Limits of Structures

Balázs Szegedy

Alfréd Rényi Institute of Mathematics

Acknowledgements

First of all, I am extremely grateful to my mentor and co-author László Lovász. Our frequent collaborations have fundamentally shaped the way I see and do mathematics. The first part of this thesis is based on our joint works on the theory of graph limits. I feel very lucky that I could participate in the birth of this beautiful subject. It was also my main motivation for many of my later research projects such as many years in the development of higher order Fourier analysis. I am also very grateful to all who supported and influenced my journey through mathematics in the past 3 decades. Without completeness this list includes Christian Szegedy, László Surányi, Miklós Abért, Péter Pál Pálfy, László Pyber, Gábor Elek, Károly Podoski. I am very grateful to the generous research grants that were fundamental for my work. Among these grants the next four contributed the most to the present thesis: NSERC grant, Sloan fellowship, Lendület grant, ERC grant. I am very grateful to Janka Fritz for the great help in the submission process of this thesis. And last but not least I am the most grateful to my parents who supported me in all possible ways.

Contents

1 Introduction

Structural limit theories provide a fruitful connection between finite mathematics and analysis. The use of infinite (measurable or continuous) structures in the study of large discrete structures is not new. For example in physics, fluids consist of finitely many particles however it is often convenient to treat them as continuous structures. In many cases, it is helpful to consider complicated large structures as approximations of simpler infinite structures. This is exactly the aim of the recently emerging structural limit theories. Our goal in this thesis is to demonstrate various aspects of such limit theories with a special emphasis on graph limits. Graph limit theory is at an interesting meeting point of algebra, probability theory, dynamical systems, combinatorics, analysis and statistical physics. We attempt to highlight many of these directions.

Structural limits can be traced back to the ancient Greeks who discovered that it is useful to approximate the circle by finite polygones. Such approximations are essential in differential and integral calculus and so they are basically everywhere in mathematics and physics. In these cases we start with a given infinite object and we approximate it by finite structures using some kind of discretization. However what we are interested here goes in the opposite direction. Assume that we have a growing sequence of fine structures (say graphs) that are meaningfully connected to each other. For example a simple (possibly randomized) rule produces the sequence. We expect large members of the sequence to be similar to each other in some sense. It is an interesting question how to measure this similarity. Furthermore it is often helpful to find some infinite structure which behaves as a limit object for the sequence. There are various approaches for the question of similarity. One approach is based on the similarity of certain algebraic invariants associated with the structures. This is the point where both algebra and statistical physics enters the picture. These invariants can often be treated as weighted homomorphism numbers or values of partition functions.

The present thesis is mostly built on the results from 4 papers. The first two are about limits of graphs [65],[51]. Both are joint works with László Lovász and one of them is also joint with Hamed Hatami. The third paper [80] is the solution of a conjecture by Freedman, Lovász and Schrijver, and it belongs to the part of graph limit theory which is mostly related to statistical physics. The fourth paper [82] connects graph limit theory with additive combinatorics and group theory.

1.1 History and basic concepts

The history of structural limits can be traced back all the way to ancient Greeks. Archimedes (287-212 BC) used polygon approximations of the circle to compute its area. Structural limit theories are routinely used in physics. Continuous limits are essential in thermodynamics and fluid dynamics where large but finite particle systems are investigated. On the other hand discrete approximations of continuous objects such as lattice gauge theory play also an important role in physics.

Many of the above limit theories are based on very simple correspondences between finite objects and continuous limit objects. Most of the time the finite approximation is directly related to a continuous space through a prescribed geometric connection. By somewhat abusing the term, we call such limits *scaling limits*. Much more mysterious and surprising limit theories emerged more recently where simple and very general structures are considered such as 0-1 sequences or graphs. In these theories there is no "prescribed" geometry to be approximated. The geometry emerges from the internal "logic" of the structure and thus a great variety of geometric, topological and algebraic structures can appear in the limit. Many of these limit theories are based on taking small random samples from large structures. We call such limit theories *local* limit theories. Some other limