can be switched.
For the cases $p = 1, 2$, if $D > 1$ this means

$$\mathbb{E}\psi_{p,M}\left(\frac{\max_{1\le i\le m}|X_i|}{D\max_{1\le i\le m}\|X_i\|_{\psi_{p,M}}}\right) \le \frac{\mathrm{e}^{(a^p)}}{M} + \frac{m}{D^p - 1}\mathrm{e}^{(a^p(1-D^p))}$$

To bound the Orlicz norm we have to choose D large enough for the right side to be less or equal to 1. Choosing $a = \log(3/2)$ and

$$D := \frac{(2+\log(m))^{1/p}}{\log(3/2)} > 1$$

leads to

$$\begin{aligned}\mathbb{E}\psi_{p,M}\left(\frac{\max_{1\le i\le m}|X_i|}{D\max_{1\le i\le m}\|X_i\|_{\psi_{p,M}}}\right) &\le \frac{\exp(a^p)}{M} + \frac{\exp(a^p - 2)}{\frac{2}{a^p} + \frac{\log(m)}{a^p} - 1} \\ &\le \exp(a^p)\left(\frac{1}{2} + \frac{1}{\mathrm{e}^2}\right) \\ &\le \frac{3}{2}\left(\frac{1}{2} + \frac{1}{\mathrm{e}^2}\right) < 1\end{aligned}$$

if $M \ge 2$.
In conclusion

$$\left\|\max_{1\le i\le m} X_i\right\|_{\psi_{p,M}} \le \frac{(2+\log(m))^{1/p}}{\log(3/2)} \max_{1\le i\le m}\|X_i\|_{\psi_{p,M}}.$$

□

At last we need

Theorem 6.6 (Bernsteins inequality)
Let $X_1, X_2, ..., X_s$ be independent real random variables with $|X_i| \leq M$ for all $1 \leq i \leq s$ and define $\sigma^2 := \sum_{i=1}^{s} \mathbb{E}\left(X_i^2\right)$ and $c = \frac{M}{3}$, then

$$\mathbb{P}\left(\sum_{i=1}^{s} (X_i - \mathbb{E}X_i) \geq x\right) \leq \mathrm{e}^{-\frac{1}{2}\frac{x^2}{\sigma^2 + cx}}$$

for all $x > 0$.

For a proof see for example [BLM13, Corollary 2.11].

With these prerequisites we can finally prove Lemma 6.4, which gives us bounds on L^1- and Orlicz norms of the maximum of an empirical process indexed by a finite set of functions. We will use this lemma later to deal with the case of an infinite set of functions.

Proof of Lemma 6.4. As given in the statement, we want to prove

$$\left\|\max_{f\in\mathcal{F}} |\mathbb{G}_s(f)|\right\|_{\psi_{1,M}} \leq K'_{\psi,M}\left(\max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{s}}(2 + \log(|\mathcal{F}|)) + \max_{f\in\mathcal{F}} \|f\|_{L^2(P)} \sqrt{2 + \log(|\mathcal{F}|)}\right)$$

with constants $K'_{\psi,M} = \frac{4(M+1)}{(M-1)\log(3/2)}$ for a finite set $\mathcal{F}$ of bounded functions.
For each $f \in \mathcal{F}$ the term $\mathbb{G}_s(f) := \frac{1}{\sqrt{s}}\left(\sum_{i=1}^{s} (f(X_i) - \mathbb{E}f(X_i))\right)$ is a sum over independent random variables with mean zero and

$$\mathbb{E}\left(\frac{1}{\sqrt{s}}(f(X_i) - \mathbb{E}f(X_i))\right)^2 = \frac{1}{s}\left(\mathbb{E}(f(X_i)^2) - (\mathbb{E}f(X_i))^2\right) \leq \frac{\mathbb{E}(f(X_1)^2)}{s}.$$

Furthermore, since expected values of random variables are less or equal to their maximums, we have

$$\frac{1}{\sqrt{s}}(f(X_i) - \mathbb{E}f(X_i)) \leq \frac{2\|f\|_\infty}{\sqrt{s}}.$$

Bernsteins inequality for such sums then results for $x > 0$ in

$$\mathbb{P}(|\mathbb{G}_s(f)| > x) \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{\mathbb{E}(f(X_1)^2) + \frac{2}{3}\frac{\|f\|_\infty}{\sqrt{s}}x}\right).$$

The factor 2 appears here because we want to calculate

$$\mathbb{P}(|\mathbb{G}_s(f)| > x) \leq \mathbb{P}(\mathbb{G}_s(f) > x) + \mathbb{P}(\mathbb{G}_s(f) < -x) = \mathbb{P}(\mathbb{G}_s(f) > x) + \mathbb{P}(-\mathbb{G}_s(f) > x).$$

As $-\mathbb{G}_s(f)$ also fulfils all prerequisites of Bernsteins inequality, we can bound $\mathbb{P}(-\mathbb{G}_s(f) > x)$ with the same term as $\mathbb{P}(\mathbb{G}_s(f) > x)$ and we get the factor 2. Now define

$$a := \max_{f\in\mathcal{F}} \frac{2}{3}\frac{\|f\|_\infty}{\sqrt{s}} \le \max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{s}},\ b := \max_{f\in\mathcal{F}} \mathbb{E}(f(X_1))^2 = \max_{f\in\mathcal{F}} \|f\|^2_{L^2(P)}$$

so that

$$\mathbb{P}(|\mathbb{G}_s(f)| > x) \le 2\mathrm{e}^{-\frac{x^2}{4b}} \quad \text{if } x \le \frac{b}{a} \tag{6.4}$$

$$\mathbb{P}(|\mathbb{G}_s(f)| > x) \le 2\mathrm{e}^{-\frac{x}{4a}} \quad \text{else.} \tag{6.5}$$

This results in

$$\mathbb{P}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| \le \frac{b}{a}\right\} > x\right) \le 2\mathrm{e}^{-\frac{x^2}{4b}} \tag{6.6}$$

$$\mathbb{P}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| > \frac{b}{a}\right\} > x\right) \le 2\mathrm{e}^{-\frac{x}{4a}}. \tag{6.7}$$

First we check (6.6): If $x > \frac{b}{a}$ then the probability is zero. If $x \le \frac{b}{a}$ then

$$\mathbb{P}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| \le \frac{b}{a}\right\} > x\right) = \mathbb{P}\left(\frac{b}{a} \ge |\mathbb{G}_s(f)| > x\right) \le \mathbb{P}(|\mathbb{G}_s(f)| > x) \le 2\mathrm{e}^{-\frac{x^2}{4b}}$$

as a result of (6.4).
Now for (6.7): If $x \le \frac{b}{a}$ then using (6.4) we get

$$\mathbb{P}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| > \frac{b}{a}\right\} > x\right) = \mathbb{P}\left(|\mathbb{G}_s(f)| > \frac{b}{a}\right) \le 2\mathrm{e}^{-\frac{\left(\frac{b}{a}\right)^2}{4b}} = 2\mathrm{e}^{-\frac{\frac{b}{a}}{4a}} \le 2\mathrm{e}^{-\frac{x}{4a}}.$$

If $x > \frac{b}{a}$ then

$$\mathbb{P}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| > \frac{b}{a}\right\} > x\right) = \mathbb{P}(|\mathbb{G}_s(f)| > x)$$

and (6.5) immediately results in the same bound.
Applying property (O2) in Lemma 6.2 to (6.6) and (6.7) leads to

$$\left\||\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| \le \frac{b}{a}\right\}\right\|_{\psi_{2,M}} \le \sqrt{\frac{4(M+1)}{M-1}b}$$

$$\left\||\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)| > \frac{b}{a}\right\}\right\|_{\psi_{1,M}} \le \frac{4(M+1)}{M-1}a.$$

Using the triangle inequality and the properties of Lemma 6.2 gives

$$\begin{aligned}\left\|\max_{f\in\mathcal{F}}|\mathbb{G}_s(f)|\right\|_{\psi_{1,M}} &\leq \left\|\max_{f\in\mathcal{F}}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|\leq\frac{b}{a}\right\}\right)\right\|_{\psi_{1,M}}\\ &\quad+\left\|\max_{f\in\mathcal{F}}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|>\frac{b}{a}\right\}\right)\right\|_{\psi_{1,M}}\\ &\leq\frac{1}{\sqrt{\log(M)}}\left\|\max_{f\in\mathcal{F}}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|\leq\frac{b}{a}\right\}\right)\right\|_{\psi_{2,M}}\\ &\quad+\left\|\max_{f\in\mathcal{F}}\left(|\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|>\frac{b}{a}\right\}\right)\right\|_{\psi_{1,M}}.\end{aligned}$$

The inequalities above and Lemma 6.5 lead to

$$\begin{aligned}\left\|\max_{f\in\mathcal{F}}|\mathbb{G}_s(f)|\right\|_{\psi_{1,M}} &\leq\frac{\sqrt{2+\log(|\mathcal{F}|)}}{\log(3/2)\sqrt{\log(M)}}\max_{f\in\mathcal{F}}\left\||\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|\leq\frac{b}{a}\right\}\right\|_{\psi_{2,M}}\\ &\quad+\frac{(2+\log(|\mathcal{F}|))}{\log(3/2)}\max_{f\in\mathcal{F}}\left\||\mathbb{G}_s(f)|\,\mathbb{1}\left\{|\mathbb{G}_s(f)|>\frac{b}{a}\right\}\right\|_{\psi_{1,M}}\\ &\leq\frac{\sqrt{2+\log(|\mathcal{F}|)}}{\log(3/2)\sqrt{\log(M)}}\sqrt{\frac{4(M+1)}{M-1}b}+\frac{(2+\log(|\mathcal{F}|))}{\log(3/2)}\frac{4(M+1)}{M-1}a\\ &\leq\frac{4(M+1)}{\log(3/2)(M-1)}\left(\sqrt{b}\sqrt{2+\log(|\mathcal{F}|)}+a(2+\log(|\mathcal{F}|))\right)\\ &\leq\frac{4(M+1)}{\log(3/2)(M-1)}\left(\left(\max_{f\in\mathcal{F}}\|f\|_{L^2}\right)\sqrt{2+\log(|\mathcal{F}|)}+\left(\max_{f\in\mathcal{F}}\frac{\|f\|_\infty}{\sqrt{s}}\right)(2+\log(|\mathcal{F}|))\right).\end{aligned}$$

□

Now all preliminaries for the proof of our main theorem are done.

6.3 Suprema over a countable set of functions

In the previous section we already derived a bound for (some) Orlicz norms of suprema over a finite set of an empirical process in Lemma 6.4. Now we want to extend this result to suprema over countable sets. The main idea for the proof can be illustrated in a toy model.

The idea here is to pick for any $q\in\mathbb{N}$ a collection of functions $(g_i^q)_{1\leq i\leq N(q)}$ that is getting progressively denser and ensures that for every $f\in\mathcal{F}$ there are $1\leq i(f,q)\leq N(q)$ such that $f=\sum_{q=2}^{\infty}\left(g^q_{i(f,q)}-g^{q-1}_{i(f,q-1)}\right)$ and such that $\left\|g^q_{i(f,q)}-g^{q-1}_{i(f,q-1)}\right\|_{L^2(P)}<2^{-q}$ is always true, meaning that the functions $(g_i^q)_{1\leq i\leq N(q)}$

at "level" q can be used to approximate any $f \in \mathcal{F}$ up to an $L^2(P)$-error of 2^{-q+1}.
If we had an expression like

$$\left\| \max_{f \in \mathcal{F}} |\mathbb{G}_s(f)| \right\|_{\Psi_1} \leq \left(\max_{f \in \mathcal{F}} \|f\|_{L^2} \right) \sqrt{2 + \log(|\mathcal{F}|)}$$

instead of Lemma 6.4 for finite $\mathcal{F}$ and were able to pick g_i^q as above then we can try to extend this result to countable sets $\mathcal{F}$ by

$$\begin{aligned} \sup_{f \in \mathcal{F}} |\mathbb{G}_s(f)| &= \sup_{f \in \mathcal{F}} \left| \mathbb{G}_s \left(\sum_{q=2}^{\infty} g_{i(f,q)}^q - g_{i(f,q-1)}^{q-1} \right) \right| \\ &\leq \sum_{q=2}^{\infty} \sup_{f \in \mathcal{F}} \left| \mathbb{G}_s \left(g_{i(f,q)}^q - g_{i(f,q-1)}^{q-1} \right) \right|. \end{aligned}$$

Since we only have a finite amount of choices for each $q \in \mathbb{N}$, we might be able to replace the supremum over $f \in \mathcal{F}$ for each term of the sum with a supremum over the functions $(g_i^q)_{1 \leq i \leq N(q)}$. This is a supremum over a finite sum, and since we assumed that $\left\| g_{i(f,q)}^q - g_{i(f,q-1)}^{q-1} \right\|_{L^2(P)} < 2^{-q}$ is always true, our assumed toy model above would lead to

$$\left\| \max_{f \in \mathcal{F}} |\mathbb{G}_s(f)| \right\|_{\Psi_1} \leq \sum_{q=2}^{\infty} 2^{-q} \sqrt{2 + \log(|\mathcal{F}_q|)}$$

where $\mathcal{F}_q$ is the number of possible combinations $\left(g_{i(f,q)}^q - g_{i(f,q-1)}^{q-1} \right)$. If this sum converges we would get the bound that we want.
Sadly Lemma 6.4 is more complicated than this toy model, but the general idea of replacing every function by a sum of small "links" with a choice from a finite set is the same, we just need to do some extra steps. The way we will measure how many "links" we need for a given set of functions $\mathcal{F}$ will be its bracketing number.

Definition 6.7 (Bracketing Cover)
Let Y be a measurable space and let P be a measure on Y. For two functions $l, u \colon Y \to \mathbb{R}$ with $l(y) \leq u(y)$ for all $y \in Y$ the ***bracket*** *$[l, u]$ is the set*

$$[l, u] = \{ f \colon Y \to \mathbb{R} \mid l(y) \leq f(y) \leq u(y) \text{ for all } y \in Y \}.$$

Here, l is called the ***lower boundary function*** *and u the* ***upper boundary function*** *of the bracket.*
Let $\mathcal{F}$ be a set of functions $f \colon Y \to \mathbb{R}$. Then $\mathcal{I} = \{[l_i, u_i], 1 \leq i \leq N\}$ is a ***monotone bracketing cover*** *of $\mathcal{F}$ if*

(BC1) For all $y \in Y$ and $1 \leq i \leq N$ we have

$$u_i(y) \geq l_i(y)$$

and for all $y \in Y$ and $1 \leq i \leq N-1$ we have

$$l_{i+1}(y) \geq u_i(y).$$

(BC2) $\mathcal{F} \subset \cup_{i=1}^{N}[l_i, u_i]$.

*A sequence $(\mathcal{I}_q)_{q\in\mathbb{N}}$ with $\mathcal{I}_q = \left\{[l_{q,i}, u_{q,i}], 1 \leq i \leq N_{[]}(q, \mathcal{F}, P)\right\}$ of monotone bracketing covers with sequences of boundary functions $(l_{q,i})_{1\leq i\leq N_{[]}(q,\mathcal{F},P)}$ and $(u_{q,i})_{1\leq i\leq N_{[]}(q,\mathcal{F},P)}$ in $L^2(P)$ is a **nested monotone bracketing cover** if*

(NBC1)

$$\|u_{q,i} - l_{q,i}\|_{L^2(P)} \leq 2^{-q}$$

for all $q \in \mathbb{N}$, $1 \leq i \leq N_{[]}(q, \mathcal{F}, P)$,

(NBC2) For all $q, t \in \mathbb{N}$ and $1 \leq i \leq N_{[]}(q, \mathcal{F}, P)$ we have

$$l_{q,i} \in (l_{q+t,j})_{1\leq j\leq N_{[]}(q+t,\mathcal{F},P)}$$
$$u_{q,i} \in (u_{q+t,j})_{1\leq j\leq N_{[]}(q+t,\mathcal{F},P)}.$$

*We call the function $q \to N_{[]}(q, \mathcal{F}, P)$ the **monotone bracketing function** of the nested monotone bracketing cover.*

Remark 6.8.

- If P is the image measure obtained from a random variable X, then condition (NBC1) means that the boundary functions for each bracket need to fulfil

$$\mathbb{E}((u_{q,i}(X) - l_{q,i}(X))^2) \leq 2^{-2q}.$$

- The brackets of a monotone bracketing cover are nearly disjoint because of the monotonicity condition (BC1), only the boundary functions may be part of multiple brackets. We can form disjoint covers by recursively defining

$$\tilde{\mathcal{I}} = \left\{ [l_i, u_i] \setminus \bigcup_{j=1}^{i-1} [l_j, u_j], 1 \leq i \leq N \right\}.$$

  We can do the same "disjointification" for every monotone bracketing cover of a nested monotone bracketing cover as well. Then the resulting covers

$$\tilde{\mathcal{I}}_{q,i} = [l_{q,i}, u_{q,i}] \setminus \bigcup_{j=1}^{i-1} [l_{q,j}, u_{q,j}]$$

are nested as well, meaning that for each non-empty $\tilde{\mathcal{I}}_{q,i}$ there is a unique k with $1 \le k \le N_{q-1}$ such that $\tilde{\mathcal{I}}_{q,i} \subset \tilde{\mathcal{I}}_{q-1,k}$.
We will prove this by contradiction and first assume that there are $f, g \in \tilde{\mathcal{I}}_{q,i}$, $f \ne g$ and $f \in \tilde{\mathcal{I}}_{q-1,k_1}$, $g \in \tilde{\mathcal{I}}_{q-1,k_2}$ with $k_1 \ne k_2$. From this assumption follows that $f, g \in [l_{q,i}, u_{q,i}]$ and $f \in [l_{q-1,k_1}, u_{q-1,k_1}]$, $g \in [l_{q-1,k_2}, u_{q-1,k_2}]$. Without loss of generality we assume $k_1 < k_2$. From property (BC1) of Definition 6.7 follows

$$l_{q,i}(y) \le f(y) \le u_{q-1,k_1}(y) \le l_{q-1,k_2}(y) \le g(y) \le u_{q,i}(y)$$

for all $y \in Y$. Property (NBC2) has the consequence that there has to be a k_1' such that $u_{q-1,k_1} = u_{q,k_1'}$. Then the inequality $l_{q,i}(y) \le u_{q,k_1'}(y)$ implies $k_1' \ge i$ but $u_{q,k_1'}(y) \le u_{q,i}(y)$ implies $k_1' \le i$, resulting in $k_1' = i$. This in turn means that $g = u_{q,i} = l_{q-1,k_2}$, but since $u_{q,i} = u_{q-1,k_1}$ is an element of $[l_{q-1,k_1}, u_{q-1,k_1}]$, it cannot be an element of $\tilde{\mathcal{I}}_{q-1,k_2}$ because of the procedure for creating a disjoint cover. Thus, we arrive at a contradiction and the created cover has to be nested.

The connection between Orlicz norms of suprema over countable sets of functions of empirical processes and a nested monotone bracketing cover of that countable set is given by the following theorem, based on Theorems 2.5.6 and 2.14.2 of [vW96] and adapted for our purposes with explicit constants.

Theorem 6.9 (Orlicz norm of the maximum over the empirical process indexed by a countable set of functions)
Let Y be a measurable space, $\mathcal{F}$ a countable set of measurable functions $f : Y \to [0,1]$ with a nested monotone bracketing cover, and let $N_{[]}(q, \mathcal{F}, P)$ be the monotone bracketing function of the nested monotone bracketing cover. Let $X_1, X_2, \ldots$ be i.i.d. random variables taking values in Y with image measure P (on Y).
Then for all $M \ge 2$ there is a constant $K''_{\psi,M} = (14\ K'_{\psi,M} + \frac{4}{\log(M)})$ such that

$$\left\| \sup_{f\in\mathcal{F}} \left| \frac{1}{\sqrt{s}} \left(\sum_{i=1}^{s} (f(X_i) - \mathbb{E} f(X_i)) \right) \right| \right\|_{\psi_{1,M}} \le K''_{\psi,M} \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log\left(N_{[]}(q, \mathcal{F}, P)\right)} \tag{6.8}$$

where $\|\cdot\|_{\psi_{1,M}}$ is the Orlicz norm associated to $\psi_{1,M} := \frac{1}{M} \mathrm{e}^x$, $K'_{\psi,M} = \frac{4(M+1)}{\log(3/2)(M-1)}$.

Instead of a nested monotone bracketing cover it is enough to just assume that for all $\varepsilon > 0$ the set $\mathcal{F}$ can be covered by a number $N'_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{L^2(P)})$ of brackets of $L^2(P)$-size less then ε in exchange for worse constants, see Remark 6.10.

Proof of Theorem 6.9. The main idea of the proof is splitting $\mathcal{F}$ into nested subsets (based on the brackets given by the nested monotone bracketing cover), picking one representative from each of these subsets and applying Lemma 6.4 for these finite sets of functions. Then we continue by choosing ever smaller subsets and making sure that both parts of the sum that result from Lemma 6.4 are of comparable magnitude.
The whole set of functions $\mathcal{F}$ is contained in the bracket $[0,1]$ with constant functions 0 and 1. For every $q \in \mathbb{N}_0$, there is a covering $\mathcal{F} \subset \cup_{i=1}^{N_{[]}(q,\mathcal{F},P)} [l_{q,i}, u_{q,i}]$ by brackets from the nested monotone bracketing cover. By Remark 6.8 we can form nested disjoint covers of $\mathcal{F}$ via

$$\mathcal{F}_{q,i} := \mathcal{F} \cap \left([l_{q,i}, u_{q,i}] \setminus \bigcup_{j=1}^{i-1} [l_{q,j}, u_{q,j}] \right).$$

Since $\mathcal{F}$ and P are unchanged in the rest of this proof, we will use the shorthand

$$N_q := N_{[]}(q, \mathcal{F}, P)$$

for the upper bound on the number of sets $\mathcal{F}_{q,i}$ (since some of these may be empty). We choose $\mathcal{F}_{q_0} := [0,1]$ with $q_0 = -1$.
For a bracket $[l,u]$ we have $\sup\limits_{f,g\in[l,u]} |f-g| = |u-l|$ pointwise, resulting in

$$\left\| \sup_{f,g\in[l,u]} |f-g| \right\|_{L^2(P)} \leq \|u-l\|_{L^2(P)} \leq 2^{-q}$$

if $[l,u]$ is a q-bracket. This means that the partition of $\mathcal{F}$ also fulfils

$$\left\| \sup_{f,g\in\mathcal{F}_{q,i}} |f-g| \right\|_{L^2(P)} \leq 2^{-q} \; \forall i \in \{1,...,N_q\}.$$

Fix an element $f_{q,i} \in \mathcal{F}_{q,i}$ for each non-empty $\mathcal{F}_{q,i}$ and define pointwise for each $\mathcal{F}_{q,i}$

$$\begin{aligned} \pi_q f &:= f_{q,i} \text{ if } f \in \mathcal{F}_{q,i} \\ \Delta_q f &:= \sup_{h,g\in\mathcal{F}_{q,i}} |h-g| \leq |u_{q,i} - l_{q,i}| \text{ if } f \in \mathcal{F}_{q,i}. \end{aligned}$$

This way $\pi_q f$ is always a function in the covering set at “level” q that contains f, and $\Delta_q f$ is bound by the difference between the upper and lower boundary functions of the bracket at “level” q that contains $\mathcal{F}_{q,i}$. Since $\mathcal{F}_{q,i}$ is at most a countable union of functions, there are also at most countable many pairs $h, g \in \mathcal{F}_{q,i}$. Since absolute values, differences and countable suprema of measurable functions are themself measurable, $\Delta_q f$ is also measurable. By definition $\pi_q f = \pi_q g$ and $\Delta_q f = \Delta_q g$ if there is an $i \in \{1,...,N_q\}$ such that $f, g \in \mathcal{F}_{q,i}$.

Furthermore, define for all $q \in \mathbb{N}_0$

$$
\begin{aligned}
a_q &:= 2^{-q}/\sqrt{2+\log(N_{q+1})}\\
a_{q_0} &:= \sqrt{2}\\
A_{q-1}f : Y \to \{0,1\},\ A_{q-1}f(y) &:= \mathbb{1}\{\Delta_k f(y) \le \sqrt{s}a_k\ \forall k \in \{q_0,...,q-1\}\}\\
B_q f : Y \to \{0,1\},\ B_q f(y) &:= \mathbb{1}\{\Delta_k f(y) \le \sqrt{s}a_k\ \forall k \in \{q_0,...,q-1\}, \Delta_q f(y) > \sqrt{s}a_q\}\\
B_{q_0} f : Y \to \{0,1\},\ B_{q_0} f(y) &:= \mathbb{1}\{\Delta_{q_0} f(y) > \sqrt{s}a_{q_0}\}
\end{aligned}
$$

All $A_{q-1}f$ and B_q are indicator functions of intersections of level sets of measurable functions and thus measurable. Note that $\Delta_{q_0} f = \sup_{h,g\in\mathcal{F}} |h-g| \le 1$ and $a_{q_0} = \sqrt{2}$ results in $B_{q_0} f \le \mathbb{1}\{1 > \sqrt{2}\sqrt{s}\} = 0$, thus $B_{q_0} f \equiv 0$.
Now decompose f pointwise:

$$
\begin{aligned}
f - \pi_{q_0} f &= (f - \pi_{q_0} f)B_{q_0} f + \sum_{q=q_0+1}^{\infty} (f - \pi_q f)B_q f + \sum_{q=q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f)A_{q-1} f\\
&= \sum_{q=q_0+1}^{\infty} (f - \pi_q f)B_q f + \sum_{q=q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f)A_{q-1} f. \qquad (6.9)
\end{aligned}
$$

To check this, notice that for each $y \in Y$ either all $B_q f(y)$ are zero, or there is a unique $q_1(y)$ such that $B_{q_1(y)} f(y) = 1$.
In the first case all $A_q f(y)$ are 1 and

$$\sum_{q=q_0+1}^{k} (\pi_q f(y) - \pi_{q-1} f(y))A_{q-1} f(y) = \pi_k f(y) - \pi_{q_0} f(y).$$

The function $\pi_k f$ is in the same $\mathcal{F}_{k,i}$ as f by definition, $|f - \pi_k f| \le \Delta_k f$ and if all $A_q f = 1$ then $\Delta_k f(y) \le \sqrt{s}a_k \searrow 0$, so $\sum_{q=q_0+1}^{k} (\pi_q f(y) - \pi_{q-1} f(y))A_{q-1} f(y)$ converges to $f(y) - \pi_{q_0} f(y)$.
In the second case there exists a unique $q_1 \in \mathbb{N}$ such that $A_q f(y) = 1$ for all $q < q_1(y)$ and $A_q f(y) = 0$ for all $q \ge q_1(y)$. As a result the second line of (6.9) is

$$(f(y) - \pi_{q_1(y)} f(y)) + \sum_{q=q_0+1}^{q_1(y)} (\pi_q f(y) - \pi_{q-1} f(y)) = f(y) - \pi_{q_0} f(y).$$

The term $(f - \pi_{q_0} f)B_{q_0} f$ would cover the case that all A_q are zero, but as shown our choice of a_q prevents this.
Both sums in the second line of (6.9) are measurable. The first sum only contains at most one term for every $y \in Y$ and is thus convergent and bounded. Limits of

measurable functions are measurable, so the first sum in the second line of (6.9) is measurable and by

$$f - \pi_{q_0} f - \sum_{q=q_0+1}^{\infty} (f - \pi_q f) B_q f = \sum_{q=q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f) A_{q-1} f.$$

the same is true for the second sum. By dominated convergence both sums are P-integrable.
Now we can apply the empirical process $\mathbb{G}_s$ as defined in (6.2) to the decomposition (6.9), take the supremum over $f \in \mathcal{F}$ and by the linearity of $\mathbb{G}_s$ apply the triangle inequality for the absolute value and the supremum, resulting in

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\mathbb{G}_s(f - \pi_{q_0} f)| \le \sup_{f\in\mathcal{F}} &\left| \mathbb{G}_s \left(\sum_{q=q_0+1}^{\infty} (f - \pi_q f) B_q f \right) \right| \\ &+ \sup_{f\in\mathcal{F}} \left| \mathbb{G}_s \left(\sum_{q=q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f) A_{q-1} f \right) \right|. \qquad (6.10)\end{aligned}$$

Now $\mathbb{G}_s$ involves a finite sum and an expected value, which can be switched with the sum over q for each $f \in \mathcal{F}$ by dominated convergence. This follows from the two cases discussed after (6.9), the dominating function of all partial sums is 2. As a result

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\mathbb{G}_s(f - \pi_{q_0} f)| \le \sup_{f\in\mathcal{F}} &\left| \sum_{q=q_0+1}^{\infty} \mathbb{G}_s \left((f - \pi_q f) B_q f \right) \right| \\ &+ \sup_{f\in\mathcal{F}} \left| \sum_{q=q_0+1}^{\infty} \mathbb{G}_s \left((\pi_q f - \pi_{q-1} f) A_{q-1} f \right) \right|. \qquad (6.11)\end{aligned}$$

To prove the theorem, we need to bound the $\psi_{1,M}$-Orlicz norm of this supremum. First we will bound the norm of the first term on the right. Since $\Delta_q f = \sup_{h,g\in\mathcal{F}_{q,i}} |h - g|$ if $f \in \mathcal{F}_{q,i}$, it follows that $|f - \pi_q f| \le \Delta_q f$. From this, property (O4) of Lemma 6.2, the monotonicity of the Orlicz norm with regard to random variables and the triangle inequality follows

$$\begin{aligned}
&\left\|\sup_{f\in\mathcal{F}}\left|\sum_{q=q_0+1}^{\infty}\mathbb{G}_s\left((f-\pi_q f)B_q f\right)\right|\right\|_{\psi_{1,M}}\\
&\leq\left\|\sum_{q=q_0+1}^{\infty}\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\Big[\left(f(X_i)-\pi_q f(X_i)\right)B_q f(X_i)\right.\right.\\
&\qquad\qquad\left.\left.-\,\mathbb{E}\left((f(X_i)-\pi_q f(X_i))B_q f(X_i)\right)\Big]\right|\right\|_{\psi_{1,M}}\\
&\leq\left\|\sum_{q=q_0+1}^{\infty}\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\left[\Delta_q f(X_i)B_q f(X_i)+\mathbb{E}\left(\Delta_q f(X_i)B_q f(X_i)\right)\right]\right|\right\|_{\psi_{1,M}}\\
&=\left\|\sum_{q=q_0+1}^{\infty}\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\Big[\Delta_q f(X_i)B_q f(X_i)-\mathbb{E}\left(\Delta_q f(X_i)B_q f(X_i)\right)\right.\right.\\
&\qquad\qquad\left.\left.+\,2\mathbb{E}\left(\Delta_q f(X_i)B_q f(X_i)\right)\Big]\right|\right\|_{\psi_{1,M}}\\
&\overset{X_i\text{ i.i.d.}}{\leq}\left\|\sum_{q=q_0+1}^{\infty}\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{s}}\sum_{i=1}^{s}\left[\Delta_q f(X_i)B_q f(X_i)-\mathbb{E}\left(\Delta_q f(X_i)B_q f(X_i)\right)\right]\right|\right.\\
&\qquad\qquad\left.+\sum_{q=q_0+1}^{\infty}\sup_{f\in\mathcal{F}}\left|2\sqrt{s}\mathbb{E}\left(\Delta_q f(X_1)B_q f(X_1)\right)\right|\right\|_{\psi_{1,M}}\\
&\leq\sum_{q=q_0+1}^{\infty}\left(\left\|\sup_{f\in\mathcal{F}}\left|\mathbb{G}_s\left(\Delta_q f B_q f\right)\right|\right\|_{\psi_{1,M}}+\frac{2}{\log(M)}\sqrt{s}\sup_{f\in\mathcal{F}}\left|\mathbb{E}\left(\Delta_q f(X_1)B_q f(X_1)\right)\right|\right).
\end{aligned}\tag{6.12}$$

Now $B_q f\leq\mathbb{1}\{\Delta_q f>\sqrt{s}a_q\}$, so we can use the fact that if Z is a positive random variable, then

$$\forall t>0:\ t\mathbb{E}\left(Z\mathbb{1}_{\{Z>t\}}\right)\leq\mathbb{E}\left(Z^2\right).$$

With the definition of $\mathcal{F}_{q,i}$ we obtain the bound

$$\begin{aligned}
0\leq\sqrt{s}a_q\mathbb{E}\left(\Delta_q f(X_1)B_q f(X_1)\right)&\leq\sqrt{s}a_q\mathbb{E}\left(\Delta_q f(X_1)\mathbb{1}\{\Delta_q f(X_1)>\sqrt{s}a_q\}\right)\\
&\leq\mathbb{E}\left((\Delta_q f(X_1))^2\right)\\
&=\|\Delta_q f\|_{L^2(P)}^2\leq 2^{-2q}
\end{aligned}\tag{6.13}$$

for all $f \in \mathcal{F}$, resulting in

$$\frac{2}{\log(M)}\sqrt{s}\sup_{f\in\mathcal{F}}\left|\mathbb{E}\left(\Delta_q f(X_1)B_q f(X_1)\right)\right| \leq \frac{2}{\log(M)\ a_q}2^{-2q}.$$

At each "level" q, the definitions of $\Delta_q f$ and $B_q f$ and the fact that the partitions are nested ensure that $\Delta_q f B_q f = \Delta_q g B_q g$ if both f and g are an element of the same $\mathcal{F}_{q,i}$. That is because in this case f and g are also part of the same $\mathcal{F}_{k,i}$ at each $k < q$, resulting in $B_q f = B_q g$. Therefore, the supremum over all $f \in \mathcal{F}$ in the first part of the right side of (6.12) can be replaced by a supremum over only N_q different functions $\Delta_q f B_q f$ and we can apply Lemma 6.4, provided we can get bounds uniform in $\mathcal{F}$ on the supremum and L^2 norms of these functions. To do this we use the nestedness of the partitions $\mathcal{F}_{q,i}$ again and gain

$$0 \leq \Delta_q f B_q f \leq \Delta_{q-1} f B_q f \leq \sqrt{s}a_{q-1}$$

pointwise by the definition of $B_q f$. This results in

$$\sup_{f\in\mathcal{F}}\|\Delta_q f B_q f\|_\infty \leq \sqrt{s}a_{q-1}$$

and combined with inequality (6.13) leads to

$$\mathbb{E}\left((\Delta_q f(X_i)B_q f(X_i))^2\right) \leq \sqrt{s}a_{q-1}\mathbb{E}(\Delta_q f(X_i)B_q f(X_i)) \leq \frac{a_{q-1}}{a_q}2^{-2q}$$

so that

$$\sup_{f\in\mathcal{F}}\|\Delta_q f B_q f\|_{L^2(P)} \leq \sqrt{\frac{a_{q-1}}{a_q}}2^{-q}.$$

Applying the last results and Lemma 6.4 to (6.12) leads to

$$\begin{aligned}
&\left\|\sup_{f\in\mathcal{F}}\left|\sum_{q=q_0+1}^{\infty}\mathbb{G}_s\left((f-\pi_q f)B_q f\right)\right|\right\|_{\psi_{1,M}} \\
&\leq \sum_{q=q_0+1}^{\infty}\left(\left\|\sup_{f\in\mathcal{F}}\left|\mathbb{G}_s\left(\Delta_q f B_q f\right)\right|\right\|_{\psi_{1,M}} + \frac{2}{\log(M)\ a_q}2^{-2q}\right) \\
&\leq \sum_{q=q_0+1}^{\infty}\left(K' a_{q-1}(2+\log(N_q)) + K'\sqrt{\frac{a_{q-1}}{a_q}}2^{-q}\sqrt{2+\log(N_q)} + \frac{2}{\log(M)\ a_q}2^{-2q}\right)
\end{aligned}$$

with the same $K' := K'_{\psi_{1,M}}$ as in Lemma 6.4.
As $a_q = 2^{-q}/\sqrt{2+\log(N_{q+1})}$ and $N_q \geq N_{q-1}$ it follows that $a_{q-1} > a_q$ and thus

$$\sqrt{\frac{a_{q-1}}{a_q}} \leq \frac{a_{q-1}}{a_q}.$$

If we use this, we arrive at

$$\begin{aligned}
&\left\|\sup_{f\in\mathcal{F}}\left|\sum_{q=q_0+1}^{\infty}\mathbb{G}_s\left((f-\pi_q f)B_q f\right)\right|\right\|_{\psi_{1,M}} \leq \sum_{q=q_0+1}^{\infty} K' 2^{-(q-1)}\sqrt{2+\log(N_q)} \\
&\quad + K'\frac{2^{-(q-1)}\sqrt{2+\log(N_{q+1})}}{2^{-q}\sqrt{2+\log(N_q)}}2^{-q}\sqrt{2+\log(N_q)} + \frac{2}{\log(M)}\cdot 2^{-q}\sqrt{2+\log(N_{q+1})} \\
&= \sum_{q=q_0+1}^{\infty}\left(2K'\cdot 2^{-q}\sqrt{2+\log(N_q)} + \left(2K' + \frac{2}{\log(M)}\right)2^{-q}\sqrt{2+\log(N_{q+1})}\right) \\
&= \sum_{q=q_0+1}^{\infty}\left(2K'\cdot 2^{-q}\sqrt{2+\log(N_q)}\right) + \sum_{q=q_0+2}^{\infty}\left(\left(2K' + \frac{2}{\log(M)}\right)2^{-(q-1)}\sqrt{2+\log(N_q)}\right) \\
&\leq \sum_{q=q_0+1}^{\infty}\left(2K'\cdot 2^{-q}\sqrt{2+\log(N_q)}\right) + \sum_{q=q_0+1}^{\infty}\left(2\left(2K' + \frac{2}{\log(M)}\right)2^{-q}\sqrt{2+\log(N_q)}\right) \\
&= \left(6\,K' + \frac{4}{\log(M)}\right)\sum_{q=q_0+1}^{\infty}2^{-q}\sqrt{2+\log(N_q)}. \qquad (6.14)
\end{aligned}$$

The last inequality is actually a slight unneeded worsening of the bound, but it is done here to get a unified expression later.
For the second sum on the right in (6.11) we can get a similar estimate. Note that there are at most N_q different functions $\pi_q f - \pi_{q-1} f$ and at most N_{q-1} functions $A_{q-1} f$ since the $\mathcal{F}_{q,i}$ are nested and two functions f, g in the same set $\mathcal{F}_{q-1,i}$ are also in the same $\mathcal{F}_{q-2,i}$ and so on, so $A_{q-1}f = A_{q-1}g$. Thus, $A_{q-1}f = A_{q-1}(\pi_q f)$ and also $\pi_{q-1}f = \pi_{q-1}(\pi_q f)$. As a consequence, we actually have

$$(\pi_q f - \pi_{q-1} f)A_{q-1} f = (\pi_q f - \pi_{q-1}(\pi_q f))A_{q-1}(\pi_q f),$$

so $(\pi_q f - \pi_{q-1} f)A_{q-1} f$ only runs over at most N_q functions if f runs over all functions in $\mathcal{F}$. Additionally

$$0 \leq \left|\pi_q f - \pi_{q-1} f\right| A_{q-1} f \leq \left(\Delta_{q-1} f\right)\left(A_{q-1} f\right) \leq \sqrt{s}a_{q-1}$$

by the definition of $A_{q-1}f$ and $\|\pi_q f - \pi_{q-1} f\|_{L^2(P)} \leq 2^{-(q-1)}$ since $\pi_q f$ is in the same bracket $\mathcal{F}_{q-1,i}$ as $\pi_{q-1}f$. Therefore, we can use Lemma 6.4, resulting in

$$\begin{aligned}
&\left\| \sup_{f\in\mathcal{F}} \left| \sum_{q=q_0+1}^{\infty} \mathbb{G}_s\left((\pi_q f - \pi_{q-1}f)A_{q-1}f\right) \right| \right\|_{\psi_{1,M}} \leq \sum_{q=q_0+1}^{\infty} \left\| \sup_{f\in\mathcal{F}} \left| \mathbb{G}_s\left((\pi_q f - \pi_{q-1}f)A_{q-1}f\right) \right| \right\|_{\psi_{1,M}} \\
&\leq \sum_{q=q_0+1}^{\infty} K' \left(a_{q-1}(2+\log(N_q)) + 2\cdot 2^{-q}\sqrt{2+\log(N_q)} \right) \\
&= K' \sum_{q=q_0+1}^{\infty} \left(2\cdot 2^{-q}\sqrt{2+\log(N_q)} + 2\cdot 2^{-q}\sqrt{2+\log(N_q)} \right) \\
&= 4K' \sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{2+\log(N_q)}.
\end{aligned} \tag{6.15}$$

Now combining (6.11) with (6.14) and (6.15) leads to

$$\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(f-\pi_{q_0}f)| \right\|_{\psi_{1,M}} \leq \left(10\ K' + \frac{4}{\log(M)} \right) \sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{2+\log(N_q)}.$$

We now just have to bound $\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(\pi_{q_0}f)| \right\|_{\psi_{1,M}}$. The supremum here is just taken over a single function and $0 \leq \pi_{q_0}f \leq 1$ as well as $\|\pi_{q_0}f\|_{L^2(P)} \leq 1$. Thus, application of Lemma 6.4 leads to

$$\begin{aligned}
\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(\pi_{q_0}f)| \right\|_{\psi_{1,M}} &\leq K' \left(\frac{1}{\sqrt{1}}(2+\log(1)) + \sqrt{2+\log(1)} \right) \\
&\leq 4K' \leq 4K' \sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{2+\log(N_q)}.
\end{aligned}$$

With the triangle inequality this results in

$$\begin{aligned}
\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(f)| \right\|_{\psi_{1,M}} &\leq \left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(f-\pi_{q_0}f)| \right\|_{\psi_{1,M}} + \left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_s(\pi_{q_0}f)| \right\|_{\psi_{1,M}} \\
&\leq \left(14\ K' + \frac{4}{\log(M)} \right) \sum_{q=q_0+1}^{\infty} 2^{-q}\sqrt{2+\log(N_q)}
\end{aligned}$$

as claimed. □

Remark 6.10. The bound can be improved if the function $f \equiv 0$ is an element of $\mathcal{F}$. In this case we can choose $\pi_{q_0} f = 0$ and skip the last step of the proof, yielding the constant

$$\mathring{K}''_{\psi,M} = \left(10\ K'_{\psi,M} + \frac{4}{\log(M)}\right)$$

instead of $K''_{\psi,M}$.
Note that the requirement of a nested monotone bracket cover can be relaxed in exchange for worse constants. For an $\varepsilon > 0$ the **bracketing number** $N'_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{L^2(P)})$ is the minimal number of brackets of $L^2(P)$-size less than ε needed to cover $\mathcal{F}$. By first forming disjoint covers for $\varepsilon = 2^{-q}$ and then intersecting we can also get nested sequences with $\log(N_q) \leq \sum_{r=q_0+1}^{q} \log(N'_{[\,]}(2^{-r}, \mathcal{F}, \|\cdot\|_{L^2(P)})$. By using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b > 0$ and some switching of sums we would only end up with a different multiplicative constant but the same general form as (6.8).
The restriction to only a countable set of functions is harder to avoid, even if it does not enter the proof directly. Since the Orlicz norm is only defined for random variables (and therefore measurable functions) it is necessary to verify that every supremum that appears in the course of the proof is measurable. In the proof we only used suprema over a countable number of functions, thus we can be sure that we do not run into any measurability problems. For the application to admissible functions we have in mind we do need a supremum over an uncountable set of functions, but we will deal with this problem in another way in Section 6.5.

6.4 Concentration inequalities

As shown by Theorem 6.9 we get a bound for the Orlicz norm of the supremum over a countable set $\mathcal{F}$ of functions of an empirical process if

$$\sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log\left(N_{[]}(q, \mathcal{F}, P)\right)}$$

converges, where $q \mapsto N_{[]}(q, \mathcal{F}, P)$ is the monotone bracketing function of a nested monotone bracketing cover. In this section we will show that this is the case if $N_{[]}(q, \mathcal{F}, P) \leq V 2^{Wq}$ for some $V, W \geq 1$, and we will then derive two different concentration inequalities based on Theorem 6.9. The first just uses property (O1) of Lemma 6.2. Since we are interested in the speed of the exponential decay with rising s, we will name the concentration inequalities by their exponential dependence on s.

Corollary 6.11 (Sub-root-exponential concentration inequality)
In the setting of Theorem 6.9 with the additional restriction $N_{[]}(q, \mathcal{F}, P) \leq V 2^{Wq}$ *for some* $V, W \geq 1$ *we have*

$$\forall \kappa > 0: \; \mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \frac{1}{s} \sum_{i=1}^{s} f(X_i) - \mathbb{E} f(X_1) \right| \geq \kappa \right) \leq M \exp\left(-\frac{\sqrt{s}\kappa}{\tilde{K}_M(V,W)}\right)$$

where

$$\tilde{K}_M(V,W) = \left(14 \frac{4(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log(V) + Wq\log(2)} < \infty.$$

Remark 6.12. Since any probability is bounded by one, we get non-trivial results for the concentration if

$$M \exp\left(-\frac{\sqrt{s}\kappa}{\tilde{K}_M(V,W)}\right) \leq 1,$$

which is equivalent to

$$s \geq \left(\frac{\log(M) \tilde{K}_M(V,W)}{\kappa}\right)^2.$$

Proof. First we will show that if we have a monotone bracketing function that is bounded by $q \mapsto V 2^{Wq}$ then the sum in the statement of Theorem 6.9 converges. By monotonicity of the logarithm and the square root we get the bound

$$\sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log\left(N_{[]}(q, \mathcal{F}, P)\right)} \leq \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log(V) + Wq\log(2)}.$$

Since

$$\frac{2^{-(q+1)}\sqrt{2 + \log(V) + W(q+1)\log(2)}}{2^{-q}\sqrt{2 + \log(V) + Wq\log(2)}} = \frac{1}{2}\sqrt{\frac{2 + \log(V) + W(q+1)\log(2)}{2 + \log(V) + Wq\log(2)}} \xrightarrow[q \to \infty]{} \frac{1}{2}$$

the sum converges by the ratio test.
In combination with Theorem 6.9 this leads to

$$\begin{aligned} \left\| \sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{s}} \left(\sum_{i=1}^{s} \left(f(X_i) - \mathbb{E} f(X_i) \right) \right) \right| \right\|_{\psi_{1,M}} &\leq K''_{\psi,M} \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log(V) + Wq\log(2)} \\ &= \tilde{K}_M(V,W) < \infty. \end{aligned}$$

Now we use property (O1) of Lemma 6.2 to get

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{s}}\left(\sum_{i=1}^{s}(f(X_i)-\mathbb{E}f(X_i))\right)\right|\geq y\right)\leq\frac{1}{\psi_{1,M}\left(\frac{y}{\tilde{K}_M(V,W)}\right)}=M\mathrm{e}^{-\frac{y}{\tilde{K}_M(V,W)}}.$$

Using $y=\kappa\sqrt{s}$ then leads to

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{s}\left(\sum_{i=1}^{s}(f(X_i)-\mathbb{E}f(X_i))\right)\right|\geq \kappa\right)\leq M\mathrm{e}^{-\frac{\kappa\sqrt{s}}{\tilde{K}_M(V,W)}}.$$

□

Thus, we get a concentration inequality of the type we wanted. However, as Schumacher, Schwarzenberger and Veselić already showed in [SSV17] (Theorem 3.29 in Section 3.5 of this thesis) there should also be a concentration inequality that has an exponential term linear in s, since $\lfloor n/m\rfloor^d$ is exactly the number of samples we get by dividing a cube with side-length n into cubes with side-length m. We would like to replicate this convergence speed. To do this we use the special case $M=2$ to get a concentration inequality. But we need an additional theorem for this, namely the following due to Massart.

Theorem 6.13 ([Mas07, Equation (5.45)])
Let T be a countable set and let $Z_1,Z_2,...,Z_s$ be independent random vectors taking values in $\mathbb{R}^T$ with $\mathbb{E}(Z_{i,t})=0$ and $|Z_{i,t}|\leq 1$ for all $1\leq i\leq s$ and $t\in T$, where $Z_{i,t}$ is the t-coordinate of Z_i. Let $\sigma^2:=\sup_{t\in T}\sum_{i=1}^{s}\mathbb{E}((Z_{i,t})^2)$, $U:=\sup_{t\in T}\sum_{i=1}^{s}(Z_{i,t})^2$ and $Z:=\sup_{t\in T}|\sum_{i=1}^{s}Z_{i,t}|$, then for any $x>0$

$$\mathbb{P}\left(Z\geq\mathbb{E}Z+2\sqrt{(\sigma^2+\mathbb{E}U)x}+2x\right)\leq e^{-x}.$$

The result we arrive at is

Corollary 6.14 (Sub-exponential concentration inequality)
In the setting of Theorem 6.9 with the additional restriction that $N_{[]}(q,\mathcal{F},P)\leq V2^{Wq}$ for some $V,W\geq 1$ we have

$$\begin{aligned}\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{s}\sum_{i=1}^{s}f(X_i)-\mathbb{E}f(X_1)\right|\geq\kappa\right)&\leq\exp\left(-\frac{1}{2}\left(\sqrt{\kappa+1}-1\right)^2s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\right)\\&\leq\exp\left(-\frac{1}{12}\kappa^2s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\right).\end{aligned}\tag{6.16}$$

for all $0 < \kappa \leq 1$ *and* $s \geq \left(\frac{\tilde{K}_2(V,W)}{\kappa}\right)^2$ *where*

$$\tilde{K}_2(V,W) = \left(\frac{168}{\log(3/2)} + \frac{4}{\log(2)}\right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + \log(V) + Wq\log(2)}.$$

Furthermore we have

$$\forall\ 1 \geq \kappa > 0 : \ \mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \frac{1}{s} \sum_{i=1}^{s} f(X_i) - \mathbb{E}f(X_1) \right| \geq \kappa \right) \leq \exp\left(-\frac{\kappa^2}{24} s\right).$$

for $s \geq \left(12 \frac{\tilde{K}_2(V,W)}{\kappa^2}\right)^2$.

Note that the second line of 6.16 gives non-trivial results as soon as $s \geq \left(6 \frac{\tilde{K}_2(V,W)}{\kappa^2}\right)^2$ and the restriction to $\kappa \leq 1$ has no practical effect since $0 \leq f \leq 1$ for all $f \in \mathcal{F}$, meaning that the supremum is always bounded by one.

Proof. From property (O3) of Lemma 6.2 and Theorem 6.9 for the special case $M = 2$ we can deduce the bound

$$\mathbb{E}\left(\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{s} (f(X_i) - \mathbb{E}f(X_i)) \right| \right) \leq \tilde{K}_2(V,W)\sqrt{s} < \infty \tag{6.17}$$

in the same way and with $\tilde{K}_2(V,W)$ as given in as in Corollary 6.11.
We want to apply Theorem 6.13 with $T = \mathcal{F}$, $Z_i = (f(X_i) - \mathbb{E}(f(X_i)))_{f \in \mathcal{F}}$ and $Z_{i,f} = f(X_i) - \mathbb{E}f(X_i)$, since then

$$Z = \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{s} (f(X_i) - \mathbb{E}f(X_i)) \right|$$

is up to normalization the supremum we are interested in. We immediately see that $\mathbb{E}(Z_{i,t}) = 0$ and because of $0 \leq f \leq 1$ we also have $|Z_{i,t}| \leq 1$. Since $X_1, ..., X_s$ are independent and all $f \in \mathcal{F}$ are measurable, we also know that $Z_{1,f}, ..., Z_{s,f}$ are independent for each f. Therefore, the vectors $Z_1, ..., Z_s$ are independent. The prerequisites of the theorem are thus satisfied and we need to find bounds on σ^2 and $\mathbb{E}U$. We have

$$\begin{aligned}\sigma^2 &= \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} \mathrm{Var}(f(X_i)) = \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} \mathbb{E}(f(X_i)^2) - (\mathbb{E}(f(X_i)))^2 \\ &\leq \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} \mathbb{E}(f(X_i)^2) \leq \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} 1 = s\end{aligned}$$

and

$$U = \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} (f(X_i) - \mathbb{E}f(X_i))^2 \leq \sup_{f \in \mathcal{F}} \sum_{i=1}^{s} 1 = s,$$

because every f is pointwise between 0 and 1. This implies $\mathbb{E}U \leq s$, and from (6.17) we get $\mathbb{E}Z \leq \tilde{K}_2(V,W)\sqrt{s}$. Applying Theorem 6.13 leads to

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \frac{1}{s} \left| \sum_{i=1}^{s} (f(X_i) - \mathbb{E}f(X_i)) \right| \geq \overbrace{\frac{\tilde{K}_2(V,W)\sqrt{s} + 2\sqrt{2s}\sqrt{x} + 2x}{s}}^{=:\kappa(x,s)} \right) \leq e^{-x}.$$

Now we need to find $x(\kappa, s)$, the inverse function of $x \to \kappa(x,s)$. Let $y := \sqrt{x}$, then the equation we have to solve is

$$\begin{aligned}
&\tilde{K}_2(V,W)\sqrt{s} + 2\sqrt{2}\sqrt{s}y + 2y^2 = \kappa s \\
\Leftrightarrow & y^2 + \sqrt{2}\sqrt{s}y + \frac{\tilde{K}_2(V,W)}{2}\sqrt{s} = \frac{\kappa}{2}s \\
\Leftrightarrow & \left(y + \frac{1}{\sqrt{2}}\sqrt{s}\right)^2 - \frac{1}{2}s + \frac{\tilde{K}_2(V,W)}{2}\sqrt{s} = \frac{\kappa}{2}s \\
\Leftrightarrow & \left(y + \frac{1}{\sqrt{2}}\sqrt{s}\right)^2 = \underbrace{\frac{\kappa}{2}s + \frac{1}{2}s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s}}_{\text{has to be } \geq 0}
\end{aligned}$$

The equation can be solved if $s \geq \left(\frac{\tilde{K}_2(V,W)}{\kappa+1}\right)^2$ and the solution is

$$y = \sqrt{\frac{\kappa+1}{2}s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s}} - \frac{1}{\sqrt{2}}\sqrt{s}$$

This has to be positive, which is fulfilled if $s \geq \left(\frac{\tilde{K}_2(V,W)}{\kappa}\right)^2$, which also implies the previous condition.
Under this condition

$$\begin{aligned}
x = y^2 &= \frac{\kappa+2}{2}s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s} - \sqrt{2s}\sqrt{\underbrace{\frac{\kappa+1}{2}s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s}}_{\leq \frac{\kappa+1}{2}s}} \\
&\geq \left(\frac{\kappa+2}{2} - \sqrt{\kappa+1}\right)s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s} \\
&= \frac{1}{2}\left(\sqrt{\kappa+1} - 1\right)^2 s - \frac{\tilde{K}_2(V,W)}{2}\sqrt{s}
\end{aligned}$$

As a result

$$\begin{aligned}
&\mathbb{P}\left(\sup_{f\in\mathcal{F}}\frac{1}{s}\left|\sum_{i=1}^{s}(f(X_i)-\mathbb{E}f(X_i))\right|\geq\kappa\right)\\
&\leq\exp\left(-\left(\frac{\kappa+2}{2}s-\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}-\sqrt{2s}\sqrt{\frac{\kappa+1}{2}s-\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}}\right)\right)\\
&\leq\exp\left(-\frac{1}{2}\left(\sqrt{\kappa+1}-1\right)^2 s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\right).
\end{aligned}$$

Next we use that $\left(\sqrt{\kappa+1}-1\right)^2\geq\kappa^2/6$ for all $0\leq\kappa\leq 1$, which can be verified by using a substitution $\vartheta=\sqrt{\kappa+1}$. This gives us

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\frac{1}{s}\left|\sum_{i=1}^{s}(f(X_i)-\mathbb{E}f(X_i))\right|\geq\kappa\right)\leq\exp\left(-\frac{1}{12}\kappa^2 s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\right).$$

We get non-trivial results if

$$-\frac{1}{12}\kappa^2 s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\leq 0$$

which is equivalent to

$$s\geq\left(6\frac{\tilde{K}_2(V,W)}{\kappa^2}\right)^2.$$

Furthermore for

$$s\geq\left(12\frac{\tilde{K}_2(V,W)}{\kappa^2}\right)^2,$$

we get

$$-\frac{1}{12}\kappa^2 s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\leq-\frac{1}{24}\kappa^2 s$$

and therefore

$$\begin{aligned}
\mathbb{P}\left(\sup_{f\in\mathcal{F}}\frac{1}{s}\left|\sum_{i=1}^{s}(f(X_i)-\mathbb{E}f(X_i))\right|\geq\kappa\right)&\leq\exp\left(-\frac{1}{12}\kappa^2 s+\frac{\tilde{K}_2(V,W)}{2}\sqrt{s}\right)\\
&\leq\exp\left(-\frac{1}{24}\kappa^2 s\right)
\end{aligned}$$

which is the last statement of the Corollary. □

The advantage of Corollary 6.11 is the fact that it gives useful results at smaller s, namely for $s \geq \left(\frac{\log(M)\tilde{K}_M(V,W)}{\kappa}\right)^2$ as opposed to 6.14, where $s \geq \left(6\frac{\tilde{K}_2(V,W)}{\kappa^2}\right)^2$ is required. Depending on κ this can be substantially larger.
Corollary 6.14 on the other hand has a scaling of e^{-Cs}, as opposed to 6.11 which only has a scaling of $e^{-C'\sqrt{s}}$. Here, the advantage is stronger for large s. Since $C = \frac{1}{24}\kappa^2$ and $C' = \frac{\kappa}{\tilde{K}_M(V,W)}$ the constant of this scaling is however better in Corollary 6.11 for small κ. This way, both corollaries cover different uses.

Now we will use the theorem we just proved for monotone increasing, right-continuous bounded functions, such as our admissible functions.

6.5 Application to monotone increasing, right-continuous bounded random functions

As shown in the last section we get useful concentration inequalities if we can show for a given set of functions $\mathcal{F}$ that there is a nested monotone bracketing cover with a monotone bracketing function satisfying $N_{[]}(q, \mathcal{F}, P) \leq V 2^{Wq}$ for some $V, W \geq 1$. The next step is now to apply these results to get uniform convergence for a class of functions that contain the admissible functions from Definition 4.3. As mentioned in the beginning of this chapter, we can translate a uniform convergence of an averaged sum to a supremum over a set of functions via

$$\sup_{x\in\mathbb{R}} \left| \frac{1}{s} \sum_{t=1}^{s} \phi(X_t)(x) - \mathbb{E}\phi(X_1)(x) \right| = \sup_{f\in\mathcal{F}'} \left| \frac{1}{s} \left(\sum_{t=1}^{s} f(X_t) - \mathbb{E}f(X_t) \right) \right|$$

where $\mathcal{F}' = \{\phi(\cdot)(x) : x \in \mathbb{R}\}$ and $X_1, ..., X_s$ is a sequence of i.i.d. random variables. The set $\mathcal{F}'$ defined here is not countable and we do not know if there is a nested monotone bracketing cover. We need to choose a suitable countable subset $\mathcal{F}$ of $\mathcal{F}'$, find a nested monotone bracketing cover for this set with a monotone bracketing function that is bounded by $V 2^{Wq}$ for some $V, W \geq 1$, and then extend the concentration inequality to the whole set $\mathcal{F}'$.
We will show that this is possible for every function fulfilling a few specific properties. Recall that $\mathbb{B}$ is the space of bounded, right-continuous functions from $\mathbb{R}$ to $\mathbb{R}$.

Definition 6.15
Let Λ be a finite set and $(\Omega_\Lambda = \mathbb{R}^\Lambda, \mathcal{B}, \mathbb{P}_\Omega)$ be a probability space.
Then $\phi\colon \Omega_\Lambda \to \mathbb{B}$ is a ***monotone increasing, right-continuous bounded random function in*** $[0,1]$ *if*

- *$\phi(\cdot)(x)\colon \Omega_\Lambda \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable for all $x \in \mathbb{R}$,*
- *$\phi(\omega)(x) \leq \phi(\omega)(y)$ if $x < y$ for all $\omega \in \Omega_\Lambda$,*

- $0 \leq \phi(\omega)(x) \leq 1$ *for all* $\omega \in \Omega_\Lambda$ *and* $x \in \mathbb{R}$.

We further define for such functions

$$\Phi\colon \mathbb{R} \to \mathbb{R}, \qquad \Phi(x) := \mathbb{E}(\phi(\cdot)(x))$$

which is also monotone increasing and right continuous. The expected value is taken with regard to the measure $\mathbb{P}_\Omega$. Additionally we define pointwise the left-continuous version of ϕ, namely

$$\phi^-(\omega)(x) := \lim_{x' \nearrow x} \phi(\omega)(x')$$

and its associated

$$\Phi^- : \mathbb{R} \to \mathbb{R}, \qquad \Phi^-(x) := \mathbb{E}(\phi^-(\cdot)(x)).$$

We further define the limits

$$\begin{aligned} \phi(\omega)(-\infty) &= \lim_{x' \searrow -\infty} \phi(\omega)(x') \\ \phi^-(\omega)(\infty) &= \lim_{x' \nearrow \infty} \phi(\omega)(x') \\ \Phi(-\infty) &= \lim_{x' \searrow -\infty} \Phi(x') \\ \Phi^-(\infty) &= \lim_{x' \nearrow \infty} \Phi(x') \end{aligned}$$

which exist for all ω by the monotonicity and boundedness of ϕ.
Next, we consider for $\varepsilon > 0$ and $j = 0, 1, \ldots, k := \lceil \frac{1}{\varepsilon} \rceil$

$$x_j(\varepsilon) := \begin{cases} -\infty & \text{if } j = 0 \\ \Phi^{-1}(j \cdot \varepsilon) & \text{if } 0 < j < k \\ \infty & \text{if } j = k \end{cases}$$

where

$$\Phi^{-1}(\alpha) := \inf\{\lambda \in \mathbb{R} \mid \Phi(\lambda) \geq \alpha\}$$

is the **generalized inverse** of Φ. Here, we use $\inf \emptyset = \infty$. See also Figure 6.1.

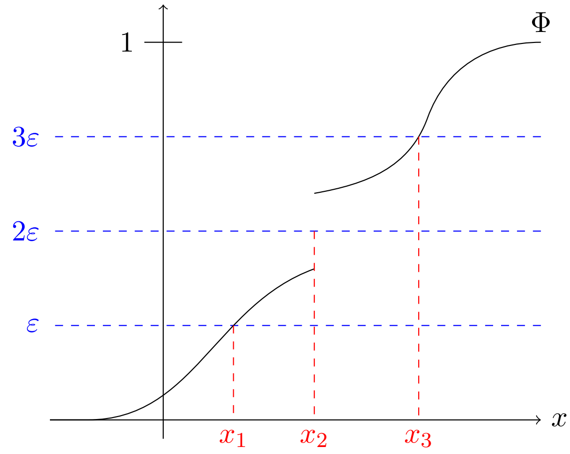

Figure 6.1: An illustration of the process of finding $x_j(\varepsilon)$

The set of functions $\mathcal{F}$ we will consider in the following is given by

$$\begin{aligned}\mathcal{F} :=& \{\phi(\cdot)(x) : \Omega_\Lambda \to \mathbb{R} \mid \exists Q \in \mathbb{Q} : x = \Phi^{-1}(Q)\} \\ & \cup \{\phi^-(\cdot)(x) : \Omega_\Lambda \to \mathbb{R} \mid \exists Q \in \mathbb{Q} : x = \Phi^{-1}(Q)\} \\ & \cup \{\phi(\cdot)(-\infty)\} \cup \{\phi^-(\cdot)(\infty)\}. \end{aligned} \tag{6.18}$$

Moreover, we define

$$\begin{aligned}\mathcal{F}_{\varepsilon,j} &:= [\phi(\cdot)(x_{j-1}(\varepsilon^2)), \phi^-(\cdot)(x_j(\varepsilon^2))] \\ &\quad (\supset \{\phi \in \mathcal{F} \mid \phi(\omega)(x_{j-1}(\varepsilon^2)) \le \phi(\omega) \le \phi^-(\omega)(x_j(\varepsilon^2)) \text{ for all } \omega \in \Omega_\Lambda\})\end{aligned}$$

for $j \in \{1, \dots, k = \lceil \frac{1}{\varepsilon^2} \rceil\}$ if $x_{j-1}(\varepsilon^2) < x_j(\varepsilon^2)$, where the brackets are defined as in Definition 6.7. Otherwise $\mathcal{F}_{\varepsilon,j} := \emptyset$.

Lemma 6.16
We have:

1. $\forall \varepsilon > 0 : \quad \mathcal{F} \subset \bigcup_{j=1}^k \mathcal{F}_{\varepsilon,j}$
2. $\forall j \in \{1, \dots, k\}$ *where* $x_{j-1}(\varepsilon^2) < x_j(\varepsilon^2)$:

$$\mathbb{E}\left(\left(\phi^-(\cdot)(x_j(\varepsilon^2)) - \phi(\cdot)(x_{j-1}(\varepsilon^2))\right)^2\right) \le \varepsilon^2$$

Proof. 1. Let $x \in \Phi^{-1}(\mathbb{Q})$ be arbitrary. Then, if $x = x_j(\varepsilon^2)$ for some $j \in \{1, \dots, k-1\}$ and $x \notin \{-\infty, \infty\}$, there are $l, m \in \{1, \dots, k-1\}$ with $x_l(\varepsilon^2) = x_m(\varepsilon^2) = x_j(\varepsilon^2)$ such that

$$\phi(\cdot)(x) = \phi(\cdot)(x_j(\varepsilon^2)) \in \mathcal{F}_{\varepsilon,l+1}$$

and

$$\phi^-(\cdot)(x) = \phi^-(\cdot)(x_j(\varepsilon^2)) \in \mathcal{F}_{\varepsilon,m}.$$

The l and m might not be equal to j since some brackets might be empty.
If $x = x_0 = -\infty$ then

$$\phi(\cdot)(x) \in \mathcal{F}_{\varepsilon,d}.$$

where $d = \min\{j \in \{1,\ldots,k\} : x_{j-1}(\varepsilon^2) < x_j(\varepsilon^2)\}$ Note that $\phi^-(\omega)(-\infty)$ is not defined. If $x = x_k = \infty$ then

$$\phi^-(\cdot)(x) \in \mathcal{F}_{\varepsilon,b}.$$

where $b = \max\{j \in \{1,\ldots,k\} : x_{j-1}(\varepsilon^2) < x_j(\varepsilon^2)\}$. As before, we do not have to check whether $\phi(\omega)(\infty) \in \mathcal{F}$.

Now let $x \in \Phi^{-1}(\mathbb{Q}) \setminus \{x_0(\varepsilon^2),\ldots,x_k(\varepsilon^2)\}$. As $x_{j-1}(\varepsilon^2) < x < x_j(\varepsilon^2)$ for some j, we have

$$\phi(\omega)(x_{j-1}(\varepsilon^2)) \leq \phi^-(\omega)(x) \leq \phi(\omega)(x) \leq \phi^-(\omega)(x_j(\varepsilon^2))\ \forall \omega \in \Omega_\Lambda$$

because of the monotonicity of ϕ. Thus,

$$\phi(\cdot)(x) \in \mathcal{F}_{\varepsilon,j} \quad \text{and} \quad \phi^-(\cdot)(x) \in \mathcal{F}_{\varepsilon,j}$$

2. Let $j \in \big\{t \in \{1,\ldots,k\} : x_{t-1}(\varepsilon^2) < x_t(\varepsilon^2)\big\}$ be arbitrary.

   Note that for each $\alpha \in [0,1]$ we have

$$\Phi^{-1}(\alpha) = \inf\{x \in \mathbb{R} \mid \Phi(x) \geq \alpha\} = \sup\{x \in \mathbb{R} \mid \Phi(x) < \alpha\}$$

   as well as

$$\begin{aligned} \forall x > \Phi^{-1}(\alpha) &: \quad \Phi(x) \geq \alpha \\ \forall x < \Phi^{-1}(\alpha) &: \quad \Phi^-(x) \leq \Phi(x) < \alpha \end{aligned}$$

   Because of this, the right-continuousness of Φ and the left-continuousness of Φ^- the inequalities

$$\Phi(\Phi^{-1}(\alpha)) \geq \alpha \quad \text{and} \quad \Phi^-(\Phi^{-1}(\alpha)) \leq \alpha \tag{6.19}$$

   hold. Note that

$$\begin{aligned} \Phi^{-1}(\alpha) = -\infty &\Rightarrow \Phi(\Phi^{-1}(\alpha)) \geq \alpha, \\ \Phi^{-1}(\alpha) = \infty &\Rightarrow \Phi^-(\Phi^{-1}(\alpha)) \leq \alpha \end{aligned}$$

holds as well, since $\Phi^{-1}(\alpha) = -\infty$ means $\Phi(x) \geq \alpha$ is true for all $x \in \mathbb{R}$ and this carries over to the limit. The second line follows the same way since $\Phi^{-1}(\alpha) = \infty$ means $\Phi(x) \leq \alpha$ is true for all $x \in \mathbb{R}$.
Applying this we end up with

$$\begin{aligned}\mathbb{E}\left(\left(\phi^{-}(\cdot)(x_j(\varepsilon^2)) - \phi(\cdot)(x_{j-1}(\varepsilon^2))\right)^2\right) &\leq \mathbb{E}\left(\phi^{-}(\cdot)(x_j(\varepsilon^2)) - \phi(\cdot)(x_{j-1}(\varepsilon^2))\right)\\ &= \Phi^{-}(x_j(\varepsilon^2)) - \Phi(x_{j-1}(\varepsilon^2))\\ &\leq j\varepsilon^2 - (j-1)\varepsilon^2 = \varepsilon^2\end{aligned}$$

where we used $0 \leq \phi(\omega)(x_{j-1}(\varepsilon^2)) \leq \phi^{-}(\omega)(x_j(\varepsilon^2)) \leq 1$ for all $\omega \in \Omega_\Lambda$ in the first line.

□

Corollary 6.17
The brackets

$$\mathcal{F}_{2^{-q},j} := \left[\phi(\cdot)\left(x_{j-1}\left(2^{-2q}\right)\right), \phi^{-}(\cdot)\left(x_j\left(2^{-2q}\right)\right)\right]$$

for $j \in \left\{t \in \{1, \dots, 2^{2q}\} : x_{t-1}(2^{-2q}) < x_t(2^{-2q})\right\}$ *form a nested monotone bracketing cover of* $\mathcal{F}$ *with regard to the measure* $\mathbb{P}_\Omega$ *and monotone bracketing function*

$$N_{[]}(q, \mathcal{F}, \mathbb{P}_\Omega) \leq 2^{2q}.$$

Proof. Lemma 6.16 shows that the brackets $\mathcal{F}_{\varepsilon,j}$ we defined in (6.18) fulfil requirements (BC2) and (NBC1) of Definition 6.7 for $\varepsilon = 2^{-q}$, since the brackets cover $\mathcal{F}$ and each bracket has an $L^2(\mathbb{P}_\Omega)$-size of less than ε. Condition (BC1) is automatically fulfilled by the monotonicity of ϕ and $x_{j-1}(\varepsilon) \leq x_j(\varepsilon)$. By the definition of $x_j(\varepsilon)$ we have $x_j((2^{-q})^2) = x_{4j}((2^{-(q+1)})^2)$ for all $q \in \mathbb{N}$ and $0 \leq j \leq 2^{2q}$, which ensures condition (NBC2). Every bracket for $\varepsilon = 2^{-q}$ contains exactly four brackets for $\varepsilon = 2^{-(q+1)}$.
We need at most $k = \lceil 1/\varepsilon^2 \rceil$ sets to cover $\mathcal{F}$ in the described way, so with $\varepsilon = 2^{-q}$ we have

$$N_{[]}(q, \mathcal{F}, P) \leq \lceil \frac{1}{2^{-2q}} \rceil = 2^{2q}.$$

□

Now we can use the two corollaries from the last section to gain results for the concentration of monotone increasing, right-continuous bounded random functions in $[0, 1]$.

Lemma 6.18
Let Λ be a finite set and $(\Omega_\Lambda = \mathbb{R}^\Lambda, \mathcal{B})$ be a measurable space. Let $\phi\colon \Omega_\Lambda \to \mathbb{B}$ (where $\mathbb{B}$ is the space of bounded, right-continuous functions from $\mathbb{R}$ to $\mathbb{R}$) be a monotone increasing, right-continuous bounded random function in $[0,1]$ as in Definition 6.15. Let $s \in \mathbb{N}$ and $X_1, ..., X_s$ be i.i.d. random variables defined on a probability space $(\Theta, \mathcal{D}, \mathbb{P})$ with values in Ω_Λ. For arbitrary $\kappa > 0$ and $2 \le M$ there is a set $A_{M,s,\kappa} \in \mathcal{D}$ such that

$$\forall \nu \in A_{M,s,\kappa}: \sup_{x\in\mathbb{R}} \left\| \frac{1}{s} \sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| \le \kappa$$

and

$$\mathbb{P}\left(A_{s,\kappa}\right) \ge 1 - M \exp\left(-\frac{\sqrt{s}\kappa}{K_M}\right) \tag{6.20}$$

where

$$K_M = \tilde{K}_M(1,2) = \left(\frac{56(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right) \sum_{q=0}^{\infty} 2^{-q}\sqrt{2 + 2q\log(2)}. \tag{6.21}$$

There is also for all $1 \ge \kappa > 0$ and $s \ge \left(12\frac{K_2}{\kappa^2}\right)^2$ (where K_2 is (6.21) for $M = 2$) a set $B_{s,\kappa} \in \mathcal{D}$ such that

$$\forall \nu \in B_{s,\kappa}: \sup_{x\in\mathbb{R}} \left\| \frac{1}{s} \sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| \le \kappa$$

and

$$\mathbb{P}\left(B_{s,\kappa}\right) \ge 1 - \exp\left(-\frac{1}{24}\kappa^2 s\right). \tag{6.22}$$

Note that by Remark 6.12 the bound on the probability in (6.20) gives non-trivial results if

$$s \ge \left(\frac{\log(M)K_M}{\kappa}\right)^2.$$

Proof. We combine Corollary 6.17 and the sub-root-exponential concentration inequality from Corollary 6.11. With $\mathbb{P}_{X_1}$, the image measure of X_1, we can form a probability space $(\Omega_\Lambda = \mathbb{R}^\Lambda, \mathcal{B}, \mathbb{P}_{X_1})$. Then Corollary 6.17 shows that the brackets

$$\mathcal{F}_{2^{-q},j} := \left[\phi(\cdot)\left(x_{j-1}\left(2^{-2q}\right)\right), \phi^{-}(\cdot)\left(x_j\left(2^{-2q}\right)\right)\right]$$

for $j \in \{t \in \{1,\dots,k\} : x_{t-1}(\varepsilon^2) < x_t(\varepsilon^2)\}$ form a nested monotone bracketing cover with

$$N_{[]}(q,\mathcal{F},P) \leq 2^{2q}.$$

Thus, we have $N_{[]}(q,\mathcal{F},P) \leq V2^{Wq}$ with $V = 1$ and $W = 2$. This leads to $K_M = \tilde{K}_M(1,2)$ with $2 \leq M$ and

$$\forall\kappa > 0: \ \mathbb{P}\left(\left\{\nu \in \Theta : \sup_{f\in\mathcal{F}} \left|\frac{1}{s}\sum_{i=1}^{s} f(X_i(\nu)) - \mathbb{E}f(X_1)\right| \geq \kappa\right\}\right) \leq M\mathrm{e}^{-\frac{\sqrt{s}\kappa}{K_M}}$$

from Corollary 6.11. For us, the important part of $\mathcal{F}$ are the functions indexed by $\Phi^{-1}(\mathbb{Q})$, and since this is a subset of $\mathcal{F}$ we can use the same bound for

$$\mathbb{P}\left(\underbrace{\left\{\nu \in \Theta : \sup_{x\in\Phi^{-1}(\mathbb{Q})} \left|\frac{1}{s}\sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x))\right| \geq \kappa\right\}}_{=:A^C_{M,s,\kappa}}\right) \leq M\mathrm{e}^{-\frac{\sqrt{s}\kappa}{K_M}}.$$

However, now we need to extend this result to all $x \in \mathbb{R}$. We will use the monotonicity of $\phi(\cdot)$ to do just that. If $x \in \mathbb{R} \setminus \Phi^{-1}(\mathbb{Q})$, then for all $J \in \mathbb{N}$ there is a $j \in \{1,...,J\}$: $x_{j-1}(1/J) < x < x_j(1/J)$
Because of this and (6.19)

$$\mathbb{E}\left(\phi(X_1)\left(x_j\left(1/J\right)\right)\right) = \Phi\left(x_j\left(1/J\right)\right) \geq \frac{j}{J},\ \mathbb{E}\left(\phi^-(X_1)\left(x_j\left(1/J\right)\right)\right) \leq \frac{j}{J},$$
$$x \in \left(x_{j-1}\left(1/J\right), x_j\left(1/J\right)\right) \Rightarrow \frac{j-1}{J} \leq \mathbb{E}\left(\phi(X_1)(x)\right) \leq \frac{j}{J}$$

As a result we have for all $\nu \in A_{M,s,\kappa}$

$$\begin{aligned}
\frac{1}{s}\sum_{i=1}^{s}\phi(X_i(\nu))(x) &\leq \frac{1}{s}\sum_{i=1}^{s}\phi^-(X_i(\nu))\left(x_j\left(1/J\right)\right) \leq \mathbb{E}\left(\phi^-(X_1)\left(x_j\left(1/J\right)\right)\right) + \kappa \\
&\leq \frac{j}{J} + \kappa \leq \mathbb{E}(\phi(X_1)(x)) + \kappa + \frac{1}{J} \\
\frac{1}{s}\sum_{i=1}^{s}\phi(X_i(\nu))(x) &\geq \frac{1}{s}\sum_{i=1}^{s}\phi(X_i(\nu))\left(x_{j-1}\left(1/J\right)\right) \geq \mathbb{E}\left(\phi(X_1)\left(x_{j-1}\left(1/J\right)\right)\right) - \kappa \\
&\geq \frac{j-1}{J} - \kappa \geq \mathbb{E}(\phi(X_1)(x)) - \kappa - \frac{1}{J}
\end{aligned}$$

and consequently

$$\forall\nu \in A_{M,s,\kappa}, J \in \mathbb{N}: \ \sup_{x\in\mathbb{R}} \left|\frac{1}{s}\sum_{i=1}^{s}\phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x))\right| \leq \kappa + \frac{1}{J}.$$

By taking $J \to \infty$ we confirm that

$$\forall \nu \in A_{M,s,\kappa} \ \sup_{x\in\mathbb{R}} \left| \frac{1}{s} \sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right| \leq \kappa$$

with $\mathbb{P}(A_{M,s,\kappa}) \geq 1 - M\mathrm{e}^{-\frac{\sqrt{s}\kappa}{K_M}}$.
We can also apply Corollary 6.14 instead of Corollary 6.11, which leads us to the second stated result. □

Remark 6.19. The restriction $0 \leq \phi \leq 1$ was chosen to fit Corollary 6.11 and our intended use case of the eigenvalue-counting functions. However, it is easily generalized to functions ϕ with $a \leq \phi \leq b$ by simply using the function $\tilde{\phi} = \frac{\phi - a}{b-a}$ leading to

$$\sup_{x\in\mathbb{R}} \left\| \frac{1}{s} \sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| \leq (b-a)\kappa \ \ \forall \nu \in A_{M,s,\kappa}$$

and $\mathbb{P}(A_{M,s,\kappa}) \geq 1 - M\mathrm{e}^{-\frac{\sqrt{s}\kappa}{K_M}}$ or equivalently

$$\sup_{x\in\mathbb{R}} \left\| \frac{1}{s} \sum_{i=1}^{s} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| \leq \kappa \ \ \forall \nu \in A_{M,s,\kappa}$$

and $\mathbb{P}(A_{M,s,\kappa}) \geq 1 - M\mathrm{e}^{-\frac{\sqrt{s}\kappa}{K_M(b-a)}}$.

Remark 6.20. If there is an $x \in \Phi^{-1}(\mathbb{Q})$ such that $\phi(\omega)(x) \equiv 0$ or $\phi(\omega)(-\infty) \equiv 0$, then by Remark 6.10 the constant K_M can be replaced by

$$\mathring{K}_M = \left(\frac{40(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)} \right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + 2q \log(2)}.$$

One additional result of the calculation is the fact that the uniform convergence is almost sure, which was already established in [SSV17] for admissible functions.

Corollary 6.21
In the setting of Lemma 6.18 there is a set Ω' of full measure such that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left\| \frac{1}{h(n)} \sum_{i=1}^{h(n)} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| = 0 \ \ \forall \nu \in \Omega'$$

for all functions $h : \mathbb{N} \to \mathbb{N}$ where there is a $c > 0$ and an $N \in \mathbb{N}$ such that $h(n) \geq cn$ for all $n > N$.

Proof. We use the Borel-Cantelli Lemma 2.9 as we know that $A^C_{M,h(n),\kappa}$ as defined in Lemma 6.18 contains all ν such that

$$\sup_{x\in\mathbb{R}} \left\| \frac{1}{h(n)} \sum_{i=1}^{h(n)} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| > \kappa \tag{6.23}$$

and consequently $\limsup_{n\to\infty} A^C_{M,h(n),\kappa}$ contains all ν such that (6.23) is true for infinitely many $n \in \mathbb{N}$. From the Borel-Cantelli Lemma follows that if

$$\sum_{n=1}^{\infty} \mathbb{P}(A^C_{M,h(n),\kappa}) \leq \sum_{n=1}^{\infty} M\mathrm{e}^{-\frac{\sqrt{h(n)}\kappa}{K_M}} < \infty \tag{6.24}$$

then $\mathbb{P}(\limsup_{n\to\infty} A^C_{M,h(n),\kappa}) = 0$. We have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\mathrm{e}^{-c\sqrt{x}}\right) = \left(2 - \frac{c}{2}\sqrt{x}\right) x\mathrm{e}^{-c\sqrt{x}}$$

for $c > 0$ and $x > 0$, and thus

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^2\mathrm{e}^{-c\sqrt{x}}\right) \geq 0 \Leftrightarrow 2 \geq \frac{c}{2}\sqrt{x} \Leftrightarrow \frac{16}{c^2} \geq x.$$

This results in

$$n^2\mathrm{e}^{-c\sqrt{n}} \leq \max_{1\leq m\leq 16/c^2}\left(m^2\mathrm{e}^{-c\sqrt{m}}\right) =: d < \infty \ \forall n \in \mathbb{N}$$

and

$$\mathrm{e}^{-c\sqrt{n}} \leq \frac{d}{n^2}.$$

The series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, which means that (6.24) is true if there is some point N such that $h(n) \geq cn$ for all $n > N$. Then $\mathbb{P}(\limsup_{n\to\infty} A^C_{M,h(n),\kappa}) = 0$ for all $\kappa > 0$. The set $(\limsup_{n\to\infty} A^C_{M,h(n),\kappa})^C$ consists of $\nu \in \Theta$ with an $n'(\nu) \in \mathbb{N}$ such that

$$\sup_{x\in\mathbb{R}} \left\| \frac{1}{h(n)} \sum_{i=1}^{h(n)} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| \leq \kappa$$

for all $n > n'$, and by the previous calculation

$$\mathbb{P}\left((\limsup_{n\to\infty} A^C_{M,h(n),\kappa})^C\right) = 1$$

for every $\kappa > 0$. The set $\Omega' = \bigcap_{m=1}^{\infty} (\limsup_{n\to\infty} A^C_{h(n),1/m})^C$ has thus also full measure and for all $\nu \in \Omega'$ we have

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left\| \frac{1}{h(n)} \sum_{i=1}^{h(n)} \phi(X_i(\nu))(x) - \mathbb{E}(\phi(X_1)(x)) \right\| = 0$$

as claimed. □

We finally have both the geometric and the probabilistic estimate at disposal and can now combine them to uniform concentration inequalities for admissible functions we want.

7 New results on quantitative uniform convergence

Now we spell out the new uniform concentration inequalities for admissible functions with explicit quantification we can achieve. The results for concentration along sequences of cubes and monotiling Følner sequences are similar, but since cubes are more intuitive and give explicitly calculable formulas we will treat them first. For eigenvalue-counting functions on cubes we also know the explicit boundary terms and the limiting function from Theorem 3.28, so we will also show a dedicated concentration inequality for this case.

7.1 Results for cubes

We consider functions in the Banach space

$$\mathbb{B} := \{F : \mathbb{R} \to \mathbb{R} : F \text{ right-continuous and bounded}\}$$

and start out with the general result for d-dimensional cubes

$$\Lambda_n = \{x \in \mathbb{Z}^d : \ 0 \leq x_i < n \ \forall 1 \leq i \leq d\}$$

on the lattice $\mathbb{Z}^d$ in the following Theorem.

Theorem 7.1 (Concentration inequality for admissible functions on cubes in $\mathbb{Z}^d$)
Let a be an admissible function as in Definition 4.3 on a measure space as in Definition 4.1 with $0 \leq a(\Lambda, \omega) \leq |\Lambda|$ for all finite $\Lambda \subset \mathbb{Z}^d$ and $\omega \in \Omega$. Let $2 \leq M$ and r be the correlation length from Definition 4.1.
Then there exists a limit function $a^* \in \mathbb{B}$ such that for all $\kappa > 0$ and $n > 2m > 4r$ there is a set $A_{M,n,m,\kappa} \in \mathcal{B}(\Omega)$ such that

$$\left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a^* \right\|_\infty \leq \frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{(\lfloor n/m \rfloor m)^d} + \frac{(3D+2E)\left(n^d - (n-2m)^d\right)}{(n-2m)^d} + 2\frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D+E)\left(m^d - (m-2r)^d\right)}{(m-2r)^d} + \kappa \tag{7.1}$$

for all $\omega \in A_{M,n,m,\kappa}$ and

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right) \geq 1 - M \exp\left(-\frac{\sqrt{\lfloor n/m \rfloor^d}\kappa}{K_M}\right), \tag{7.2}$$

where

$$K_M = \left(\frac{56(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)} \right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + 2q\log(2)}$$
$$< 16 \left(\frac{14(M+1)}{\log(3/2)(M-1)} + \frac{1}{\log(M)} \right)$$

If $\lfloor n/m \rfloor^d \geq 144(K_2)^2 \kappa^{-4}$ *there is moreover a set* $B_{n,m,\kappa} \in \mathcal{B}(\Omega)$ *such that* (7.1) *is true for all* $\omega \in B_{n,m,\kappa}$ *and*

$$\mathbb{P}\left(B_{n,m,\kappa}\right) \geq 1 - \exp\left(-\frac{1}{24}\kappa^2 \lfloor n/m \rfloor^d \right).$$

Furthermore

$$\lim_{n\to\infty} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a^* \right\|_\infty = 0$$

for almost all $\omega \in \Omega$.

Remark 7.2. The almost sure convergence was already established in [SSV17], but the explicit bound in (7.2) was not achieved in previous results.
If $0 \leq a(\Lambda, \omega) \leq |\Lambda|$ for all $\omega \in \Omega$ is not assumed, we still know from (4.3) that

$$\|a(\Lambda, \omega)\|_\infty \leq (D + E)\,|\Lambda|$$

with D and E as defined in (A1) and (A4). Therefore, we have

$$-(D+E) \leq \frac{a(\Lambda, \omega)}{|\Lambda|} \leq D + E$$

and we can use Remark 6.19 to gain the bound

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right) \geq 1 - M \exp\left(-\frac{\sqrt{\lfloor n/m \rfloor^d}\kappa}{2K_M(D+E)} \right)$$

instead of (7.2). The parallel result for the second concentration inequality is

$$\mathbb{P}\left(B_{n,m,\kappa}\right) \geq 1 - \exp\left(-\frac{1}{24} \frac{\kappa^2}{4(D+E)^2} \lfloor n/m \rfloor^d \right)$$

in this case.
By Lemma 6.18 the probability in (7.2) is positive if

$$\lfloor n/m \rfloor^d \geq \left(\frac{\log(M) K_M}{\kappa} \right)^2.$$

Proof of Theorem 7.1. We know from Lemma 5.1 that

$$\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r+t,\omega)}{|\Lambda_m|}\right\|_\infty \le \frac{b\left(\Lambda_{\lfloor n/m\rfloor m}\right)}{\left|\Lambda_{\lfloor n/m\rfloor m}\right|} + \frac{(3D+2E)\left(|\Lambda_n|-|\Lambda_n^m|\right)}{|\Lambda_n^m|} + \frac{b(\Lambda_m)+b(\Lambda_m^r)+(2D+E)\left(|\Lambda_m|-|\Lambda_m^r|\right)}{|\Lambda_m|}$$

for $n > 2m > 4r$ and all $\omega \in \Omega$, and by Lemma 5.2 we know that there exists a limit function $a^* \in \mathbb{B}$ such that

$$\left\| \frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|} - a^*\right\|_\infty \le \frac{b(\Lambda_m)+b(\Lambda_m^r)+(2D+E)\left(|\Lambda_m|-|\Lambda_m^r|\right)}{|\Lambda_m|}.$$

Combining these results leads to

$$\begin{aligned}
\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - a^*\right\|_\infty &\le \left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r+t,\omega)}{|\Lambda_m|}\right\|_\infty \\
&\quad + \left\| \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r+t,\omega)}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty \\
&\quad + \left\| \frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|} - a^*\right\|_\infty \\
&\le \frac{b\left(\Lambda_{\lfloor n/m\rfloor m}\right)}{\left|\Lambda_{\lfloor n/m\rfloor m}\right|} + \frac{(3D+2E)\left(|\Lambda_n|-|\Lambda_n^m|\right)}{|\Lambda_n^m|} \\
&\quad + 2\frac{b(\Lambda_m)+b(\Lambda_m^r)+(2D+E)\left(|\Lambda_m|-|\Lambda_m^r|\right)}{|\Lambda_m|} \\
&\quad + \left\| \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r+t,\omega)}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty .
\end{aligned} \tag{7.3}$$

This is where we need to apply the concentration inequalities for monotone right-continuous bounded random functions derived in Lemma 6.18 in the last chapter. First we note that $|T_{m,n}| = \lfloor n/m\rfloor^d$ and since $\mathrm{d}_{\mathrm{set}}\left(\Lambda_m^r+t_1, \Lambda_m^r+t_2\right) > r$ for $t_1 \neq t_2 \in T_{m,n}$ our assumptions on the probability measure from Definition 4.1 mean that the random variables $\left(\Pi_{\Lambda_m^r+t}\right)_{t\in T_{m,n}}$ are independent and identically distributed. Thus, $\left(\Pi_{\Lambda_m^r}\circ\gamma_t\right)_{t\in T_{m,n}}$ are independent and identically distributed, as mentioned in

Remark 4.2. Using a_Λ defined in (4.2) in Remark 4.4, we can write

$$\begin{aligned}&\left\|\frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}}\frac{a(\Lambda_m^r+t,\omega)}{|\Lambda_m|}-\frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty\\&\quad=\left\|\frac{1}{\lfloor n/m\rfloor^d}\sum_{t\in T_{m,n}}\frac{a_{\Lambda_m^r+t}(\Pi_{\Lambda_m^r+t}(\omega))}{|\Lambda_m|}-\frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty\\&\quad=\left\|\frac{1}{\lfloor n/m\rfloor^d}\sum_{t\in T_{m,n}}\frac{a_{\Lambda_m^r}(\Pi_{\Lambda_m^r}(\gamma_t(\omega)))}{|\Lambda_m|}-\frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty.\end{aligned}$$

The function $a_{\Lambda_m^r}:(\Omega_{\Lambda_m^r},\mathcal{B}(\Lambda_m^r))\to\mathbb{B}$ maps into right-continuous and bounded functions, and the function $a_{\Lambda_m^r}(x):(\Omega_{\Lambda_m^r},\mathcal{B}(\Lambda_m^r))\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ is measurable for each $x\in\mathbb{R}$ as detailed in Remark 4.4. Thus, $a_{\Lambda_m^r}$ satisfies the requirements from Lemma 6.18 and additionally we have $\mathbb{E}a_{\Lambda_m^r}(\Pi_{\Lambda_m^r}(\cdot))=\mathbb{E}a(\Lambda_m^r,\cdot)$. The random variables $\Pi_{\Lambda_m^r}\circ\gamma_t:(\Omega,\mathcal{B}(\Omega))\to(\Omega_{\Lambda_m^r},\mathcal{B}(\Lambda_m^r))$ take the place of the X_i in Lemma 6.18. Since $0\le a(\Lambda,\omega)\le|\Lambda|$ for all $\omega\in\Omega$ we have $0\le\frac{a(\Lambda_m^r,\omega)}{|\Lambda_m|}\le 1$ and we can immediately apply Lemma 6.18 to get that there exists a set $A_{M,n,m,\kappa}$ such that

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right)\ge 1-M\exp\left(-\frac{\sqrt{\lfloor n/m\rfloor^d}\kappa}{K_M}\right)\tag{7.4}$$

and

$$\left\|\frac{1}{\lfloor n/m\rfloor^d}\sum_{t\in T_{m,n}}\frac{a_{\Lambda_m^r}(\Pi_{\Lambda_m^r}(\gamma_t(\omega)))}{|\Lambda_m|}-\frac{\mathbb{E}a(\Lambda_m^r,\cdot)}{|\Lambda_m|}\right\|_\infty\le\kappa\tag{7.5}$$

for all $\omega\in A_{M,n,m,\kappa}$ with K_M as specified in Theorem 7.1 and $2\le M$.The same lemma shows the existence of a set $B_{n,m,\kappa}$ such that (7.5) is true for all $\omega\in B_{n,m,\kappa}$ and

$$\mathbb{P}\left(B_{n,m,\kappa}\right)\ge 1-\exp\left(-\frac{1}{24}\kappa^2\lfloor n/m\rfloor^d\right)$$

provided $\lfloor n/m\rfloor^d\ge\left(12\frac{K_2}{\kappa^2}\right)^2$. Combining (7.3) with (7.5) leads to

$$\begin{aligned}\left\|\frac{a(\Lambda_n,\omega)}{|\Lambda_n|}-a^*\right\|_\infty\le\frac{b\left(\Lambda_{\lfloor n/m\rfloor m}\right)}{\left|\Lambda_{\lfloor n/m\rfloor m}\right|}+\frac{(3D+2E)\left(|\Lambda_n|-|\Lambda_n^m|\right)}{|\Lambda_n^m|}\\+2\frac{b(\Lambda_m)+b(\Lambda_m^r)+(2D+E)\left(|\Lambda_m|-|\Lambda_m^r|\right)}{|\Lambda_m|}+\kappa.\end{aligned}$$

Since $|\Lambda_n|=n^d$ and $|\Lambda_n^r|=(n-2r)^d$ for cubes Λ_n we arrive at (7.1) and the probability estimate is provided in (7.4).

We have $n > 2m$ and therefore

$$h(n) := \left\lfloor \frac{n}{m} \right\rfloor^d \geq \left\lfloor \frac{n}{m} \right\rfloor \geq \frac{n}{m} - 1 \geq \frac{n}{m} - \frac{n}{2m} = \frac{n}{2m}.$$

From Corollary 6.21 follows almost sure uniform convergence, i.e.

$$\lim_{n\to\infty} \left\| \frac{1}{\lfloor n/m \rfloor^d} \sum_{t\in T_{m,n}} \frac{a_{\Lambda_m^r}(\Pi_{\Lambda_m^r}(\gamma_t(\omega)))}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty = 0$$

for almost all $\omega \in \Omega$ and $m > 2r$. As shown in Lemma 5.1 we have

$$\lim_{n\to\infty} \frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} + \frac{(3D + 2E)\left(|\Lambda_n| - |\Lambda_n^m|\right)}{|\Lambda_n^m|} = 0$$

for $m > 2r$ and thus (7.1) leads to

$$\lim_{n\to\infty} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a^* \right\|_\infty \leq 2\frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D + E)\left(|\Lambda_m| - |\Lambda_m^r|\right)}{|\Lambda_m|}$$

for $m > 2r$ and almost all $\omega \in \Omega$. Since Lemma 5.1 also showed

$$\lim_{m\to\infty} \frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D + E)\left(|\Lambda_m| - |\Lambda_m^r|\right)}{|\Lambda_m|} = 0$$

we arrive at the claimed almost sure convergence. □

The main application we have in mind are the eigenvalue-counting functions from Definition 3.25, and since we know their boundary terms from Lemma 4.5 we can make the bounds from Theorem 7.1 explicit. This allows us to choose κ and m as functions of n. We also know that the limit function is the integrated density of states if the conditions of Theorem 3.28 are satisfied.

Corollary 7.3 (Sub-root-exponential concentration inequality for eigenvalue-counting functions)
Let N_ω^Λ be the normalized eigenvalue-counting function on a finite set $\Lambda \subset \mathbb{Z}^d$ as in Definition 3.25 for an Anderson operator satisfying the requirements of Theorem 3.28, and r the constant from those requirements. Let N be the integrated density of states as defined in Theorem 3.26 and $2 \leq M$. Then there is a set $A_{M,n} \in \mathcal{B}(\Omega)$ such that the uniform distance between the normalized eigenvalue-counting function on cubes Λ_n and the integrated density of states is bound by

$$\begin{aligned} \left\|N_\nu^{\Lambda_n} - N\right\|_\infty \leq{}& 32d\frac{1}{n} + 52\left(2^d - 1\right)\frac{1}{n^{1-(1/k)}} \\ &+ \left(8d + 4r(2^d - 1) + 72dr + 1\right)\frac{1}{\sqrt[k]{n} - 1} \end{aligned} \tag{7.6}$$

for all $\nu \in A_{M,n}$ *and*

$$\mathbb{P}\left(A_{M,n}\right) \geq 1 - M \exp\left(-\frac{\sqrt{\lfloor n/\lfloor \sqrt[k]{n} \rfloor \rfloor^d}}{\lfloor \sqrt[k]{n} \rfloor \mathring{K}_M}\right) \tag{7.7}$$

provided $n > (2r+1)^k$ *and* $n > 4$*, where*

$$k = \begin{cases} 4 & d = 1 \\ 3 & d = 2 \\ 2 & d \geq 3 \end{cases}$$

and

$$\begin{aligned} \mathring{K}_M &= \left(\frac{40(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + 2q \log(2)} \\ &< 16 \left(\frac{10(M+1)}{\log(3/2)(M-1)} + \frac{1}{\log(M)}\right). \end{aligned}$$

Remark 7.4. From Remark 7.2 follows that we only get non-trivial results for the probability if

$$\frac{\lfloor n/\lfloor \sqrt[k]{n} \rfloor \rfloor^d}{\lfloor \sqrt[k]{n} \rfloor^2} \geq \left(\log(M) \mathring{K}_M\right)^2.$$

Since $x - 1 \leq \lfloor x \rfloor \leq x$, we have

$$\frac{\lfloor n/\lfloor \sqrt{n} \rfloor \rfloor^d}{\lfloor \sqrt{n} \rfloor^2} \geq \lfloor \sqrt{n} \rfloor^{d-2} \geq \left(\sqrt{n} - 1\right)^{d-2}$$

for $d \geq 3$, and thus we are guaranteed non-trivial results if

$$n > \left(\left(\log(M) \mathring{K}_M\right)^{2/(d-2)} + 1\right)^2.$$

For $d = 1, 2$ we use

$$\frac{\lfloor n/\lfloor \sqrt[k]{n} \rfloor \rfloor^d}{\lfloor \sqrt[k]{n} \rfloor^2} \geq \frac{\lfloor n^{1-1/k} \rfloor^d}{\lfloor \sqrt[k]{n} \rfloor^2} \geq \frac{\left(n^{1-1/k} - 1\right)^d}{n^{2/k}}. \tag{7.8}$$

Thus, we need to ensure

$$n^{1-1/k} - 1 \geq n^{2/(dk)} \left(\log(M) \mathring{K}_M\right)^{2/d},$$

and since

$$n^{1-1/k-2/(dk)} - n^{-2/(dk)} \geq n^{1-1/k-2/(dk)} - 1$$

we can ensure a non-trivial result if

$$n^{1-1/k-2/(dk)} \geq \left(\log(M)\mathring{K}_M\right)^{2/d} + 1.$$

Therefore, equation (7.7) gives guaranteed non-trivial results for

$$n > \begin{cases} \left(\left(\log(M)\mathring{K}_M\right)^2 + 1\right)^4 & d = 1, \\ \left(\left(\log(M)\mathring{K}_M\right) + 1\right)^3 & d = 2, \\ \left(\left(\log(M)\mathring{K}_M\right)^{2/(d-2)} + 1\right)^2 & d \geq 3. \end{cases}$$

Proof. We know from Lemma 4.5 that the eigenvalue-counting functions are admissible with boundary function $b(\Lambda) = 8\left|\Lambda \setminus \Lambda^1\right|$, $D = 8$ and $E = 1$. Additionally we have $|\Lambda_n| = n^d$ and $|\Lambda_n^r| = (n-2r)^d$ for cubes Λ_n, which allows us to compute

$$\frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} = 8\frac{\left|\Lambda_{\lfloor n/m \rfloor m}\right| - \left|\Lambda^1_{\lfloor n/m \rfloor m}\right|}{\left|\Lambda_{\lfloor n/m \rfloor m}\right|} = 8\left(1 - \left(\frac{\lfloor n/m \rfloor m - 2}{\lfloor n/m \rfloor m}\right)^d\right).$$

Application of Theorem 7.1 with $\frac{a(\Lambda_n,\nu)}{|\Lambda_n|} = N_\nu^{\Lambda_n}$ leads for $n > 2m > 4r$ to the existence of a limit function a^* and a set $A_{M,n,m,\kappa}$ such that

$$\begin{aligned} \left\|\frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - a^*\right\|_\infty \leq{}& \frac{b\left(\Lambda_{\lfloor n/m \rfloor m}\right)}{(\lfloor n/m \rfloor m)^d} + \frac{(3D+2E)\left(n^d - (n-2m)^d\right)}{(n-2m)^d} \\ &+ 2\frac{b(\Lambda_m) + b(\Lambda_m^r) + (2D+E)\left(m^d - (m-2r)^d\right)}{(m-2r)^d} + \kappa \end{aligned} \tag{7.9}$$

for all $\nu \in A_{M,n,m,\kappa}$ and

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right) \geq 1 - M\exp\left(-\frac{\sqrt{\lfloor n/m \rfloor^d}\kappa}{\mathring{K}_M}\right).$$

The improved constant $\mathring{K}_M$ instead of K_M follows from Remark 6.20 and the fact that $\lim\limits_{x \searrow -\infty} N_\nu^{\Lambda_n} = 0$ for all $\nu \in \Omega$.

Now we just calculate the right side of (7.9) to get

$$\begin{aligned}
\left\|N_\nu^{\Lambda_n} - a^*\right\|_\infty \le{}& 8\left(1 - \left(\frac{\lfloor n/m\rfloor m - 2}{\lfloor n/m\rfloor m}\right)^d\right) + 26\left(\left(\frac{n}{n-2m}\right)^d - 1\right) \\
&+ 2\frac{m^d - (m-2)^d + (m-2r)^d - (m-2r-2)^d + 17(m^d - (m-2r)^d)}{m^d} + \kappa \\
={}& 8\left(1 - \left(1 - \frac{2}{\lfloor n/m\rfloor m}\right)^d\right) + 26\left(\left(1 + \frac{2m/n}{1-2m/n}\right)^d - 1\right) \\
&+ 2\left(1 - \left(1 - \frac{2}{m}\right)^d\right) + 2\frac{(m-2r)^d - (m-2r-2)^d}{m^d} \\
&+ 34\left(1 - \left(1 - \frac{2r}{m}\right)^d\right) + \kappa
\end{aligned} \tag{7.10}$$

This expression actually only holds for $m > 2r+1$ in contrast to $2m > 4r$ used in the previous lemmas and theorems, since we exclude the edge case where $(\Lambda_m^r)^1 = \emptyset$ but $m - 2r - 2 < 0$ in favour of a unified expression.
Now we use Bernoulli's inequality

$$1 - (1-y)^d \le yd \tag{7.11}$$

for $y \le 1$ for the first, third and fifth term on the right of (7.10). Since $\lfloor n/m\rfloor m \ge n - m$ we have

$$\frac{1}{\lfloor n/m\rfloor m} \le \frac{1}{n-m} \le \frac{1}{n}\frac{1}{1-m/n} \le \frac{2}{n}.$$

if $n > 2m$.
If $n \ge 4$ we can apply Bernoulli's inequality and gain the bounds

$$\begin{aligned}
\left(1 - \left(1 - \frac{2}{\lfloor n/m\rfloor m}\right)^d\right) &\le \frac{2d}{\lfloor n/m\rfloor m} \le \frac{4d}{n} \\
\left(1 - \left(1 - \frac{2}{m}\right)^d\right) &\le \frac{2d}{m} \\
\left(1 - \left(1 - \frac{2r}{m}\right)^d\right) &\le \frac{2dr}{m}
\end{aligned}$$

since $m > 2r+1$ and thus $m \ge 2$ as well as $m \ge 2r$ was already required. Next we note that by the binomial theorem

$$(1+x)^d = \sum_{j=0}^{d} \binom{d}{j} x^j \tag{7.12}$$

we have

$$(1+x)^d - 1 \leq |x| \sum_{j=1}^{d} \binom{d}{j} |x|^{j-1} \leq |x| \sum_{j=1}^{d} \binom{d}{j} = |x|\left(2^d - 1\right)$$

for $|x| \leq 1$. From

$$\frac{x}{1-x} \leq 2x \leq 1 \Leftrightarrow x \leq \frac{1}{2}$$

follows

$$\left(\left(1 + \frac{2m/n}{1-2m/n}\right)^d - 1\right) \leq \frac{2m}{n}\left(2^d - 1\right)$$

as long as $4m < n$.
For the fourth term we use both (7.11) and (7.12) for

$$\left(1 - \frac{2r}{m}\right)^d - \left(1 - \frac{2r+2}{m}\right)^d \leq \left(2^d - 1\right)\frac{2r}{m} + \frac{d(2r+2)}{m} = \frac{2r2^d + 2d(r+1) - 2r}{m}.$$

In conclusion we get

$$\left\|N_\nu^{\Lambda_n} - a^*\right\|_\infty \leq 32d\frac{1}{n} + 52\left(2^d - 1\right)\frac{m}{n} + \left(8d + 4r(2^d - 1) + 72dr\right)\frac{1}{m} + \kappa$$

by applying all the previous bounds to (7.10).
Now we want to choose m and κ as functions of n with the fastest convergence of the uniform error possible. Since $n > m$ the first term on the right side does not dominate the error, the relevant terms for the convergence are only $\frac{m}{n}$, $\frac{1}{m}$ and κ. The optimal choice for the first two of these is $m = \sqrt{n}$, since any other choice will either increase the first or the second. As m needs to be a natural number we will choose $m(n) = \lfloor\sqrt{n}\rfloor$. From this follows that we can choose $\kappa(n) = \frac{1}{\lfloor\sqrt{n}\rfloor}$ without slowing the convergence.
But now we need to check that $\sqrt{\lfloor n/m(n)\rfloor^d}\kappa(n)$ still grows in n, otherwise there is no concentration. We have

$$\sqrt{\lfloor n/m(n)\rfloor^d}\kappa(n) \geq \sqrt{\lfloor n/\sqrt{n}\rfloor^d}\frac{1}{\lfloor\sqrt{n}\rfloor} = \lfloor\sqrt{n}\rfloor^{d/2-1},$$

so our choice works for $d \geq 3$, but not for $d = 1, 2$.
To check what we have to change for these cases we start out with a general case of $m(n) \sim n^j$ for some $j < 1/2 \in \mathbb{R}$ and $\kappa(n) = 1/m(n)$. The convergence of the uniform bound will then be dominated by n^{-j}, since $\frac{m}{n} \sim n^{-(1-j)}$. We have to ensure that

$$\sqrt{\lfloor n/m(n)\rfloor^d}\kappa(n) \sim n^{\frac{d(1-j)}{2} - j}$$

grows in n, i.e.

$$j < \frac{d}{2(1+d/2)} = \begin{cases} \frac{1}{3} & d=1 \\ \frac{1}{2} & d=2 \end{cases}.$$

Thus, we choose

$$\begin{aligned} d=1: m(n) &= \lfloor \sqrt[4]{n} \rfloor,\ \kappa(n) = \frac{1}{\lfloor \sqrt[4]{n} \rfloor} \\ d=2: m(n) &= \lfloor \sqrt[3]{n} \rfloor,\ \kappa(n) = \frac{1}{\lfloor \sqrt[3]{n} \rfloor}. \end{aligned}$$

and arrive at

$$\begin{aligned} \left\| N_\nu^{\Lambda_n} - a^* \right\|_\infty &\leq 32d\frac{1}{n} + 52\left(2^d-1\right)\frac{\lfloor \sqrt[k]{n} \rfloor}{n} + \left(8d + 4r(2^d-1) + 72dr + 1\right)\frac{1}{\lfloor \sqrt[k]{n} \rfloor} \\ &\leq 32d\frac{1}{n} + 52\left(2^d-1\right)\frac{1}{n^{1-1/k}} + \left(8d + 4r(2^d-1) + 72dr + 1\right)\frac{1}{\sqrt[k]{n}-1} \end{aligned}$$

where

$$k = \begin{cases} 4 & d=1 \\ 3 & d=2 \\ 2 & d \geq 3 \end{cases}.$$

We required $m > 2r+1$ before, so the bound is valid if $n > 4$ and $n > (2r+1)^k$. We also need to ensure $n > 2m$, but since

$$\frac{n}{m} = \frac{n}{\lfloor \sqrt[k]{n} \rfloor} \geq n^{1-\frac{1}{k}} \geq \sqrt{n}$$

this is already covered by $n > 4$.
This shows the bound in (7.6), but the identification of a^* with N is still left. As shown in Theorem 7.1 the uniform convergence to a^* is almost sure, and Theorem 3.28 states that $N_\nu^{\Lambda_n}$ also converges almost surely pointwise to N for all continuity points of N, thus N and a^* are identical on the continuity points of N. The integrated density of states is monotone increasing, bounded and right-continuous by Remark 3.27. Such a function can only have countably many discontinuities, see for example [Kir07, Lemma 5.10]. As a result we get that the points of continuity of N are dense in $\mathbb{R}$. Thus, we can find for each point of discontinuity x of N a sequence $(x_n)_{n\in\mathbb{N}}$ in the points of continuity of N such that $x_n \searrow x$. Since both N and a^* are right-continuous, we have $N(x) = \lim_{n\to\infty} N(x_n) = \lim_{n\to\infty} a^*(x_n) = a^*(x)$. Therefore, $N = a^*$. □

We can use the sub-exponential concentration inequality of Corollary 6.14 instead of the sub-root-exponential one as well, but we will need larger cubes for non-trivial results and only achieve the best uniform convergence at higher dimensions.

Corollary 7.5 (Sub-exponential concentration inequality for eigenvalue-counting functions)
In the setting of Corollary 7.3 there is a set $B_n \in \mathcal{B}(\Omega)$ *such that* (7.6) *is true for all* $\nu \in B_n$ *and*

$$\mathbb{P}\left(B_n\right) \geq 1 - \exp\left(-\frac{1}{24}\frac{\lfloor n/\lfloor \sqrt[k]{n}\rfloor\rfloor^d}{\lfloor \sqrt[k]{n}\rfloor}\right)$$

with

$$k \begin{cases} > \frac{4+d}{d} & d < 5 \\ = 2 & d \geq 5 \end{cases},$$

provided

$$n > \max\left\{4, (2r+1)^k, \left(\left(12\mathring{K}_2\right)^{2/d} + 1\right)^{\frac{dk}{dk-d-4}}\right\}$$

where $\mathring{K}_2 = \left(\frac{120}{\log(3/2)} + \frac{4}{\log(2)}\right)\sum_{q=0}^{\infty} 2^{-q}\sqrt{2+2q\log(2)} < \left(\frac{480}{\log(3/2)} + \frac{16}{\log(2)}\right).$

Proof. The bound in (7.6) follows from Theorem 7.1 just as in the proof of Corollary 7.3, but using the statement for $B_{n,m,\kappa}$ instead of $A_{M,n,m,\kappa}$. Just as in that proof the best choice for $m(n)$ arising from the uniform bound would be $m(n) = \lfloor\sqrt{n}\rfloor$, but we also need to ensure that

- $\lfloor n/m(n)\rfloor^d/m(n)^2$ is growing in n (which would be satisfied by the same $m(n)$ calculated for Corollary 7.3),
- $\lfloor n/m\rfloor^d/m(n)^4 \geq \left(12\mathring{K}_2\right)^2$ for the formulas from Theorem 7.1 to be applicable.

By setting $m(n) \sim n^{1/k}$ for some yet to be determined k we get

$$\frac{\lfloor n/m\rfloor^d}{m(n)^4} \sim n^{d(1-1/k)-4/k},$$

and thus need to ensure that $d(1-1/k) - 4/k > 0$ which is equivalent to $k > \frac{4+d}{d}$. Our favoured case $k = 2$ satisfies this condition only for $d \geq 5$. Just as in (7.8) we have

$$\frac{\lfloor n/m\rfloor^d}{m(n)^4} \geq \frac{\left(n^{1-1/k} - 1\right)^d}{n^{4/k}}$$

and ensure the applicability if

$$n^{1-1/k} - 1 \geq n^{4/(dk)}\left(12\mathring{K}_2\right)^{2/d}$$

or equivalently

$$n^{1-1/k-4/(dk)} - n^{-4/(dk)} \geq \left(12\mathring{K}_2\right)^{2/d}.$$

As in Remark 7.4 this is guaranteed if

$$n^{1-1/k-4/(dk)} - 1 \geq \left(12\mathring{K}_2\right)^{2/d}$$

or equivalently

$$n > \left(\left(12\mathring{K}_2\right)^{2/d} + 1\right)^{\frac{d}{d-\frac{(d+4)}{k}}}$$

as claimed. □

7.2 Results for monotiles

The results for cubes on $\mathbb{Z}^d$ can nearly directly be applied to other monotiling Følner sequences as well. Recall that for a monotiling Følner sequence $(\Lambda_n)_{n\in\mathbb{N}}$ we defined $\rho(m) := \max_{x,y\in\Lambda_m} \mathrm{d}_{\mathbb{Z}^d}(x,y)$ and

$$\Lambda_{m,n} := \bigcup_{t\in T_{m,n}} (\Lambda_m + t)$$

with $T_{m,n} := \{t \in T_m \mid \Lambda_m + t \subseteq \Lambda_n\}$ and the tiling T_m from Definition 5.3.

Theorem 7.6 (Sub-root-exponential concentration inequality for admissible functions on monotiling Følner sequences)
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence as in Definition 5.3 and $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be a generalized admissible function as in Definition 5.6 with $0 \leq a(\Lambda,\omega) \leq |\Lambda|$ for all $\omega \in \Omega$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space fulfilling the requirements of Definition 4.1 with the associated constant r. Let further $2 \leq M$ and $W(r), \rho(m)$ as defined in Remark 5.4 and Lemma 5.5. Then there exists a limit function $a^* \in \mathbb{B}$ such that for all $\kappa > 0$, $m > W(r)$ and $n > \max\{W(r), W(\rho(m))\}$ there is a set $A_{M,n,m,\kappa} \in \mathcal{B}(\Omega)$ such that

$$\begin{aligned}\left\|\frac{a(\Lambda_n,\nu)}{|\Lambda_n|} - a^*\right\|_\infty \leq{}& (3D+2E)\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} \\ &+ 2\left(\frac{b(\Lambda_m)}{|\Lambda_m|} + \frac{b(\Lambda_m^r)}{|\Lambda_m|} + (2D+E)\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|}\right) + \kappa\end{aligned}$$

for all $\nu \in A_{M,n,m,\kappa}$.
The probability of $A_{M,n,m,\kappa}$ can be bounded by

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right) \geq 1 - M \exp\left(-\frac{\sqrt{|T_{m,n}|}\kappa}{K_M}\right),$$

where

$$\begin{aligned} K_M &= \left(\frac{56(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right) \sum_{q=0}^{\infty} 2^{-q}\sqrt{2+2q\log(2)} \\ &\leq 16\left(\frac{14(M+1)}{\log(3/2)(M-1)} + \frac{1}{\log(M)}\right). \end{aligned}$$

Furthermore if there is an $n_0 \in \mathbb{N}$ and $c > 0$ such that $|\Lambda_n| \geq cn$ for all $n \geq n_0$, then

$$\lim_{n\to\infty} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a^* \right\|_\infty = 0$$

for almost all $\omega \in \Omega$.

The lower bound $|\Lambda_n| \geq cn$ can be satisfied by passing to a subsequence since $|\Lambda_n| \to \infty$ by (b) of Lemma 5.5. Of course there is also a sub-exponential concentration inequality in this case, which follows in exactly the same way as in Theorem 7.1.
It is also possible to extend the result to generalized admissible functions without assuming additionally $0 \leq a(\Lambda, \omega) \leq |\Lambda|$ just as in Remark 7.2.

Proof. We can follow the proof of Theorem 7.1, but instead of Lemmas 5.1 and 5.2 we use the monotile equivalents from Lemmas 5.7 and 5.8 to show the existence of a limit function a^* and

$$\begin{aligned} \left\| \frac{a(\Lambda_n, \omega)}{|\Lambda_n|} - a^* \right\|_\infty &\leq (3D+2E)\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + \frac{b(\Lambda_{m,n})}{|\Lambda_{m,n}|} \\ &\quad + 2\left(\frac{b(\Lambda_m)}{|\Lambda_m|} + \frac{b(\Lambda_m^r)}{|\Lambda_m|} + (2D+E)\frac{|\partial^r(\Lambda_m)|}{|\Lambda_m|}\right) \\ &\quad + \left\| \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty . \end{aligned} \tag{7.13}$$

Again we need to apply the sub-root-exponential concentration inequality of Lemma 6.18. Because we excluded the r-boundaries we can ensure $\mathrm{d}_{\mathrm{set}}\left(\Lambda_m^r + t_1, \Lambda_m^r + t_2\right) > r$ for $t_1 \neq t_2 \in T_{m,n}$ in any tiling. Just as in the proof of Theorem 7.1 we can show that $\left(\Pi_{\Lambda_m^r} \circ \gamma_t\right)_{t\in T_{m,n}}$ is an independent and identically distributed sequence. The restricted admissible functions a_Λ as defined in (4.2) in Remark 4.4 satisfy all needed properties in the same way as in the proof of Theorem 7.1, and we can again use

$$\begin{aligned}&\left\| \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a(\Lambda_m^r + t, \omega)}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty \\ &= \left\| \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a_{\Lambda_m^r + t}(\Pi_{\Lambda_m^r + t}(\omega))}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty \\ &= \left\| \frac{1}{|T_{m,n}|} \sum_{t \in T_{m,n}} \frac{a_{\Lambda_m^r}(\Pi_{\Lambda_m^r}(\gamma_t(\omega)))}{|\Lambda_m|} - \frac{\mathbb{E}a(\Lambda_m^r, \cdot)}{|\Lambda_m|} \right\|_\infty .\end{aligned}$$

and apply Lemma 6.18 with the random variables $\Pi_{\Lambda_m^r} \circ \gamma_t$ taking the place of the X_i. The existence and probability of $A_{M,n,m,\kappa}$ follows just as in the proof of Theorem 7.1. The almost sure convergence follows if

$$\sum_{n=1}^\infty \exp\left(- \frac{\sqrt{|T_{m,n}|\,\kappa}}{K_M} \right) < \infty.$$

As shown in Corollary 6.21 we have

$$\sum_{n=1}^\infty \mathrm{e}^{-c\sqrt{n}} < \infty$$

for any $c > 0$ and thus it suffices to show that there exists an $n_0' \in \mathbb{N}$ and a $b > 0$ such that $|T_{m,n}| \geq bn$ for all $n \geq n_0'$. One thing we can use here is (b) of Lemma 5.5 to see that

$$\lim_{n \to \infty} \frac{|T_{m,n}|}{|\Lambda_n|} = \frac{1}{|\Lambda_m|}$$

and because of that there is an $\tilde{n}_0 \in \mathbb{N}$ such that $|T_{m,n}| \geq \frac{|\Lambda_n|}{2|\Lambda_m|}$ for all $n \geq \tilde{n}_0$. If there is an $n_0 \in \mathbb{N}$ such that $|\Lambda_n| \geq cn$ for all $n \geq n_0$, we can now set $n_0' = \max\{\tilde{n}_0, n_0\}$ and combine both previous bounds to $|T_{m,n}| \geq \frac{cn}{2|\Lambda_m|}$ for all $n \geq n_0'$ just as required. Then the almost sure convergence follows from Corollary 6.21. □

The error terms for monotiling Følner sequences are nearly identical to those for cubes, but the main difference is the fact that for a general sequence the tilings $T_{m,n}$ and the sets $\Lambda_{m,n}$ do not need to behave as nicely as for cubes, so it might be harder to calculate error terms. However, we can say something about the limit object, namely that it is the same for all (fast enough growing) monotiling Følner sequences.

Corollary 7.7
The limit function $a^* \in \mathbb{B}$ from Theorem 7.6 is the same for all monotiling Følner sequences $(\Lambda_n)_{n\in\mathbb{N}}$ for which there exists an $n_0 \in \mathbb{N}$ and $c > 0$ such that $|\Lambda_n| \geq cn$ for all $n \geq n_0$.

Proof. Any two monotiling Følner sequences can be combined into a single one by alternately taking elements from both sequences, and if c_1 and $n_{0,1}$ are the associated coefficients of the first sequence and c_2 and $n_{0,2}$ those of the second, then both sequences almost surely converge and $N_a := \max\{2n_{0,1}+1, 2n_{0,2}+1\}$ and $c_a := \min\{c_1, c_2\}$ form fitting coefficients for the combined alternating sequence. The combined sequence has therefore also an almost-sure uniform limit, which has to coincide with the limits of both original sequences, and consequently these two limits have to be the same. □

This means in particular that the limit function a^* that arises as a limit of admissible functions on cubes in Theorem 7.1 is not special and choosing for example rectangles would not change it.
The previous results for monotiles are not just valid on $\mathbb{Z}^d$ and can be adapted for any Cayley graph of a finitely generated amenable group. However, this hinges on the existence of a monotiling Følner sequence for the amenable group under consideration. Such a sequence exists for (among others) all elementary amenable groups as well as all residually finite amenable groups [LSV10]. However, it is not clear if any Cayley graph of a finitely generated amenable group has a sequence of monotiling Følner sequences [SSV18] [LSV10]. The next chapter will first define finitely generated amenable groups and their Cayley graph and then show a way to nevertheless define and prove an analogous result that does not depend on the existence of a monotiling Følner sequence.

8 More general geometries and interactions

8.1 Generalization to Cayley graphs of finitely generated amenable groups

The previous results all relied on the specific structure of $\mathbb{Z}^d$ and the resulting tiling properties. However, Schumacher, Schwarzenberger and Veselić also extended their setting from [SSV17], including the results described in Chapter 4, to a much more general case. In [SSV18] they proved concentration inequalities for admissible functions on Cayley graphs of amenable groups, but again without explicit values for some parameters. Instead of tilings by cubes of rising side-length, they applied quantitative results by Pogorzelski and Schwarzenberger[PS16] for ε-quasi tilings developed by Ornstein and Weiß [OW87]. The methods of Chapter 6 can also be applied in this case, and this chapter contains the resulting concentration inequalities based on [SSV18].

Definition 8.1 (Cayley graph of a finitely generated amenable group, [SSV18])
*Let G be a group generated by a symmetric generator set $S = S^{-1} \subseteq G \setminus id$, i.e. for every $x \in G$ there exists an $n \in \mathbb{N}$ and $y_i \in S$, $1 \leq i \leq n$ such that $x = y_1 ... y_{n-1} y_n$. Since G need not be abelian, we will write the group action multiplicative instead of additive like in Chapter 4. G is called **amenable** if there exists a sequence $(\Lambda_n)_{n\in\mathbb{N}}$ in the set $\mathcal{C}$ of all finite subsets of G such that for each $K \in \mathcal{C}$*

$$\frac{|\Lambda_n \Delta K \Lambda_n|}{|\Lambda_n|} \to 0$$

*where $K\Lambda_n := \{kl \mid k \in K, l \in \Lambda_n\}$ is the pointwise group multiplication of sets and Δ is the symmetric difference. Such a sequence is called a **Følner sequence**.*
*The group G and the generator set S allow the construction of an undirected graph (G, E) via $(x, y) \in E \Leftrightarrow xy^{-1} \in S$. This graph is called the **Cayley graph of G with respect to S**.*
The graph metric induced by this graph is

$$\mathrm{d}(x, y) = \inf\{n \in \mathbb{N} \mid \exists z_1, ..., z_n \in S : y = xz_1...z_n\}$$

and distances

$$\mathrm{d}_{set}(A, B) := \min \{\mathrm{d}(x, y) \mid x \in A, y \in B\}.$$

of sets and r-boundaries

$$\partial^r(A) := \{x \in A : \mathrm{d}_{set}(x, G \setminus A) \le r\} \cup \{x \in G \setminus A : \mathrm{d}_{set}(x, A) \le r\}$$

can be defined analogous to Section 4.1.
If $(\Lambda_n)_{n\in\mathbb{N}}$ is a Følner sequence, then the sequence $\frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|}$ converges to 0 and $\frac{|\Lambda_n^r|}{|\Lambda_n|}$ converges to 1 for all $r > 0$.
It is sufficient to show that either

$$\frac{|S\Lambda_n \setminus \Lambda_n|}{|\Lambda_n|} \to 0 \qquad \text{or} \qquad \forall r > 0 : \frac{|\partial^r(\Lambda_n)|}{|\Lambda_n|} \to 0 \tag{8.1}$$

to prove that $(\Lambda_n)_{n\in\mathbb{N}}$ is a Følner sequence [Sch14, Lemmas 2.7, 2.8].
*Furthermore every amenable group contains a **nested Følner sequence**, i.e. a Følner sequence $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{C}$ such that for all $n \in \mathbb{N}$ the sequence fulfils $\{id\} \subseteq Q_n \subseteq Q_{n+1}$ [PS16, Lemma 2.6].*

The lattice $\mathbb{Z}^d$ with edges between next neighbours falls under this definition, since its vertex set is the group $\mathbb{Z}^d$ generated by the set S containing the positive and negative unit vectors in all d directions. As shown in (4.7) the cubes with side length n fulfil (8.1) and thus form a Følner sequence.
Just as in the case of $\mathbb{Z}^d$ a coloring/potential on the graph can be introduced via a set of colors $\mathcal{A} \in \mathcal{B}(\mathbb{R})$ and the sample set $\Omega = \mathcal{A}^G$. We can also define a translation on Ω for each $g \in G$ via

$$\gamma_g \colon \Omega \to \Omega,\ (\gamma_g\omega)_x = \omega_{xg}.$$

With these definitions we can also define a probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ and require that $\mathbb{P}$ again satisfies (M1) and (M2) from Definition 4.1. Furthermore we can again define generalized admissible functions $a \colon \mathcal{C} \times \Omega \to \mathbb{B}$ that fulfil the requirements from Definition 5.6, with the formal change of using multiplication instead of addition as the group action on the graph. The setup is thus very similar to the case $\mathbb{Z}^d$. The next step would be to define a tiling of a Λ_n from the Følner sequence with a smaller Λ_m, but this is sadly not guaranteed to exist. However, there is a weaker form, an ε-quasi tiling.

Definition 8.2 (ε-quasi tiling (as used in [SSV18]))
Let G be a finitely generated group, $\alpha, \varepsilon \in (0,1)$ and I some index set.

- *Sets $K_i \in \mathcal{C},\ i \in I$ are said to α-**cover** a set $\Lambda \in \mathcal{C}$ if $\bigcup_{i\in I} K_i \subset \Lambda$ and $\left|\bigcup_{i\in I} K_i\right| \ge \alpha\,|\Lambda|$.*

- *Sets* $K_i \in \mathcal{C},\ i \in I$ *are said to be* ε***-disjoint*** *if there are subsets* $\mathring{K}_i \subseteq K_i,\ i \in I$ *such that* $\left|K_i \setminus \mathring{K}_i\right| \leq \varepsilon |K_i|$ *for all* $i \in I$ *and* $\mathring{K}_i$ *is disjoint from* $\bigcup_{j\in I\setminus\{i\}} \mathring{K}_j$ *for all* $i \in I$.
- *Sets* $K_i \in \mathcal{C},\ i \in I$ *are said to be an* ε***-quasi tiling*** *of* $\Lambda \in \mathcal{C}$ *if there are* $T_i \in \mathcal{C},\ i \in I$ *such that*
  1. *the elements of* $\{K_iT_i \mid i \in I\}$ *are pairwise disjoint,*
  2. *the family* $\{K_iT_i \mid i \in I\}$ $(1-2\varepsilon)$*-covers* Λ*,*
  3. *for each* $i \in I$*, the elements of* $\{K_it \mid t \in T_i\}$ *are* ε*-disjoint.*

  The set T_i *is called the* ***center set*** *for the* ***tile*** K_i.

The important result for amenable groups is the following theorem taken from [SSV18], which is a reformulation of more general results proven in [PS16].

Theorem 8.3 ([SSV18, Theorem 3.2])
Let G *be a finitely generated amenable group,* $(Q_n)_{n\in\mathbb{N}}$ *a nested Følner sequence and* $\varepsilon \in (0, 1/10)$*. Then there exists a finite and strictly increasing selection of sets* $K_i \in \{Q_n \mid n \in \mathbb{N}\}$*,* $i \in \{1, ..., N(\varepsilon)\}$ *where*

$$N(\varepsilon) := \left\lceil \frac{\log(\varepsilon)}{\log(1-\varepsilon)} \right\rceil,$$

with the following quasi-tiling property:
For each Følner sequence $(\Lambda_n)_{n\in\mathbb{N}}$*, there exists* $j_0(\varepsilon) \in \mathbb{N}$ *such that for all* $j \geq j_0(\varepsilon)$*, the sets* K_i*,* $i \in \{1, ..., N(\varepsilon)\}$ *are an* ε*-quasi tiling of* Λ_j*.*
For all $j \geq j_0(\varepsilon)$ *and all* $i \in \{1, ..., N(\varepsilon)\}$*, the proportion of* Λ_j *covered by the tile* K_i *satisfies*

$$\left| \frac{\left|K_iT_i^j\right|}{|\Lambda_j|} - \eta_i(\varepsilon) \right| \leq \frac{\varepsilon^2}{N(\varepsilon)} \tag{8.2}$$

where T_i^j *is the center set of the tile* K_i *for the* ε*-quasi tiling of* Λ_j *and where*

$$\eta_i(\varepsilon) := \varepsilon(1-\varepsilon)^{N(\varepsilon)-i}.$$

There are also some important properties of $\eta_i(\varepsilon)$ and $N(\varepsilon)$ that were shown in [SSV18] and [PS16], namely

Lemma 8.4
With the definitions of Theorem 8.3 holds:

- *[PS16, Remark 4.3] For each* $\varepsilon \in (0,1)$

$$\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) = 1 - (1-\varepsilon)^{N(\varepsilon)} \leq 1$$

- *[SSV18, Lemma 3.3] For each* $\varepsilon \in (0, 1/10)$ *and* $i \in \{1, ..., N(\varepsilon)\}$

$$\frac{\varepsilon}{N(\varepsilon)} \leq \eta_i(\varepsilon) \leq \varepsilon$$

- *[SSV18, Lemma 3.3] For a bounded sequence* $(a_i)_{i\in\mathbb{N}}$ *and* $\varepsilon \in (0, 1/10)$*, the inequality*

$$\left|\sum_{i=1}^{N(\varepsilon)} a_i \eta_i(\varepsilon)\right| \leq A\sqrt{\varepsilon} + A_\varepsilon \tag{8.3}$$

with $A := \sup\{|a_i| \mid i \in \mathbb{N}\}$ *and* $A_\varepsilon := \sup\{|a_i| \mid i \in \mathbb{N}, i \geq \varepsilon^{-1/2}\}$ *holds.*

To achieve good bounds it is necessary to choose a suitable subsequence of the nested Følner sequence $(Q_n)_{n\in\mathbb{N}}$ in Theorem 8.3, namely we want the following requirements also used in [SSV18] and partially in [PS16]:

Remark 8.5 (Requirements on the tiling Følner sequence). With the correlation length r from Definition 4.1, the sequences

$$\left(\frac{b(Q_n)}{|Q_n|}\right)_{n\in\mathbb{N}}, \left(\frac{b(Q_n^r)}{|Q_n|}\right)_{n\in\mathbb{N}}, \left(\frac{|\partial^r(Q_n)|}{|Q_n|}\right)_{n\in\mathbb{N}}$$

are required to converge *monotonically* to 0. All three sequences converge to 0 for any Følner sequence, but the monotonicity may only be true for a subsequence.
We define

$$\beta_n' := \max\left\{\left(\frac{b(Q_n)}{|Q_n|}\right), \left(\frac{b(Q_n^r)}{|Q_n|}\right), \left(\frac{|\partial^r(Q_n)|}{|Q_n|}\right)\right\}$$

and

$$\beta(\varepsilon) := \beta_1'\sqrt{\varepsilon} + \beta_{\lceil 1/\sqrt{\varepsilon}\rceil}'$$

for $n \in \mathbb{N}$ and $\varepsilon \in (0, 1/10)$, such that $(\beta'_n)_{n\in\mathbb{N}}$ is a monotone sequence converging to 0 and by (8.3) of Lemma 8.4

$$\sum_{i=1}^{N(\varepsilon)} \beta'_i \eta_i(\varepsilon) \leq \beta(\varepsilon) \xrightarrow[\varepsilon \searrow 0]{} 0.$$

We also require $\beta'_n \leq \frac{1}{2n}$, which is again achievable by choosing a suitable subsequence of any given Følner sequence. This requirement ensures that

$$\beta(\varepsilon) = \beta'_1 \sqrt{\varepsilon} + \beta'_{\lceil 1/\sqrt{\varepsilon} \rceil} \leq \frac{1}{2}\sqrt{\varepsilon} + \frac{1}{2\lceil 1/\sqrt{\varepsilon} \rceil} \leq \sqrt{\varepsilon}.$$

Recall

$$\mathbb{B} := \{F : \mathbb{R} \to \mathbb{R} : F \text{ right-continuous and bounded}\}.$$

With these results we can formulate the analogue of Lemma 5.1 for Cayley graphs of finitely generated amenable groups, which quantifies the uniform distance between a normalized admissible function on a large set and the averaged sum of this function on the ε-quasi tiling of the set.

Lemma 8.6 ([SSV18, Lemma 4.3])
Let G be a finitely generated amenable group, let $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be a generalized admissible function with a sub-additive boundary function as defined in Definition 5.6 and let $(\Lambda_n)_{n\in\mathbb{N}}$ and $(Q_n)_{n\in\mathbb{N}}$ be Følner sequences, where $(Q_n)_{n\in\mathbb{N}}$ is additionally nested and satisfies the requirements of Remark 8.5. Then for $j \geq j_0$ from Theorem 8.3 and $\varepsilon \in (0, 1/10)$

$$\left\| \frac{a(\Lambda_j, \omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon)t, \omega\right)}{|K_i(\varepsilon)|} \right\|_\infty$$
$$\leq (24D + 9E)\varepsilon + 12(2 + 2D + E)\beta(\varepsilon)$$

for all $\omega \in \Omega$, where $K_i(\varepsilon)$, $i \in \{1, ..., N(\varepsilon)\}$ are given by Theorem 8.3. Here $K_i^r = K_i \setminus \partial^r(K_i)$.

In comparison to Lemma 5.1, there are two additional problems that prevent the strategy used before from being immediately applicable. First there is not just one tiling, but multiple partial tilings. Additionally there is some overlap between the tiles of just one tiling, which destroys the independence that was used for the probabilistic part before.
The first problem can be solved by considering all tilings separately, the second is more

complicated. The solution found in [SSV18] is the resampling of overlapping tiles. For fixed $\varepsilon > 0$, $i \in \{1, ..., N(\varepsilon)\}$, $j \geq j_0(\varepsilon)$ and $t \in T_i^j(\varepsilon)$ the set

$$U^{i,j,t} := (K_i^r t) \setminus \left(K_i \left(T_i^j(\varepsilon) \setminus \{t\}\right)\right)$$

is the subset of the tile $K_i^r t$ that is not covered by some other part of the same tiling. For $t_1 \neq t_2$ there is a distance

$$\mathrm{d}_{\mathrm{set}}\left(U^{i,j,t_1}, U^{i,j,t_2}\right) \geq \mathrm{d}_{\mathrm{set}}\left(K_i^r t_1, G \setminus K_i t_1\right) > r$$

so the colors on those sets are independent. But since these partial tiles are no longer just translates of each other, the missing parts need to be resampled to gain identical distributions. The result is the following lemma that also quantifies the error made by this resampling.

Lemma 8.7 ([SSV18, Lemmas 5.2, 5.3])
Let $\varepsilon > 0$ and

$$I := \bigcup_{i=1}^{N(\varepsilon)} \bigcup_{j=j_0(\varepsilon)}^{\infty} \{(i,j)\} \times T_i^j(\varepsilon).$$

There exists a probability space $(\underline{\Omega}, \mathcal{B}(\underline{\Omega}), \underline{\mathbb{P}})$ and random variables

$$X, X^{i,j,t} : \underline{\Omega} \to \Omega, \ (i,j,t) \in I$$

such that for all $(i,j,t) \in I$:

1. *X and $X^{i,j,t}$ have distribution $\mathbb{P}$*
2. *X and $X^{i,j,t}$ agree on $U^{i,j,t}$ $\underline{\mathbb{P}}$-almost surely*
3. *the random variables in the set $\left\{X^{i,j,t}\right\}_{t \in T_i^j(\varepsilon)}$ are $\underline{\mathbb{P}}$-independent*

From this follows that in the setting of Lemma 8.6 for almost all $\underline{\omega} \in \underline{\Omega}$

$$\begin{aligned} &\left\| \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} a(K_i^r(\varepsilon)t, X(\underline{\omega})) - \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} a(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})) \right\|_\infty \\ &\qquad \leq 2b(K_i^r(\varepsilon)) + 2(3D+E)\varepsilon \left|K_i^r(\varepsilon)\right|. \end{aligned}$$

With the now independent random variables it is possible to apply the concentration inequalities for monotone increasing, right-continuous bounded function from Lemma 6.18 to get a result for the case of Cayley graphs of finitely generated amenable groups. This is achieved by integrating it into a proof in [SSV18].

Theorem 8.8 (based on Proposition 5.4 in [SSV18])
Let G be a finitely generated amenable group, $\mathcal{A} \in \mathcal{B}(\mathbb{R})$, $\Omega = \mathcal{A}^G$, $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space fulfilling (M1) and (M2) from Definition 4.1, $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be a generalized admissible function with a sub-additive boundary function as defined in Definition 5.6 and $0 \leq a(\Lambda, \omega) \leq |\Lambda|$ for all $\Lambda \in \mathcal{C}$.
Let $(\Lambda_n)_{n\in\mathbb{N}}$ and $(Q_n)_{n\in\mathbb{N}}$ be Følner sequences, where $(Q_n)_{n\in\mathbb{N}}$ is nested and fulfils the requirements of Remark 8.5.
For $\varepsilon \in (0, 1/10)$ let $K_i(\varepsilon)$, $i \in \{1, ..., N(\varepsilon)\}$ and $j_0(\varepsilon)$ be given by Theorem 8.3.
Then for all $\kappa > 0$ and $2 \leq M$ there exists for all $j \geq j_0(\varepsilon)$ an event $\Omega_{M,j,\varepsilon,\kappa} \in \mathcal{B}(\Omega)$ such that

$$\mathbb{P}(\Omega_{M,j,\varepsilon,\kappa}) \geq 1 - 2N(\varepsilon)M \exp\left(-\sqrt{\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|}|\Lambda_j|}\frac{\kappa}{K_M}\right),$$

where $K_M = \left(\frac{56(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right)\sum_{q=0}^{\infty} 2^{-q}\sqrt{2 + 2q\log(2)}$.
For all $\omega \in \Omega_{M,j,\varepsilon,\kappa}$ we have

$$\begin{aligned}&\left\|\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a\left(K_i^r(\varepsilon)t,\omega\right)}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)}{|K_i(\varepsilon)|}\right\|_\infty \\ &\quad \leq 2\beta(\varepsilon) + 2(3D+E)\varepsilon + \kappa.\end{aligned}$$

Proof. For $\varepsilon \in (0, 1/10)$, $j \in \mathbb{N}$, $\omega \in \Omega$ the triangle inequality leads to

$$\begin{aligned}&\left\|\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a\left(K_i^r(\varepsilon)t,\omega\right)}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)}{|K_i(\varepsilon)|}\right\|_\infty \\
&\leq \sum_{i=1}^{N(\varepsilon)}\frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|}\left\|\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t,\omega\right) - \mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)\right\|_\infty \\
&\leq \inf_{\underline{\omega}\in X^{-1}(\omega)}\left(\sum_{i=1}^{N(\varepsilon)}\frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|}\left(\frac{1}{\left|T_i^j(\varepsilon)\right|}\left\|\sum_{t\in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t,X(\underline{\omega})\right) - \sum_{t\in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t,X^{i,j,t}(\underline{\omega})\right)\right\|_\infty\right.\right. \\
&\quad\left.\left. + \left\|\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t,X^{i,j,t}(\underline{\omega})\right) - \underline{\mathbb{E}}a\left(K_i^r(\varepsilon),X(\underline{\cdot})\right)\right\|_\infty\right)\right).\end{aligned}$$

Note that the expected value $\underline{\mathbb{E}}$ in the last line is taken with regard to the probability measure $\underline{\mathbb{P}}$ from the space $(\underline{\Omega}, \mathcal{B}(\underline{\Omega}), \underline{\mathbb{P}})$ of Lemma 8.7.

From Lemma 8.7 and the conditions of Remark 8.5 follows that there exists a set $B \in \mathcal{B}(\underline{\Omega})$ with $\underline{\mathbb{P}}(B) = 1$ such that for all $\underline{\omega} \in B$

$$\begin{aligned}
&\frac{1}{|K_i(\varepsilon)| \left|T_i^j(\varepsilon)\right|} \left\| \sum_{t \in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t, X(\underline{\omega})\right) - \sum_{t \in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})\right) \right\|_\infty \\
&\quad \le \frac{2b(K_i^r(\varepsilon)) + 2(3D+E)\varepsilon \left|K_i^r(\varepsilon)\right|}{|K_i^r(\varepsilon)|} \\
&\quad \le \frac{2b(Q_i^r(\varepsilon))}{|Q_i|} + 2(3D+E)\varepsilon
\end{aligned}$$

holds true. Together with Lemma 8.4 follows

$$\begin{aligned}
&\sum_{i=1}^{N(\varepsilon)} \frac{\eta_i(\varepsilon)}{|K_i(\varepsilon)|} \frac{1}{\left|T_i^j(\varepsilon)\right|} \left\| \sum_{t \in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t, X(\underline{\omega})\right) - \sum_{t \in T_i^j(\varepsilon)} a\left(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})\right) \right\|_\infty \\
&\quad \le \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{2b(Q_i^r(\varepsilon))}{|Q_i|} + \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) 2(3D+E)\varepsilon \\
&\quad \le 2\beta(\varepsilon) + 2(3D+E)\varepsilon.
\end{aligned} \tag{8.4}$$

The random variables $\left\{X^{i,j,t}\right\}_{t \in T_i^j(\varepsilon)}$ are independent, and thus the same is true for $\left\{\gamma_t \circ X^{i,j,t}\right\}_{t \in T_i^j(\varepsilon)}$. All $X^{i,j,t}$ have the same distribution $\mathbb{P}$ and $\mathbb{P}$ is translation invariant, so the random variables $\left\{\gamma_t \circ X^{i,j,t}\right\}_{t \in T_i^j(\varepsilon)}$ are identically distributed as well. As a result of the requirements of the statement,

$$\frac{a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} : (\Omega, \mathcal{B}(\Omega)) \to \mathbb{B}$$

is a monotone increasing, right-continuous bounded random function in $[0,1]$ apart from Ω not being finite dimensional, which is taken care of by locality of a. We can thus use Lemma 6.18 like in Theorem 7.1 to show that there is a set $A_{i,j,\varepsilon,\kappa} \in \mathcal{B}(\underline{\Omega})$ with

$$\underline{\mathbb{P}}(A_{i,j,\varepsilon,\kappa}) \ge 1 - M \exp\left(-\frac{\sqrt{\left|T_i^j(\varepsilon)\right|}\kappa}{K_M} \right)$$

and M and K_M as defined in Lemma 6.18 and

$$\begin{aligned}
&\left\| \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon), \gamma_t \circ X^{i,j,t}(\underline{\omega})\right)}{|K_i(\varepsilon)|} - \frac{\underline{\mathbb{E}} a\left(K_i^r(\varepsilon), X(\underline{\cdot})\right)}{|K_i(\varepsilon)|} \right\|_\infty \\
&= \left\| \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})\right)}{|K_i(\varepsilon)|} - \frac{\underline{\mathbb{E}} a\left(K_i^r(\varepsilon), X(\underline{\cdot})\right)}{|K_i(\varepsilon)|} \right\|_\infty \leq \kappa
\end{aligned}$$

for all $\underline{\omega} \in A_{i,j,\varepsilon,\kappa}$. This bound needs to hold for all $i \in \{1, ..., N(\varepsilon)\}$ at the same time and we need to stay in B as well, so we form

$$A_{j,\varepsilon,\kappa} := \bigcap_{i=1}^{N(\varepsilon)} A_{i,j,\varepsilon,\kappa} \cap B.$$

From (8.2) and Lemma 8.4 we gain with $\left|T_i^j(\varepsilon)\right| |K_i(\varepsilon)| \geq \left|T_i^j(\varepsilon) K_i(\varepsilon)\right|$ that

$$\left|T_i^j(\varepsilon)\right| \geq \left(\eta_i(\eta) - \frac{\varepsilon^2}{N(\varepsilon)}\right) \frac{|\Lambda_j|}{|K_i(\varepsilon)|} \geq \frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\,|K_i(\varepsilon)|} |\Lambda_j| \geq \frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|} |\Lambda_j|,$$

and thus

$$\begin{aligned}
\underline{\mathbb{P}}(A_{j,\varepsilon,\kappa}) &= 1 - \underline{\mathbb{P}}\left(\bigcup_{i=1}^{N(\varepsilon)} A_{i,j,\varepsilon,\kappa}^C \cup B^C\right) \geq 1 - \sum_{i=1}^{N(\varepsilon)} \underline{\mathbb{P}}(A_{i,j,\varepsilon,\kappa}^C) \\
&\geq 1 - \sum_{i=1}^{N(\varepsilon)} M \exp\left(-\frac{\sqrt{\left|T_i^j(\varepsilon)\right|}\kappa}{K_M}\right) \\
&\geq 1 - N(\varepsilon) M \exp\left(-\sqrt{\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|} |\Lambda_j|}\frac{\kappa}{K_M}\right).
\end{aligned}$$

The set $A_{j,\varepsilon,\kappa}$ lies in $\mathcal{B}(\underline{\Omega})$, but we seek a set in $\mathcal{B}(\Omega)$. A candidate is $X\left(A_{j,\varepsilon,\kappa}\right)$, since if $\omega \in X\left(A_{j,\varepsilon,\kappa}\right)$ then there is a $\underline{\omega} \in X^{-1}(\{\omega\}) \cap A_{j,\varepsilon,\kappa}$ and

$$\begin{aligned}
&\inf_{\underline{\omega} \in X^{-1}(\omega)} \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \left\| \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon)t, X^{i,j,t}(\underline{\omega})\right)}{|K_i(\varepsilon)|} - \frac{\underline{\mathbb{E}} a\left(K_i^r(\varepsilon), X(\underline{\cdot})\right)}{|K_i(\varepsilon)|} \right\|_\infty \\
&\leq \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon)\kappa \leq \kappa. \qquad (8.5)
\end{aligned}$$

However, $X(A_{j,\varepsilon,\kappa})$ is not guaranteed to be measurable, only analytic.
But as argued in [SSV18], there still exists at least a compact subset $\Omega_{M,j,\varepsilon,\kappa}$ of $X(A_{j,\varepsilon,\kappa})$ with

$$\mathbb{P}(\Omega_{M,j,\varepsilon,\kappa}) \geq 1 - 2N(\varepsilon)M\exp\left(-\sqrt{\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|}\left|\Lambda_j\right|}\frac{\kappa}{K_M}\right).$$

By combining (8.4) with (8.5) it follows that for all $\omega \in \Omega_{M,j,\varepsilon,\kappa}$ we have

$$\left\|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a\left(K_i^r(\varepsilon)t,\omega\right)}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)}{|K_i(\varepsilon)|}\right\|_\infty$$
$$\leq 2\beta(\varepsilon) + 2(3D+E)\varepsilon + \kappa$$

as claimed. □

The last needed step is a result analogous to Lemma 5.2 for a convergence of the expected values. There is a result like this in [SSV18] using further tilings of tiles, but a short proof is also available using Lemma 8.6.

Corollary 8.9
Let G be a finitely generated amenable group, $\mathcal{A} \in \mathcal{B}(\mathbb{R})$, $\Omega = \mathcal{A}^G$, $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space fulfilling (M1) and (M2) from Definition 4.1, let a be a generalized admissible function with a sub-additive boundary function as defined in Definition 5.6, $(\Lambda_n)_{n\in\mathbb{N}}$ and $(Q_n)_{n\in\mathbb{N}}$ Følner sequences, where $(Q_n)_{n\in\mathbb{N}}$ is nested and fulfils the requirements of Remark 8.5. The limit $a^*(x) := \lim_{n\to\infty}\frac{\mathbb{E}a(\Lambda_n,\cdot)(x)}{|\Lambda_n|}$ exists uniformly in $x \in \mathbb{R}$, and for all $\varepsilon \in (0, 1/10)$ the estimate

$$\left\|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)}{|K_i(\varepsilon)|} - a^*\right\| \leq (24D+9E)\varepsilon + 12(2+2D+E)\beta(\varepsilon) \tag{8.6}$$

holds true with $\eta_i(\varepsilon), \beta(\varepsilon), K_i(\varepsilon)$ as defined in Theorem 8.3 and Remark 8.5.

Proof. From Lemma 8.6 follows for $\varepsilon \in (0, 1/10)$, $j \geq j(\varepsilon)$ and all $\omega \in \Omega$

$$\sup_{x\in\mathbb{R}}\left|\left(\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a\left(K_i^r(\varepsilon)t,\omega\right)(x)}{|K_i(\varepsilon)|} - \frac{a(\Lambda_j,\omega)(x)}{|\Lambda_j|}\right)\right| \tag{8.7}$$
$$\leq (24D+9E)\varepsilon + 12(2+2D+E)\beta(\varepsilon). \tag{8.8}$$

Since a is admissible we can use the translation invariance (A1) (see Definition 4.3) of a and the translation invariance (M1) (see Definition 4.1) of the measure to get

$$\begin{aligned}\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)(x) &= \frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\mathbb{E}a(K_i^r(\varepsilon),\cdot)(x)\\ &= \frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\mathbb{E}\left(a(K_i^r(\varepsilon),\cdot)(x)\circ\gamma_t\right)\\ &= \frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\mathbb{E}a(K_i^r(\varepsilon)t,\cdot)(x)\end{aligned}$$

for all $x\in\mathbb{R}$.
Now first let $j\geq j(\varepsilon)$ and by (8.7) $\forall x\in\mathbb{R}$

$$\begin{aligned}&\left|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)(x)}{|K_i(\varepsilon)|}-\frac{\mathbb{E}a(\Lambda_j,\cdot)(x)}{|\Lambda_j|}\right|\\ &=\left|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{\mathbb{E}a(K_i^r(\varepsilon)t,\cdot)(x)}{|K_i(\varepsilon)|}-\frac{\mathbb{E}a(\Lambda_j,\cdot)(x)}{|\Lambda_j|}\right|\\ &=\left|\mathbb{E}\left(\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a(K_i^r(\varepsilon)t,\cdot)(x)}{|K_i(\varepsilon)|}-\frac{a(\Lambda_j,\cdot)(x)}{|\Lambda_j|}\right)\right|\\ &\leq\mathbb{E}\left|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{1}{\left|T_i^j(\varepsilon)\right|}\sum_{t\in T_i^j(\varepsilon)}\frac{a(K_i^r(\varepsilon)t,\cdot)(x)}{|K_i(\varepsilon)|}-\frac{a(\Lambda_j,\cdot)(x)}{|\Lambda_j|}\right|\\ &\leq(24D+9E)\varepsilon+12(2+2D+E)\beta(\varepsilon)\end{aligned}$$

holds.
Since this is true for all $x\in\mathbb{R}$ this also shows

$$\sup_{x\in\mathbb{R}}\left(\left|\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)(x)}{|K_i(\varepsilon)|}-\frac{\mathbb{E}a(\Lambda_j,\cdot)(x)}{|\Lambda_j|}\right|\right)\tag{8.9}$$

$$\leq(24D+9E)\varepsilon+12(2+2D+E)\beta(\varepsilon).\tag{8.10}$$

We will use this equation now to prove that $\sum_{i=1}^{N(\varepsilon)}\eta_i(\varepsilon)\frac{\mathbb{E}a\left(K_i^r(\varepsilon),\cdot\right)(x)}{|K_i(\varepsilon)|}$ satisfies the Cauchy property for $\varepsilon\to 0$.

For this let $0 < \delta < \varepsilon < 1/10$. Then choose a $j \geq \max(j(\delta), j(\varepsilon))$. Now

$$\begin{aligned}
&\left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\delta)} \eta_i(\delta) \frac{\mathbb{E}a\left(K_i^r(\delta), \cdot\right)}{|K_i(\delta)|} \right\|_\infty \\
&\qquad \leq \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} - \frac{\mathbb{E}a(\Lambda_j, \cdot)}{|\Lambda_j|} \right\|_\infty && (8.11) \\
&\qquad\quad + \left\| \frac{\mathbb{E}a(\Lambda_j, \cdot)}{|\Lambda_j|} - \sum_{i=1}^{N(\delta)} \eta_i(\delta) \frac{\mathbb{E}a\left(K_i^r(\delta), \cdot\right)}{|K_i(\delta)|} \right\|_\infty \\
&\qquad \leq (24D + 9E)(\varepsilon + \delta) + 12(2 + 2D + E)(\beta(\varepsilon) + \beta(\delta)) && (8.12) \\
&\qquad \leq 2\left((24D + 9E)\varepsilon + 12(2 + 2D + E)\beta(\varepsilon)\right) \to 0
\end{aligned}$$

since $\beta(\varepsilon)$ is monotone increasing in ε and $\beta(\varepsilon) \to 0$ if $\varepsilon \to 0$, so $\sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)(x)}{|K_i(\varepsilon)|}$ converges to a limit $\tilde{a} = \lim_{\varepsilon \to 0} \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|}$.

With the same idea, let $\varepsilon \in (0, 1/10)$ and $k, l \geq j(\varepsilon)$, so that

$$\begin{aligned}
\left\| \frac{\mathbb{E}a(\Lambda_k, \cdot)}{|\Lambda_k|} - \frac{\mathbb{E}a(\Lambda_l, \cdot)}{|\Lambda_l|} \right\|_\infty &\leq \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} - \frac{\mathbb{E}a(\Lambda_k, \cdot)}{|\Lambda_k|} \right\| \\
&\quad + \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} - \frac{\mathbb{E}a(\Lambda_l, \cdot)}{|\Lambda_l|} \right\| \\
&\leq 2\left((24D + 9E)\varepsilon + 12(2 + 2D + E)\beta(\varepsilon)\right) \to 0
\end{aligned}$$

which proves that $\left(\frac{\mathbb{E}a(\Lambda_n, \cdot)}{|\Lambda_n|}\right)_{n \in \mathbb{N}}$ is also a Cauchy sequence for $n \to \infty$ that converges uniformly to some $a^* := \lim_{n \to \infty} \frac{\mathbb{E}a(\Lambda_n, \cdot)}{|\Lambda_n|}$. Since (8.9) is true for all $j \geq j(\varepsilon)$ the two limits coincide, so $\tilde{a} = a^*$ and (8.6) follows by taking the limit $\delta \to 0$ in (8.11). ☐

Now we just need to combine the previous results to obtain

Theorem 8.10 (Concentration inequality for admissible functions on Cayley graphs of finitely generated amenable groups)
Let G be a finitely generated amenable group, $\mathcal{A} \in \mathcal{B}(\mathbb{R})$, $\Omega = \mathcal{A}^G$, $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ a probability space satisfying (M1) and (M2) from Definition 4.1, $a : \mathcal{C} \times \Omega \to \mathbb{B}$ be a generalized admissible function with a sub-additive boundary function as defined in Definition 5.6 and $0 \leq a(\Lambda, \omega) \leq |\Lambda|$ for all $\Lambda \in \mathcal{C}$.
Let $(\Lambda_n)_{n \in \mathbb{N}}$ and $(Q_n)_{n \in \mathbb{N}}$ be Følner sequences, where $(Q_n)_{n \in \mathbb{N}}$ is nested and fulfils the

requirements of Remark 8.5.
For $\varepsilon \in (0, 1/10)$ let $K_i(\varepsilon)$, $i \in \{1, ..., N(\varepsilon)\}$ and $j_0(\varepsilon)$ be given by Theorem 8.3. Then there exists an $a^* \in \mathbb{B}$ such that for all $\kappa > 0$ and $2 \le M$ there exists for all $j \ge j_0(\varepsilon)$ an event $\Omega_{M,j,\varepsilon,\kappa} \in \mathcal{B}(\Omega)$ such that

$$\mathbb{P}(\Omega_{M,j,\varepsilon,\kappa}) \ge 1 - 2N(\varepsilon) M \exp\left(-\sqrt{\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|} |\Lambda_j|}\frac{\kappa}{K_M} \right),$$

where $K_M = \left(\frac{56(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)} \right) \sum_{q=0}^{\infty} 2^{-q} \sqrt{2 + 2q\log(2)}$, and for all $\omega \in \Omega_{M,j,\varepsilon,\kappa}$ we have

$$\left\| \frac{a(\Lambda_j, \omega)}{|\Lambda_l|} - a^* \right\|_\infty \le (54D + 20E)\varepsilon + (50 + 48D + 24E)\beta(\varepsilon) + \kappa \tag{8.13}$$

$$\le (50 + 102D + 44E)\sqrt{\varepsilon} + \kappa. \tag{8.14}$$

Proof. With $K_i(\varepsilon)$, $i \in \{1, ..., N(\varepsilon)\}$, $T_i^j(\varepsilon)$, $i \in \{1, ..., N(\varepsilon)\}$ and $j_0(\varepsilon)$ given by Theorem 8.3 we use the triangle inequality for

$$\begin{aligned}\left\| \frac{a(\Lambda_j, \omega)}{|\Lambda_l|} - a^* \right\|_\infty \le & \left\| \frac{a(\Lambda_j, \omega)}{|\Lambda_j|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon)t, \omega\right)}{|K_i(\varepsilon)|} \right\|_\infty \\ & + \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{1}{\left|T_i^j(\varepsilon)\right|} \sum_{t \in T_i^j(\varepsilon)} \frac{a\left(K_i^r(\varepsilon)t, \omega\right)}{|K_i(\varepsilon)|} - \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} \right\|_\infty \\ & + \left\| \sum_{i=1}^{N(\varepsilon)} \eta_i(\varepsilon) \frac{\mathbb{E}a\left(K_i^r(\varepsilon), \cdot\right)}{|K_i(\varepsilon)|} - a^* \right\|\end{aligned}$$

and apply Lemma 8.6 to the first, Theorem 8.8 to the second and Corollary 8.9 to the third term on the right to get the event $\Omega_{M,j,\varepsilon,\kappa} \in \mathcal{B}(\Omega)$ as claimed and the bound

$$\begin{aligned}\left\| \frac{a(\Lambda_j, \omega)}{|\Lambda_l|} - a^* \right\|_\infty \le & (24D + 9E)\varepsilon + 12(2 + 2D + E)\beta(\varepsilon) \\ & + 2\beta(\varepsilon) + 2(3D + E)\varepsilon + \kappa \\ & + (24D + 9E)\varepsilon + 12(2 + 2D + E)\beta(\varepsilon) \\ \le & (54D + 20E)\varepsilon + (50 + 48D + 24E)\beta(\varepsilon) + \kappa\end{aligned}$$

for all $\omega \in \Omega_{M,j,\varepsilon,\kappa}$. The inequalities $\varepsilon \le \sqrt{\varepsilon}$ and $\beta(\varepsilon) \le \sqrt{\varepsilon}$ from Remark 8.5 finish the claimed statement. □

Remark 8.11. The concentration inequality given here is of the sub-root-exponential-type, however we can of course also formulate a matching sub-exponential-type inequality. By using the same proofs with inequality (6.22) in Lemma 6.18 instead of (6.20) we get that in the setting of Theorem 8.10 with the additional condition that $\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|}|\Lambda_j| > \left(12\frac{K_2}{\kappa^2}\right)^2$ where

$$K_2 = \left(\frac{168}{\log(3/2)} + \frac{4}{\log(2)}\right)\sum_{q=0}^{\infty} 2^{-q}\sqrt{2+2q\log(2)},$$

there is a set $\Omega'_{j,\varepsilon,\kappa}$ with

$$\mathbb{P}(\Omega'_{j,\varepsilon,\kappa}) \geq 1 - 2N(\varepsilon)\exp\left(-\frac{\kappa^2}{24}\frac{(1-\varepsilon)\varepsilon}{N(\varepsilon)\left|K_{N(\varepsilon)}(\varepsilon)\right|}|\Lambda_j|\right)$$

such that (8.13) holds for all $\omega \in \Omega'_{j,\varepsilon,\kappa}$.

8.2 Long-range percolation Hamiltonians

Another extension of Theorem 7.1 allows for unbound hopping range operators in the context of long range percolation. Instead of using a nearest neighbour Laplace operator combined with a random potential at every point in $\mathbb{Z}^d$ we will construct a graph by randomly connecting any two points in $\mathbb{Z}^d$ and then apply a Laplace operator on the resulting graph. We will see that it is nevertheless possible to show a concentration inequality for the eigenvalue-counting function of this Laplace operator. To implement long-range percolation in our framework, we need to ensure that the needed properties are fulfilled. For this we follow [Sch12b].

8.2.1 The probability space of long-range percolation

We will only consider the group $\mathbb{Z}^d$ here, but the procedure easily generalises to finitely generated groups with a finite and symmetric set of generators and a monotiling Følner sequence. As in Chapter 4 let $\mathcal{C}$ denote the set of all finite subsets of $\mathbb{Z}^d$. Let $\Gamma = (V, E)$ be the graph consisting of the vertex set $V = \mathbb{Z}^d$ and the edge set E consisting of all unordered pairs of vertices $x, y \in V$, denoted by $e = [x, y]$. Our sample space is $\Omega = \{0,1\}^E$, so for $\omega \in \Omega$ each edge $e \in E$ between two points is randomly either activated (if $\omega_e = 1$) or deactivated (if $\omega_e = 0$). Each $\omega \in \Omega$ then gives rise to a subgraph $\Gamma_\omega = (V, E_\omega)$, where E_ω contains only the activated edges. A σ-algebra $\mathcal{B}(\Omega)$ for this space can be constructed using cylinder sets. Now a probability space for this measurable space can be constructed from any sequence $p = (p(x))_{x\in\mathbb{Z}^d} \in \ell^1(\mathbb{Z}^d)$ that fulfils $0 \leq p(x) \leq 1$ and $p(x) = p(-x)$ for all $x \in \mathbb{Z}^d$. The way to do this is by

choosing this as the probability measure for all links between 0 and all other points in $\mathbb{Z}^d$. Then this probability function can be extended to all other marginals by mapping each edge to an edge connecting to 0 in the following way:
Let $e = [x, y] \in E$ and define first for $\omega \in \Omega$ a probability measure $\mathbb{P}_e$ on $\{0, 1\}$ for ω_e by

$$\mathbb{P}_e(\omega_e = 1) = p(x - y) \text{ and } \mathbb{P}_e(\omega_e = 0) = 1 - p(x - y)$$

where $\omega \in \Omega$. From these measures we can construct a product measure $\mathbb{P} = \prod_{e \in E} \mathbb{P}_e$ on Ω.
Since $\mathbb{Z}^d$ is finitely generated with a symmetric set of generators we can use the graph metric d as in Definition 8.1.
Here we will construct some random variables for later use. First we define the set E^R of edges of length equal or less than a given number R, i.e.

$$E^R := \{e = [x, y] \in E \mid d(x, y) \leq R\}$$

and let p^R be the projection from Ω to $\Omega^R = \{0, 1\}^{E^R}$. For a finite subset $Q \in \mathcal{C}$ we further define the space $E^R_Q := \left\{e = [x, y] \in E^R \mid x, y \in Q\right\}$ of edges of length equal or smaller than R that connect vertices in Q, with an associated projection Π^R_Q from Ω^R to $\Omega^R_Q = \{0, 1\}^{E^R_Q}$.
In combination we get a projection

$$X^R_Q \colon \Omega \to \Omega^R_Q, \ X^R_Q := \Pi^R_Q \circ p^R.$$

As projections these maps are measurable and since E^R_Q is finite the induced σ-algebra on Ω^R_Q by $X^R_Q = \Pi^R_Q \circ p^R$ is just the power set of Ω^R_Q. We can also define a subgraph $\Gamma_{X^R_Q(\omega)} = (Q, E^R_{Q,\omega})$ with

$$E^R_{Q,\omega} = \{e \in E^R_Q \mid \left(X^R_Q(\omega)\right)_e = 1\}.$$

Thus, $\Gamma_{X^R_Q(\omega)}$ is the graph with vertex set Q and the edges between them that are contained in E_ω and have a length of at most R. We need to check now if we can verify the assumptions on $\mathbb{P}$ required by Definition 4.1.

Lemma 8.12
The probability space $(\Omega, \mathcal{B}(\Omega), \mathbb{P})$ *defined above fulfils:*

1. $\mathbb{P} = \mathbb{P} \circ \gamma_z$ *for all translations* γ_z*, defined as*

$$\gamma_z \colon \Omega \to \Omega, \ (\gamma_z \omega)_{[x,y]} := \omega_{[x+z, y+z]}, \ x, y, z \in \mathbb{Z}^d$$

2. *For all* $n \in \mathbb{N}$ *and* $Q_1, ..., Q_n \in \mathcal{C}$ *with* $|Q_i| \neq 0$ *for all* $1 \leq i \leq n$ *and* $\min\{\mathrm{d}_{set}(Q_i, Q_j) \mid i \neq j\} > 0$ *we have that the projections* $\left(X^R_{Q_i}\right)_{1 \leq i \leq n}$ *are independent.*

Proof. 1. We have

$$\mathbb{P}((\gamma_z\omega)_{[x,y]} = 1) = \mathbb{P}(\omega_{[x+z,y+z]} = 1) = p\left((x+z) - (y+z)\right) \\ = p(x-y) = \mathbb{P}(\omega_{[x,y]} = 1)$$

and therefore $\mathbb{P} \circ \gamma_z = \mathbb{P}$.

2. All ω_e are independent by construction, and if $Q_1, ..., Q_n \in \mathcal{C}$ with $|Q_i| \neq 0$ for all $1 \leq i \leq n$ and $\min\{\mathrm{d}_{\mathrm{set}}(Q_i, Q_j) \mid i \neq j\} > 0$ then there is no edge that will be part of more than one of the edge sets $E^R_{Q_i}$, $1 \leq i \leq n$. Since there are no shared edges, the projections $\left(X^R_{Q_i}\right)_{1\leq i\leq n}$ are independent.

□

The next step would be to find an admissible function as defined in Definition 4.3, but this will not be possible on the whole space Ω. First we need to exclude all random graphs with vertices with infinitely many neighbours, where a neighbour of a vertex x is any vertex y where $[x, y] \in E_\omega$. With our chosen probability measure those graphs have zero measure, which can be shown using the Borel-Cantelli-Lemma 2.9:
Define for each $x \in \mathbb{Z}^d$ the vertex degree of x in Γ_ω by

$$m_\omega(x) := \left|\{y \in \mathbb{Z}^d \mid [x, y] \in E_\omega\}\right| \in [0, \infty]. \tag{8.15}$$

Now for $x, y \in \mathbb{Z}^d$ define the event $A_{[x,y]} := \{[x, y] \in E_\omega\}$ that x and y are linked. Then

$$\sum_{y\in\mathbb{Z}^d} \mathbb{P}(A_{[x,y]}) = \sum_{y\in\mathbb{Z}^d} p(x-y) < \infty$$

since $p \in \ell^1(\mathbb{Z}^d)$. The Borel-Cantelli-Lemma then guarantees the existence of a set $\Omega_x \subset \Omega$ of full measure where $m_\omega(x)$ is finite for each $\omega \in \Omega_x$, and by

$$\mathbb{P}(\{\exists x \in \mathbb{Z}^d \ : m_\omega(x) = \infty\}) = \mathbb{P}\left(\bigcup_{x\in\mathbb{Z}^d} \{m_\omega(x) = \infty\}\right) \\ \leq \sum_{x\in\mathbb{Z}^d} \mathbb{P}\left(\{m_\omega(x) = \infty\}\right) \\ \leq \sum_{x\in\mathbb{Z}^d} \mathbb{P}(\Omega \setminus \Omega_x) = 0$$

it follows that this is also almost surely true for all $x \in \mathbb{Z}^d$ at the same time since $\mathbb{Z}^d$ is countable. We will call this subset

$$\Omega_{lf} = \bigcap_{x\in\mathbb{Z}^d} \Omega_x$$

and the graph Γ_ω is locally finite for every $\omega \in \Omega_{lf}$.
However, to continue we need to restrict the probability space further by only considering events where there are not too many long edges.
With the metric defined in Section 4.1 we can define again for $R \in \mathbb{N}$ an R-boundary for any given set $Q \subset \mathbb{Z}^d$ as

$$\partial^R Q := \Big\{ x \in Q \mid \exists h \in \mathbb{Z}^d \setminus Q : d(x,h) \leq R \Big\} \cup \Big\{ x \in \mathbb{Z}^d \setminus Q \mid \exists h \in Q : d(x,h) \leq R \Big\}$$

and the R-interior of Q, i.e. the remainder of Q without the R-boundary as

$$Q^R := Q \setminus \partial^R Q.$$

For a fixed $R \in \mathbb{N}$ and a finite subset $Q = \{x_1, ..., x_{|Q|}\} \in \mathcal{C}$ the random variables $Y_i, i = 1, ..., |Q|$ with

$$Y_i(\omega) := \sum_{y \in M_i^R} \omega_{[x_i, y]}$$

where

$$\begin{aligned} M_i^R &:= \{x \in \mathbb{Z}^d \mid d(x, x_i) > R, x \neq x_j \forall j < i\} \\ &= B_R(x_i)^C \setminus \{x_1, ..., x_{i-1}\} \end{aligned}$$

count all edges incident to vertices in Q with length greater than R without double counting edges between vertices in Q so that all Y_i are independent of each other. In [Sch12b] it was shown that with

$$c := \prod_{y \in \mathbb{Z}^d} (1 + p(y)(\mathrm{e} - 1)) \tag{8.16}$$

the estimate

$$\mathbb{P}(Y_i \geq t) \leq c\mathrm{e}^{-t}$$

holds for all $t \in \mathbb{N}$ and all $i \in \{1, ..., |Q|\}$. This can now be used to apply a Bernstein inequality. First choose $T \in \mathbb{N}$ large enough for

$$\sum_{t=T+1}^{\infty} t^2 \mathrm{e}^{-t} \leq \frac{1}{3c}$$

with c as in (8.16). Now define the monotonely decreasing function

$$\varepsilon\colon \mathbb{R} \to \mathbb{R},\ \varepsilon(R) := \mathbb{E}(Y_1) = \sum_{y \in \mathbb{Z}^d \setminus B_R} p(y) \tag{8.17}$$

and let $R_0 \in \mathbb{N}$ be such that for all $R \geq R_0$

$$\varepsilon(R) \leq \varepsilon(R_0) \leq -\frac{1}{2} \log \left(1 - \left(3 \sum_{t=1}^{T} t^2 \right)^{-1} \right).$$

Since $\sum_{t=1}^{T} t^2 \geq 1$, this implies $\varepsilon(R) \leq \frac{1}{3}$ for all $R \geq R_0$ and therefore also $p(y) \leq \frac{1}{2}$ for all $y \in \mathbb{Z}^d \setminus B_{R_0}$. Based on this the following Lemma was proved in [Sch12b].

Lemma 8.13 ([Sch12b, Corollary 3.3])
Let R_0 and c be as defined above, $\delta > 0$ and $Q \in \mathcal{C}$. Then for all $R \geq R_0$

$$\mathbb{P}(\Omega_1(\delta, R, Q)) \leq \begin{cases} \exp(-\frac{\delta^2 |Q|}{4}), & 0 \leq \delta \leq \frac{1}{6c}, \\ \exp(-\frac{\delta |Q|}{24c}), & \delta > \frac{1}{6c}, \end{cases}$$

where

$$\Omega_1(\delta, R, Q)) := \left\{ \omega \in \Omega \mid \sum_{i=1}^{|Q|} Y_i(\omega) \geq |Q| \left(\varepsilon(R) + \delta \right) \right\}$$

is the set of all $\omega \in \Omega$ so that there are at least $|Q| (\varepsilon(R) + \delta)$ edges in E_ω connecting to vertices in Q that are longer than R.

In the following we will also define

$$\Omega_2(\delta, R, Q)) = \Omega \setminus \Omega_1(\delta, R, Q))$$

8.2.2 The Laplace operator on a percolation graph

Now we can take the next step and define the operator we want to consider, the Laplace operator Δ_ω on our random graph as defined in [Sch12b]. To make sure the operator is well defined we will only consider $\omega \in \Omega_{lf}$, random graphs which are locally finite. As was shown before, this set has measure one.
Let $\ell_0(\mathbb{Z}^d) \subset \ell^2(\mathbb{Z}^d)$ be the dense subset of complex-valued functions on $\mathbb{Z}^d$ with finite support and define the operator $\tilde{\Delta}_\omega \colon \ell_0(\mathbb{Z}^d) \to \ell^2(\mathbb{Z}^d)$ by

$$\tilde{\Delta}_\omega \varphi(x) := m_\omega(x)\varphi(x) - \sum_{y:[x,y] \in E_\omega} \varphi(y) = \sum_{y:[x,y] \in E_\omega} (\varphi(x) - \varphi(y)).$$

This operator is known to be essentially selfadjoint [Web10, Theorem 2.1], so there is a domain D_ω such that $\Delta_\omega : D_\omega \to \ell^2(\mathbb{Z}^d)$ is the unique selfadjoint extension of $\tilde{\Delta}_\omega$.

For a given subgraph $S = (V_S, E_S)$ of the complete graph $\Gamma = (V, E)$ where V_S is finite we similarly define an operator $\Delta_S \colon \ell^2(V_S) \to \ell^2(V_S)$ via

$$\Delta_S \varphi(x) := \sum_{y \in V_S : [x,y] \in E_S} (\varphi(x) - \varphi(y)).$$

Let $\mathcal{S}$ be the set of all subgraphs of the complete graph Γ with a finite vertex set and $\mathcal{S}(Q)$ the subset of $\mathcal{S}$ consisting of all subgraphs with a fixed finite vertex set Q. If we already have a subgraph $S = (V_S, E_S)$ of Γ and a set $Q \subset V_S$ we further construct a subgraph $S[Q]$ of S on Q by setting as the edge set all edges in E_S that connect points in Q.

Furthermore we now define restrictions of Δ_ω and Δ_S to subsets by projections and inclusions analogous to Definition 3.24. The space $\ell^2(Q)$ can be identified with a subspace of $\ell^2(\mathbb{Z}^d)$ consisting of all functions whose support is a subset of Q. Then we project on this subspace via $p_Q \colon \ell^2(\mathbb{Z}^d) \to \ell^2(Q)$ with $p_Q(\varphi)(x) = \varphi(x)$ for $x \in Q$. We also define the inclusion $i_Q \colon \ell^2(Q) \to \ell^2(\mathbb{Z}^d)$ by

$$i_Q(u)(x) := \begin{cases} u(x) \text{ if } x \in Q, \\ 0 \text{ else} \end{cases} .$$

For finite $Q \subset \mathbb{Z}^d$ and $U \subset V_S$ we now form the restricted operators

$$\Delta_\omega[Q] := p_Q \Delta_\omega i_Q$$
$$\Delta_S[U] := p_U i_{V_S} \Delta_S p_{V_S} i_U.$$

These restricted operators are symmetric matrices with real entries and therefore hermitian. As such they have real eigenvalues and the number of eigenvalues is $|Q|$ in the first and $|U|$ in the second case if we take into account multiplicities.

What we are interested in are specifically the eigenvalue-counting function (evcf) of our operators, defined analogous to Definition 3.25 but this time we need to be more specific about the operator whose eigenvalues we are counting.

Definition 8.2.1. Recall the Banach space

$$\mathbb{B} := \{F : \mathbb{R} \to \mathbb{R} : F \text{ right-continuous and bounded}\}.$$

The eigenvalue counting function (evcf) $n(A) \colon \mathbb{R} \to \mathbb{R}$ of a selfadjoint operator A on a finite dimensional Hilbert space V is

$$n(A)(x) := |\{i \in \mathbb{N} \mid \lambda_i \leq x\}|$$

for all $x \in \mathbb{R}$, where $\lambda_i,\ i \in 1, ..., \dim V$ are the eigenvalues of A counted with multiplicities.

The evcf has some useful properties that we already used in Lemma 4.5:

Lemma 8.14 ([LSV10, Proposition 7.1])
Let A and C be selfadjoint operators in a finite dimensional Hilbert space. Then

$$|n(A)(x) - n(A+C)(x)| \leq \operatorname{rank}(C)$$

for all $x \in \mathbb{R}$.

Lemma 8.15 ([LSV10, Proposition 7.2])
Let V be a finite dimensional Hilbert space and U a subspace of V. Then

$$|n(A)(x) - n(pAi)(x)| \leq 4\operatorname{rank}(\mathrm{id}_V - i \circ p)$$

for all selfadjoint operators A on V and all $x \in \mathbb{R}$ where $i : U \to V$ is the inclusion and $p : V \to U$ is the orthogonal projection.

8.2.3 Constructing the finite range operator

We will now investigate the influence of long edges on the evcf. We want to apply the concentration inequalities of Lemma 6.18 later, but we need to find suitable independent random variables. Since the graph might contain edges of arbitrary length, we cannot isolate the Laplace operator on some finite subgraph of Γ from the rest of the graph. Thus, we need to modify the Laplace operator to remove these edges and quantify the disturbance this causes. We follow [Sch12b] in part, but will construct a slightly different operator. First we compare for $\omega \in \Omega_{lf}$, $Q \in \mathcal{C}$ and $R \in \mathbb{N}$ the operator $\Delta_\omega[Q^R]$, the Laplace operator restricted to a finite subset Q^R, with the operator $\Delta_{\Gamma_\omega[Q]}[Q^R]$, which is the Laplace operator restricted to the sub*graph* $\Gamma_\omega[Q]$ and then restricted to Q^R. Both operators are defined on $\ell^2(Q^R)$. Since the embedding of any function in $\ell^2(Q^R)$ into $\ell^2(\mathbb{Z}^d)$ vanishes outside of Q^R, these functions can be written as

$$\begin{aligned}
\Delta_\omega[Q^R]\varphi(z) &= m_\omega(z)\varphi(z) - \sum_{y \in Q^R : [z,y] \in E_\omega} \varphi(y) \\
\Delta_{\Gamma_\omega[Q]}[Q^R]\varphi(z) &= \tilde{m}_\omega(z)\varphi(z) - \sum_{y \in Q^R : [z,y] \in E_\omega} \varphi(y)
\end{aligned}$$

where $m_\omega(z) = \left|\{y \in \mathbb{Z}^d \mid [z,y] \in E_\omega\}\right|$ as defined in (8.15) and $\tilde{m}_\omega(z) = |\{y \in Q \mid [z,y] \in E_\omega\}|$. This means that $m_\omega(z)$ counts the number of edges that end in z if we look at every possible edge to points in $\mathbb{Z}^d$ and $\tilde{m}_\omega(z)$ only counts edges to other points in Q. Thus, the difference between $\Delta_\omega[Q^R]$ and $\Delta_{\Gamma_\omega[Q]}[Q^R]$ is the multiplication operator $D_\omega^{R,1}[Q] : \ell^2(Q^R) \to \ell^2(Q^R)$ with

$$(D_\omega^{R,1}[Q]\varphi)(z) := \left|\left\{y \in \mathbb{Z}^d \setminus Q \mid [z,y] \in E_\omega\right\}\right| \varphi(z).$$

We further define an operator $D_\omega^{R,2}[Q] : \ell^2(Q^R) \to \ell^2(Q^R)$ as

$$(D_\omega^{R,2}[Q]\varphi)(z) := |\{y \in Q \mid d(z,y) > R, [z,y] \in E_\omega\}| \, \varphi(z).$$

This operator multiplies every function value at a point z with the number of edges of length greater than R between z and other points in Q.
The sum of both multiplication operators is

$$\left((D_\omega^{R,1}[Q] + D_\omega^{R,2}[Q])\varphi\right)(z) = \left|\left\{y \in \mathbb{Z}^d \mid d(z,y) > R, [z,y] \in E_\omega\right\}\right| \varphi(z)$$

since $z \in Q^R$, and therefore any edge connecting to $\mathbb{Z}^d \setminus Q$ has a length greater than R. This is a diagonal matrix operator whose number of non-zero entries is limited by the number of edges of length greater than R that are connected to a point in Q^R. Thus,

$$\operatorname{rank}\left(D_\omega^{R,1}[Q] + D_\omega^{R,2}[Q]\right) \leq |Q| \left(\varepsilon(R) + \delta\right)$$

if $\omega \in \Omega_2(\delta, R, Q))$. By Lemma 8.14 this means

$$\left\|n(\Delta_\omega[Q^R]) - n\left(\Delta_\omega[Q^R] - D_\omega^{R,1}[Q] - D_\omega^{R,2}[Q]\right)\right\|_\infty \leq |Q| \left(\varepsilon(R) + \delta\right) \tag{8.18}$$

if $\omega \in \Omega_2(\delta, R, Q))$.
At last we define for $\omega \in \Omega_{lf}$, $Q \in \mathcal{C}$ and $R \in \mathbb{N}$ an operator $L_\omega[Q] : \ell^2(Q) \to \ell^2(Q)$ as

$$(L_\omega[Q]\varphi)(z) := \sum_{y \in Q : [z,y] \in E_\omega, d(z,y) > R} \varphi(y).$$

This operator contains the off-diagonal influence of edges between points in Q that are longer than R. If $\omega \in \Omega_2(\delta, R, Q))$ there are at most $|Q| (\varepsilon + \delta)$ edges of length bigger than R that connect to points in Q, so the matrices $L_\omega[Q]$ and $L_\omega[Q^R]$ contain at most $2 |Q| (\varepsilon + \delta)$ non-zero elements and thus

$$\begin{aligned} \operatorname{rank}(L_\omega[Q]) &\leq 2 |Q| \left(\varepsilon(R) + \delta\right) \\ \operatorname{rank}(L_\omega[Q^R]) &\leq 2 |Q| \left(\varepsilon(R) + \delta\right). \end{aligned} \tag{8.19}$$

We can now form the **finite range Laplace operator** $\Delta^R_{\Gamma_\omega[Q]}[Q^R] : \ell^2(Q^R) \to \ell^2(Q^R)$ with

$$\Delta^R_{\Gamma_\omega[Q]}[Q^R] := \Delta_{\Gamma_\omega[Q]}[Q^R] + L_\omega[Q^R] - D_\omega^{R,2}[Q] = \Delta_\omega[Q^R] - D_\omega^{R,1}[Q] + L_\omega[Q^R] - D_\omega^{R,2}[Q]$$

which does not depend on any edge longer than R at all. We can write this operator explicitly as

$$\Delta^R_{\Gamma_\omega[Q]}[Q^R]\varphi(z) = m'_\omega(z)\varphi(z) - \sum_{y \in Q^R : [z,y] \in E_\omega, d(z,y) \leq R} \varphi(y)$$

with $m'_\omega(z) = \left|\left\{y \in \mathbb{Z}^d \mid d(z,y) \le R, [z,y] \in E_\omega\right\}\right|$ Furthermore there is also again a restricted operator

$$\Delta^R_{\Gamma_\omega[Q]}[U] = p_U \, i_{Q^R} \, \Delta^R_{\Gamma_\omega[Q]}[Q^R] \, p_{Q^R} \, i_U$$

for subsets $U \subset Q^R$.
Lemma 8.14 in combination with (8.18) and (8.19) now leads to

$$\begin{aligned}
&\left\| n\left(\Delta_\omega[Q^R]\right) - n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right) \right\|_\infty \\
&\quad \le \left\| n\left(\Delta_\omega[Q^R]\right) - n\left(\Delta_\omega[Q^R] - D^{R,1}_\omega[Q] - D^{R,2}_\omega[Q]\right) \right\|_\infty \\
&\qquad + \left\| n\left(\Delta_\omega[Q^R] - D^{R,1}_\omega[Q] - D^{R,2}_\omega[Q]\right) - n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right) \right\|_\infty \\
&\quad \le |Q| \left(\varepsilon(R) + \delta\right) + 2\,|Q| \left(\varepsilon(R) + \delta\right) = 3\,|Q| \left(\varepsilon(R) + \delta\right)
\end{aligned} \tag{8.20}$$

if $\omega \in \Omega_2(\delta, R, Q)$).

8.2.4 Admissible percolation functions

Next we want to establish a uniform convergence for $n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)$. To continue with the strategy we used for the eigenvalue-functions of Anderson operators, we need to ensure that we still have all necessary properties of admissible functions from Definition 4.3. However, since the probability space is slightly different, we get

Lemma 8.16
The eigenvalue-counting function $n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)$ of the finite range Laplace operator on the percolation sub-graph $\Gamma_\omega[Q]$ has the following properties:

(AP1) translation invariance: For $Q \in \mathcal{C}$ such that $Q^R \neq \emptyset$, $z \in \mathbb{Z}^d$ and $\omega \in \Omega$ we have

$$n\left(\Delta^R_{\Gamma_\omega[Q+z]}[Q^R + z]\right) = n\left(\Delta^R_{\Gamma_{(\gamma_z \omega)}[Q]}[Q^R]\right).$$

(AP2) locality: For all $Q \in \mathcal{C}$ such that $Q^R \neq \emptyset$ and $\omega, \omega' \in \Omega$ with $X^R_Q(\omega) = X^R_Q(\omega')$ we have

$$n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right) = n\left(\Delta^R_{\Gamma_{\omega'}[Q]}[Q^R]\right).$$

(AP3) almost additivity: For each $\omega \in \Omega$, pairwise disjoint $Q_1, ..., Q_n \in \mathcal{C}$ such that $Q_i^R \neq \emptyset$ for $1 \le i \le n$ and $Q := \bigcup_{i=1}^n Q_i$ we have

$$\left\| n\left(\Delta^R_{\Gamma_\omega[Q]}\left[Q^R\right]\right) - \sum_{i=1}^k n\left(\Delta^R_{\Gamma_\omega[Q_i]}\left[Q_i^R\right]\right) \right\|_\infty \le 4 \sum_{i=1}^k \left|\partial^R Q_i\right|. \tag{8.21}$$

(AP4) boundedness: For $Q \in \mathcal{C}$ such that $Q^R \neq \emptyset$ we have

$$\sup_{\omega\in\Omega} \left\| n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)\right\|_\infty \leq |Q^R|$$

(AP5) monotonicity: the function $n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)$ is monotone increasing at all points, i.e. for all $Q \in \mathcal{C}$ such that $Q^R \neq \emptyset$ and all $\omega \in \Omega$ we have

$$\forall x_1, x_2 : x_1 < x_2 \Rightarrow n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)(x_1) \leq n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right)(x_2).$$

(AP6) point-wise measurability: the function $n\left(\Delta^R_{\Gamma_\cdot[Q]}[Q^R]\right)(x) : \Omega \to \mathbb{R}$ is X^R_Q-measurable for all $x \in \mathbb{R}$ and $Q \in \mathcal{C}$, with the usual Borel σ-algebra as the σ-algebra of $\mathbb{R}$.

Note that almost additivity was already proven in [Sch12b].

Proof. We show the properties in the ordering given above:

To (AP1) Here, we will show that the sub-graphs $\Gamma_\omega[Q+z]$ and $\Gamma_{(\gamma_z\omega)}[Q]$ are isomorphic and thus the eigenvalues of $\Delta^R_{\Gamma_\omega[Q+z]}[Q^R+z]$ and $\Delta^R_{\Gamma_{(\gamma_z\omega)}}[Q]$ are identical. The vertex set of $\Gamma_{(\gamma_z\omega)}[Q]$ is Q and the vertex set of $\Gamma_\omega[Q+z]$ is $Q+z$, so the map $\beta\colon Q \to Q+z, \beta(y) = y+z$ is bijective. We need to check that $a, b \in Q$ are adjacent in $\Gamma_{(\gamma_z\omega)}[Q]$ if and only if $\beta(a)$ and $\beta(b)$ are adjacent in $\Gamma_\omega[Q+z]$.
The points a and b being adjacent in $\Gamma_{(\gamma_z\omega)}[Q]$ is equivalent to $1 = (\gamma_z\omega)_{[a,b]} = \omega_{[a+z,b+z]} = \omega_{[\beta(a),\beta(b)]}$, which is equivalent to $\beta(x)$ and $\beta(y)$ being adjacent in $\Gamma_\omega[Q+z]$. Therefore, β is a graph isomorphism as required.

To (AP2) If $X^R_Q(\omega) = X^R_Q(\omega')$ then any edge of length less or equal to R that is activated/deactivated in $\Gamma_\omega[Q]$ is also activated/deactivated in $\Gamma'_\omega[Q]$ respectively. These are the only edges that influence the finite range Laplace operator, thus the eigenvalue-counting functions have to be identical.

To (AP3) Let $Q = \bigcup_{i=1}^{k} Q_i$ with disjoint sets $Q_i,\ i = 1, ..., k$. Since $\Delta^R_{\Gamma_\omega[Q]}[Q^R]$ has only range R,

$$\Delta^R_{\Gamma_\omega[Q]}\left[\bigcup_{i=1}^{k} Q_i^R\right] = \bigoplus_{i=1}^{k} \Delta^R_{\Gamma_\omega[Q]}\left[Q_i^R\right] = \bigoplus_{i=1}^{k} \Delta^R_{\Gamma_\omega[Q_i]}\left[Q_i^R\right] \tag{8.22}$$

where the last equality follows from the fact that $\Delta^R_{\Gamma_\omega[Q]}\left[Q_i^R\right]$ only depends on function values in Q_i^R and edges connecting to Q_i^R with length less than or equal

to R, so everything outside of Q_i can be ignored.
The result from (8.22) for the evcfs is

$$n\left(\Delta^R_{\Gamma_\omega[Q]}\left[\bigcup_{i=1}^{k} Q_i^R\right]\right) = \sum_{i=1}^{k} n\left(\Delta^R_{\Gamma_\omega[Q_i]}\left[Q_i^R\right]\right), \tag{8.23}$$

parallel to Lemma 4.5.
Lemma 8.15 allows us to estimate the error that arises from restricting the operator to $\ell^2\left(\bigcup_{i=1}^{k} Q_i^R\right)$ instead of $\ell^2\left(Q^R\right)$ as

$$\left\| n\left(\Delta^R_{\Gamma_\omega[Q]}\left[Q^R\right]\right) - n\left(\Delta^R_{\Gamma_\omega[Q]}\left[\bigcup_{i=1}^{k} Q_i^R\right]\right)\right\|_\infty \leq 4\sum_{i=1}^{k}\left|\partial^R Q_i\right| \tag{8.24}$$

resulting in

$$\left\| n\left(\Delta^R_{\Gamma_\omega[Q]}\left[Q^R\right]\right) - \sum_{i=1}^{k} n\left(\Delta^R_{\Gamma_\omega[Q_i]}\left[Q_i^R\right]\right)\right\|_\infty \leq 4\sum_{i=1}^{k}\left|\partial^R Q_i\right| \tag{8.25}$$

as required. See also the proof of Lemma 4.5.

To (AP4) The boundedness of the eigenvalue-counting function follows from the fact that the restricted Laplace operators are real symmetric matrices.

To (AP5) The monotonicity follows from the definition of the eigenvalue-counting function.

To (AP6) Since the finite range Laplace operator on a sub-graph does not depend on edges with length longer than R we have

$$n\left(\Delta^R_{\Gamma_\cdot[Q]}[Q^R]\right)(x) = n\left(\Delta_{\Gamma_{X_Q^R(\cdot)}}[Q^R]\right)(x) \tag{8.26}$$

with the subgraph $\Gamma^R_{X_Q^R(\omega)}$ defined in 8.2.1. Thus, $n\left(\Delta^R_{\Gamma_\cdot[Q]}[Q^R]\right)(x)$ only depends on X_Q^R. Consequently it is X_Q^R-measurable, as the induced σ-algebra of $X_Q^R(\cdot)$ on Ω_Q^R is the power set and every function on Ω_Q^R is X_Q^R-measurable. This includes $n\left(\Delta_{\Gamma^R_{X_Q^R(\cdot)}}[Q^R]\right)(x)$.

□

8.2.5 Geometric bounds

Now that we have established the necessary properties, we can continue on to the geometric bounds that result from them. Here we will proceed along the proofs of Lemma 5.7 and Lemma 5.8.
Recall that

$$T_{m,n} := \{t \in T_m : \Lambda_m + t \subseteq \Lambda_n\},$$

$\rho(m) = \max_{x,y\in\Lambda_m} \mathrm{d}_{\mathbb{Z}^d}(x,y)$ and

$$\Lambda_{m,n} := \bigcup_{t\in T_{m,n}} (\Lambda_m + t)\,.$$

Lemma 8.17
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence in $\mathbb{Z}^d$ as in Definition 5.3. Then we have for all $\omega \in \Omega$, $n > W(\rho(m))$ with W from Remark 5.4 and any $R \in \mathbb{N}$ such that $\Lambda_n^R \neq \emptyset$, $\Lambda_m^R \neq \emptyset$

$$\left\| \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda_n^R]\right)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m+t]}[\Lambda_m^R+t]\right)}{|\Lambda_m|}\right\|_\infty$$
$$\leq 2\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{R}(\Lambda_{m,n})\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{\rho(m)+R}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{R}(\Lambda_m)\right|}{|\Lambda_m|}$$

and

$$\lim_{m\to\infty}\lim_{n\to\infty}\left\| \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda_n^R]\right)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m]}[\Lambda_m^R]\right)}{|\Lambda_m|}\right\|_\infty = 0.$$

Proof. We can mostly follow the proof of Lemma 5.7 with $a(\Lambda_n,\omega) = n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda_n^R]\right)$ and use the triangle inequality for

$$\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m+t,\omega)}{|\Lambda_m|}\right\|_\infty$$
$$\leq \left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|}\right\|_\infty + \left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|}\right\|_\infty \tag{8.27}$$
$$+ \left\| \frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{a(\Lambda_m+t,\omega)}{|\Lambda_m|}\right\|_\infty.$$

We use (5.20) and (AP4) for

$$\left\| \frac{a(\Lambda_n,\omega)}{|\Lambda_n|} - \frac{a(\Lambda_n,\omega)}{|\Lambda_{m,n}|} \right\|_\infty \leq \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|\,|\Lambda_n|} \|a(\Lambda_n,\omega)\|_\infty \leq \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|}. \tag{8.28}$$

For the second term we use (AP3) and

$$\hat{\Lambda}_{m,n} = \Lambda_n \setminus \Lambda_{m,n}$$

for

$$\left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} \right\|_\infty \leq \frac{4\left|\partial^R(\Lambda_{m,n})\right| + 4\left|\partial^R(\hat{\Lambda}_{m,n})\right| + \left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty}{|\Lambda_{m,n}|}. \tag{8.29}$$

For the next step we need to establish a further property of monotiling Følner sequences along the lines of Lemma 5.5, namely that if we have $\Lambda', \Lambda \in \mathcal{C}$ with $\Lambda' \subseteq \partial^r(\Lambda) \cap \Lambda$ for an $r \in \mathbb{N}$ then $\partial^R(\Lambda') \subseteq \partial^{r+R}(\Lambda)$ for any $R \in \mathbb{N}$.
To see this let $z \in \partial^R(\Lambda')$. Either $z \in \Lambda'$ and therefore $z \in \partial^r(\Lambda)$, or $z \in {\Lambda'}^C$ and there is a $y \in \Lambda'$ such that $d_{\mathbb{Z}^d}(z,y) \leq R$. Here we can distinguish two cases:

- $z \in \Lambda^C$: Then we have $d_{\mathrm{set}}(z,\Lambda) \leq R$ since $\Lambda' \subseteq \Lambda$.
- $z \in \Lambda \setminus \Lambda'$: Since $y \in \Lambda' \subseteq \partial^r(\Lambda)$ we have $d_{\mathrm{set}}(y,\Lambda^C) \leq r$ and thus

  $$d_{\mathrm{set}}(z,\Lambda^C) \leq d_{\mathbb{Z}^d}(z,y) + d_{\mathrm{set}}(y,\Lambda^C) \leq R + r$$

  resulting in $z \in \partial^{r+R}(\Lambda)$.

From $\partial^h(\Lambda) \subseteq \partial^k(\Lambda)$ for all $h \leq k$ then follows $\partial^R(\Lambda') \subseteq \partial^{r+R}(\Lambda)$ as claimed. We use this for $\Lambda' = \hat{\Lambda}_{m,n}$ By (a) of Lemma 5.5 we have $\Lambda_n \subset \partial^{\rho(m)}(\Lambda_n) \cup \Lambda_{m,n}$ and consequently $\hat{\Lambda}_{m,n} \subseteq \partial^{\rho(m)}(\Lambda_n) \cap \Lambda_n$. Thus,

$$\left|\partial^R(\hat{\Lambda}_{m,n})\right| \leq \left|\partial^{\rho(m)+R}(\Lambda_n)\right|$$

and

$$\left\|a(\hat{\Lambda}_{m,n},\omega)\right\|_\infty \leq \left|\hat{\Lambda}_{m,n}\right| \leq \left|\partial^{\rho(m)}(\Lambda_n)\right|.$$

We can therefore bound (8.29) by

$$\left\| \frac{a(\Lambda_n,\omega) - a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} \right\|_\infty \leq \frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^R(\Lambda_{m,n})\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{\rho(m)+R}(\Lambda_n)\right|}{|\Lambda_{m,n}|}. \tag{8.30}$$

For the third term of (8.27) we use (AP3) for

$$\begin{aligned}
&\left\| \frac{a(\Lambda_{m,n},\omega)}{|\Lambda_{m,n}|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{a(\Lambda_m + t,\omega)}{|\Lambda_m|} \right\|_\infty \\
&\quad = \frac{1}{|T_{m,n}|\,|\Lambda_m|} \left\| a(\Lambda_{m,n},\omega) - \sum_{t\in T_{m,n}} a(\Lambda_m + t,\omega) \right\|_\infty \\
&\quad \le \frac{1}{|T_{m,n}|\,|\Lambda_m|} \sum_{t\in T_{m,n}} 4\left|\partial^R(\Lambda_m + t)\right| = 4\frac{\left|\partial^R(\Lambda_m)\right|}{|\Lambda_m|}. \qquad (8.31)
\end{aligned}$$

Combining (8.28), (8.30) and (8.31) leads to the claimed bound

$$\begin{aligned}
&\left\| \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda^R_n]\right)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m t]}[\Lambda^R_m t]\right)}{|\Lambda_m|} \right\|_\infty \\
&\le 2\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^R(\Lambda_{m,n})\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{\rho(m)+R}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^R(\Lambda_m)\right|}{|\Lambda_m|}
\end{aligned}$$

and the convergence

$$\lim_{m\to\infty}\lim_{n\to\infty} \left\| \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda^R_n]\right)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|} \sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m t]}[\Lambda^R_m t]\right)}{|\Lambda_m|} \right\|_\infty = 0.$$

follows from the properties of monotiling Følner sequences listed in Lemma 5.5. □

We can use this Lemma to establish convergence of the expected value with basically the same proof as Lemma 5.8, leading to

Lemma 8.18
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence in $\mathbb{Z}^d$ as in Definition 5.3. Then $\frac{\mathbb{E}n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda^R_n]\right)}{|\Lambda_n|}$ forms a $\|\cdot\|_\infty$-Cauchy sequence and there exists a limit function $a^* \in \mathbb{B}$ that is also monotone increasing. Furthermore for any $R \in \mathbb{N}$ such that $\Lambda^R_n \neq \emptyset$ we have

$$\left\| \frac{\mathbb{E}n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda^R_n]\right)}{|\Lambda_n|} - a^* \right\|_\infty \le 4\frac{\left|\partial^R(\Lambda_n)\right|}{|\Lambda_n|}$$

8.2.6 Concentration inequality on percolation graphs

After this preparation we are able to apply the concentration inequalities for right-continuous, bounded random functions of Lemma 6.18 to the Laplace operator on long-range percolation graphs as well.
Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence in $\mathbb{Z}^d$. Recall $c = \prod_{y\in\mathbb{Z}^d} (1+p(y)(\mathrm{e}-1))$ as defined in (8.16) and

$$\varepsilon(R) = \sum_{y\in\mathbb{Z}^d\setminus B_R} p(y),\ \rho(m) = \max_{z,y\in\Lambda_m} \mathrm{d}_{\mathbb{Z}^d}(z,y),\ \Lambda_{m,n} := \bigcup_{t\in T_{m,n}} (\Lambda_m + t)$$

with $T_{m,n} := \{t \in T_m \mid \Lambda_m + t \subseteq \Lambda_n\}$ and the tiling T_m defined in Definition 5.3.

Theorem 8.19 (Concentration inequality for the eigenvalue counting function of the Laplace operator on a long-range percolation graph along monotiling Følner sequences) *Let $(\Lambda_n)_{n\in\mathbb{N}}$ be a monotiling Følner sequence in $\mathbb{Z}^d$. Then there exists a limit function $a^* \in \mathbb{B}$ such that for all $\kappa > 0$, $\delta < \frac{1}{6c}$, $n > W(\rho(m))$ with W from Remark 5.4, $2 \le M$ and any $R \in \mathbb{N}$ such that $\Lambda_n^R \neq \emptyset$, $\Lambda_m^R \neq \emptyset$ there is a set $A_{M,n,m,\kappa,\delta,R} \in \mathcal{B}(\Omega)$ such that*

$$\begin{aligned}\left\| \frac{n\left(\Delta_\omega[\Lambda_n^R]\right)}{|\Lambda_n|} - a^* \right\|_\infty \le{}& 2\frac{\left|\partial^{\rho(m)}(\Lambda_n)\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{R}(\Lambda_{m,n})\right|}{|\Lambda_{m,n}|} + 4\frac{\left|\partial^{\rho(m)+R}(\Lambda_n)\right|}{|\Lambda_{m,n}|} \\ &+ 8\frac{\left|\partial^{R}(\Lambda_m)\right|}{|\Lambda_m|} + \kappa + \varepsilon(R) + \delta \end{aligned} \tag{8.32}$$

for all $\omega \in A_{M,n,m,\kappa,\delta,R}$.
The probability of $A_{M,n,m,\kappa,\delta,R}$ can be bounded by

$$\mathbb{P}\left(A_{M,n,m,\kappa,\delta,R}\right) \ge 1 - \exp\left(-\frac{\delta^2\,|\Lambda_n|}{4}\right) - M\exp\left(-\frac{\sqrt{|T_{m,n}|}\kappa}{\mathring{K}_M}\right),$$

where $\mathring{K}_M = \left(\frac{40(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right)\sum_{q=0}^{\infty} 2^{-q}\sqrt{2+2q\log(2)}$.

Proof. The existence of a limit a^* follows from Lemma 8.18. We use the triangle inequality for

$$
\begin{aligned}
\left\| \frac{n\left(\Delta_\omega[\Lambda_n^R]\right)}{|\Lambda_n|} - a^* \right\|_\infty \leq & \frac{\left\| n\left(\Delta_\omega[\Lambda_n^R]\right) - n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda_n^R]\right)\right\|_\infty}{|\Lambda_n|} \\
& + \left\| \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_n]}[\Lambda_n^R]\right)}{|\Lambda_n|} - \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m+t]}[\Lambda_m^R+t]\right)}{|\Lambda_m|}\right\|_\infty \\
& + \left\| \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m+t]}[\Lambda_m^R+t]\right)}{|\Lambda_m|} - \frac{\mathbb{E}n\left(\Delta^R_{\Gamma_\omega[\Lambda_m]}[\Lambda_m^R]\right)}{|\Lambda_m|}\right\|_\infty \\
& + \left\| \frac{\mathbb{E}n\left(\Delta^R_{\Gamma_\omega[\Lambda_m]}[\Lambda_m^R]\right)}{|\Lambda_m|} - a^*\right\|_\infty . \qquad (8.33)
\end{aligned}
$$

Because of (8.20) the first term on the right side of (8.33) is bound by $3(\varepsilon(R)+\delta)$ if $\omega \in \Omega_2(\delta, R, \Lambda_n)$. The second and fourth term can be bound by the terms given in Lemmas 8.17 and 8.18. Consequently, it remains only to bound the third term.
As already mentioned in (8.26) we have

$$
n\left(\Delta^R_{\Gamma_\omega[Q]}[Q^R]\right) = n\left(\Delta_{\Gamma_{X_Q^R(\omega)}}[Q^R]\right)
$$

with the subgraph $\Gamma_{X_Q^R(\omega)} = (Q, E^R_{Q,\omega})$ with

$$
E^R_{Q,\omega} = \{e \in E_Q^R \mid \left(X_Q^R(\omega)\right)_e = 1\}.
$$

If we also use (AP1) we get

$$
n\left(\Delta^R_{\Gamma_\omega[\Lambda_m+t]}[\Lambda_m^R+t]\right) = n\left(\Delta^R_{\Gamma_{\gamma_t\omega}[\Lambda_m]}[\Lambda_m^R]\right) = n\left(\Delta_{\Gamma_{X^R_{\Lambda_m}(\gamma_t\omega)}}[\Lambda_m^R]\right).
$$

The random variables $\left(X^R_{\Lambda_m+t}\right)_{t\in T_{m,n}}$ are independent, since $\min\{\mathrm{d_{set}}(\Lambda_m+t_1, \Lambda_m+t_2) \mid t_1 \neq t_2\} > 0$ by the definition of monotilings. Thus, the random variables $\left(X^R_{\Lambda_m}\circ\gamma_t\right)_{t\in T_{m,n}}$ are also independent, and by the translation invariance of $\mathbb{P}$ also identically distributed. We define

$$
\varphi_\Delta \colon \Omega^R_{\Lambda_m} \to \mathbb{B},\ \varphi_\Delta(\nu) := n\left(\Delta_{(\Lambda_m,\nu)}[\Lambda_m^R]\right),
$$

i.e. $\varphi_\Delta(X^R_{\Lambda_m}(\omega)) = n\left(\Delta^R_{\Gamma_\cdot[\Lambda_m]}[\Lambda_m^R]\right)$. Lemma 8.16 ensures that φ_Δ is a right-continuous, monotone, bounded random function and we can apply Lemma 6.18 with $s = |T_{m,n}|$, $(X_i)_{1\leq i\leq s} = \left(X^R_{\Lambda_m}\circ\gamma_t\right)_{t\in T_{m,n}}$ and $\varphi = \varphi_\Delta$ to get a set $A_{M,n,m,\kappa} \in \mathcal{B}(\Omega)$ such that

$$
\left\| \frac{1}{|T_{m,n}|}\sum_{t\in T_{m,n}} \frac{n\left(\Delta^R_{\Gamma_\omega[\Lambda_m+t]}[\Lambda_m^R+t]\right)}{|\Lambda_m|} - \frac{\mathbb{E}n\left(\Delta^R_{\Gamma_\omega[\Lambda_m]}[\Lambda_n^R]\right)}{|\Lambda_m|}\right\|_\infty \leq \kappa
$$

for all $\omega \in A_{M,n,m,\kappa} \in \mathcal{B}(\Omega)$ and

$$\mathbb{P}\left(A_{M,n,m,\kappa}\right) \geq 1 - M\mathrm{e}^{-\frac{\sqrt{|T_{m,n}|}\kappa}{\mathring{K}_M}},$$

where $\mathring{K}_M = \left(\frac{40(M+1)}{\log(3/2)(M-1)} + \frac{4}{\log(M)}\right)\sum_{q=0}^{\infty} 2^{-q}\sqrt{2+2q\log(2)}$ and $2 \leq M$. The better constant $\mathring{K}_M$ instead of K_M follows from Remark 6.20 and the fact that $\lim_{x\to-\infty} n\left(\Delta_{(\Lambda_m,\nu)}[\Lambda_m^R]\right)(x) = 0$ for all $\nu \in \Omega_{\Lambda_m}^R$.
Therefore, we can get the claimed bound (8.32) for

$$\omega \in A_{M,n,m,\kappa,\delta,R} := \Omega_2(\delta, R, \Lambda_n) \cap A_{M,n,m,\kappa} \cap \Omega_{lf}.$$

With Lemma 8.13 we can calculate

$$\begin{aligned}\mathbb{P}\left(A^C_{M,n,m,\kappa,\delta,R}\right) &= \mathbb{P}\left(\Omega_2(\delta, R, \Lambda_n)^C \cup A^C_{M,n,m,\kappa} \cup \Omega^C_{lf}\right)\\ &\leq \mathbb{P}\left(\Omega_2(\delta, R, \Lambda_n)^C\right) + \mathbb{P}\left(A^C_{M,n,m,\kappa}\right) + 0\\ &\leq \exp\left(-\frac{\delta^2\left|\Lambda_n\right|}{4}\right) + M\exp\left(-\frac{\sqrt{|T_{m,n}|}\kappa}{\mathring{K}_M}\right)\end{aligned}$$

and arrive at the wanted estimate for the probability of $A_{M,n,m,\kappa,\delta,R}$. □

Remark 8.20. This is again just a Sub-root-exponential concentration inequality, but just as before we can also use the same proofs with the Sub-exponential concentration inequalities to get that in the setting of Theorem 8.19 provided that $|T_{m,n}| \geq \left(12\frac{\mathring{K}_2}{\kappa^2}\right)^2$ there is also a set $B_{n,m,\kappa,\delta,R} \in \mathcal{B}(\Omega)$ such that

$$\mathbb{P}\left(B_{n,m,\kappa,\delta,R}\right) \geq 1 - \exp\left(-\frac{\delta^2\left|\Lambda_n\right|}{4}\right) - \exp\left(-\frac{1}{24}\kappa^2\left|T_{m,n}\right|\right),$$

and (8.32) holds for all $\omega \in B_{n,m,\kappa,\delta,R}$.

Bibliography

[AG93] M. A. Arcones and E. Gine. Limit Theorems for U-Processes. *The Annals of Probability*, 21(3):1494 – 1542, 1993. doi:10.1214/aop/1176989128.

[And58] P. W. Anderson. Absence of Diffusion in Certain Random Lattices. *Phys. Rev.*, 109:1492–1505, Mar 1958. doi:10.1103/PhysRev.109.1492.

[Bha13] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 2013. doi:10.1007/978-1-4612-0653-8.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 02 2013. doi:10.1093/acprof:oso/9780199535255.001.0001.

[CS83] W. Craig and B. Simon. Log Hölder continuity of the integrated density of states for stochastic Jacobi matrices. *Communications in Mathematical Physics*, 90(2):207 – 218, 1983. doi:10.1007/BF01205503.

[DG27] C. Davisson and L. H. Germer. Diffraction of electrons by a crystal of nickel. *Phys. Rev.*, 30:705–740, Dec 1927. doi:10.1103/PhysRev.30.705.

[DS84] F. Delyon and B. Souillard. Remark on the continuity of the density of states of ergodic finite difference operators. *Communications in Mathematical Physics*, 94:289–291, 1984. URL: https://projecteuclid.org/journals/communications-in-mathematical-physics/volume-94/issue-2/Remark-on-the-continuity-of-the-density-of-states-of/cmp/1103941286.full, doi:10.1007/BF01209306.

[Dud78] R. M. Dudley. Central Limit Theorems for Empirical Measures. *The Annals of Probability*, 6(6):899 – 929, 1978. doi:10.1214/aop/1176995384.

[Ele06a] G. Elek. Aperiodic order, integrated density of states and the continuous algebras of John von Neumann. *arXiv: Mathematical Physics*, 2006. arXiv:math-ph/0606061v1.

[Ele06b] G. Elek. L^2-spectral invariants and quasi-crystal graphs. *arXiv: Mathematical Physics*, 2006. arXiv:math/0607198.

[GGAGP12] J.D. Garcia, A. Galindo, L. Alvarez-Gaume, and P. Pascual. *Quantum Mechanics I*. Theoretical and Mathematical Physics. Springer Berlin Heidelberg, 2012. doi:10.1007/978-3-642-83854-5.

[Hal13] B.C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Springer New York, 2013. doi:10.1007/978-1-4614-7116-5.

[Kal21] O. Kallenberg. *Foundations of Modern Probability*. Springer Cham, 3 edition, 2021. doi:10.1007/978-3-030-61871-1.

[Kir07] W. Kirsch. An Invitation to Random Schroedinger operators. *arXiv: Mathematical Physics*, 2007. arXiv:0709.3707.

[Kle20] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer Cham, 3 edition, 2020. doi:10.1007/978-3-030-56402-5.

[Len02] D. Lenz. Uniform ergodic theorems on subshifts over a finite alphabet. *Ergodic Theory and Dynamical Systems*, 22(1):245–255, 2002. arXiv:math/0005067, doi:10.1017/S0143385702000111.

[LMV08] D. Lenz, P. Müller, and I. Veselić. Uniform existence of the integrated density of states for models on $\mathbb{Z}^d$. *Positivity*, 12:571–589, 11 2008. arXiv:math-ph/0607063, doi:10.1007/s11117-008-2238-3.

[LS05] D. Lenz and P. Stollmann. An ergodic theorem for Delone dynamical systems and existence of the integrated density of states. *Journal d Analyse Mathématique*, 97:1–24, 11 2005. arXiv:math-ph/0310017, doi:10.1007/BF02807400.

[LSV10] D. Lenz, F. Schwarzenberger, and I. Veselić. A Banach space-valued ergodic theorem and the uniform approximation of the integrated density of states. *Geometriae Dedicata*, 150(1):1–34, 2010. arXiv:1003.3620, doi:10.1007/s10711-010-9491-x.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Classics in Mathematics. Springer Berlin Heidelberg, 1991. doi:10.1007/978-3-642-20212-4.

[LV09] D. Lenz and I. Veselić. Hamiltonians on discrete structures: Jumps of the integrated density of states and uniform convergence. *Mathematische Zeitschrift*, 263:813–835, 12 2009. arXiv:0709.2836v2, doi:10.1007/s00209-008-0441-3.

[Mas07] P. Massart. *Concentration Inequalities and Model Selection: Ecole D'Eté de Probabilités de Saint-Flour XXXIII - 2003*. Lecture Notes in Mathematics. Springer, 2007. doi:10.1007/978-3-540-48503-2.

[MK82] F. Martinelli and W. Kirsch. On the ergodic properties of the specrum of general random operators. *Journal für die reine und angewandte Mathematik*, 1982(334):141–156, 1982. doi:10.1515/crll.1982.334.141.

[Oss87] M. Ossiander. A Central Limit Theorem Under Metric Entropy with L_2 Bracketing. *The Annals of Probability*, 15(3):897 – 919, 1987. doi:10.1214/aop/1176992072.

[OW87] D. S. Ornstein and B. Weiss. Entropy and isomorphism theorems for actions of amenable groups. *Journal d'Analyse Mathématique*, 48:1–141, 1987. doi:10.1007/BF02790325.

[Pas73] L. A. Pastur. Spectra of random self adjoint operators. *Russian Mathematical Surveys*, 28(1):1, feb 1973. doi:10.1070/RM1973v028n01ABEH001396.

[Pas80] L. A. Pastur. Spectral properties of disordered systems in the one-body approximation. *Communications in Mathematical Physics*, 75(2):179 – 196, 1980. doi:10.1007/BF01222516.

[PF91] L. A. Pastur and A. Figotin. *Spectra of Random and Almost-Periodic Operators*. Springer Berlin, Heidelberg, 1991. URL: https://link.springer.com/book/9783642743481.

[Pol90] D. Pollard. Empirical Processes: Theory and Applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 2:i–86, 1990. doi:10.1214/cbms/1462061091.

[PS16] F. Pogorzelski and F. Schwarzenberger. A Banach space-valued ergodic theorem for amenable groups and applications. *Journal d'Analyse Mathématique*, 130, 11 2016. arXiv:1205.3649, doi:10.1007/s11854-016-0028-6.

[RS80] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: Functional analysis*. Number Bd. 1 in Methods of Modern Mathematical Physics. Academic Press, 1980. doi:10.1016/B978-0-12-585001-8.X5001-6.

[Sch26] E. Schrödinger. Quantisierung als Eigenwertproblem. *Annalen der Physik*, 384(4):361–376, 1926. doi:10.1002/andp.19263840404.

[Sch12a] K. Schmüdgen. *Unbounded Self-adjoint Operators on Hilbert Space*. Graduate Texts in Mathematics. Springer Netherlands, 2012. doi:10.1007/978-94-007-4753-1.

[Sch12b] F. Schwarzenberger. Uniform Approximation of the Integrated Density of States for Long-Range Percolation Hamiltonians. *J Stat Phys*, 146:1156–1183, 2012. `arXiv:1011.4192v1`, `doi:10.1007/s10955-012-0431-z`.

[Sch14] F. Schwarzenberger. *The Integrated Density of States for Operators on Groups.* Dissertation, TU Chemnitz, 05 2014. URL: `https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa-123241`.

[See19] A. Seelmann. Semidefinite perturbations in the subspace perturbation problem. *Journal of Operator Theory*, 81(2):321–333, 2019. `arXiv:1708.02463v2`, `doi:10.7900/jot.2018feb07.2186`.

[Shu79] M. A. Shubin. The spectral theory and the index of elliptic operators with almost periodic coefficients. *Russian Mathematical Surveys*, 34(2):109, apr 1979. `doi:10.1070/RM1979v034n02ABEH002908`.

[SSV17] C. Schumacher, F. Schwarzenberger, and I. Veselić. A Glivenko–Cantelli theorem for almost additive functions on lattices. *Stochastic Processes and their Applications*, 127(1):179–208, 2017. `arXiv:1606.07664v1`, `doi:10.1016/j.spa.2016.06.005`.

[SSV18] C. Schumacher, F. Schwarzenberger, and I. Veselić. Glivenko–Cantelli theory, Ornstein–Weiss quasi-tilings, and uniform ergodic theorems for distribution-valued fields over amenable groups. *The Annals of Applied Probability*, 28(4):2417 – 2450, 2018. `doi:10.1214/17-AAP1361`.

[SSV20] C. Schumacher, F. Schwarzenberger, and I. Veselić. *Uniform Existence of the IDS on Lattices and Groups*, page 445–478. London Mathematical Society Lecture Note Series. Cambridge University Press, 2020. `arXiv:1901.08834`, `doi:10.1017/9781108615259.017`.

[Tal94] M. Talagrand. Sharper Bounds for Gaussian and Empirical Processes. *The Annals of Probability*, 22(1):28 – 76, 1994. `doi:10.1214/aop/1176988847`.

[vW96] A. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes. With applications to statistics.* New York, NY: Springer, 1996. `doi:10.1007/978-3-031-29040-4`.

[Web10] A. Weber. Analysis of the physical Laplacian and the heat flow on a locally finite graph. *Journal of Mathematical Analysis and Applications*, 370(1):146–158, 2010. `doi:10.1016/j.jmaa.2010.04.044`.

[Wei80] J. Weidmann. *Linear Operators in Hilbert Spaces.* Graduate texts in mathematics. Springer, New York ; Heidelberg [u.a.], 2. [pr.] edition, 1980. `doi:10.1007/978-1-4612-6027-1`.

[Wer07] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2007. doi:10.1007/978-3-662-55407-4.

[Wri81] F. T. Wright. The Empirical Discrepancy Over Lower Layers and a Related Law of Large Numbers. *The Annals of Probability*, 9(2):323 – 329, 1981. doi:10.1214/aop/1176994475.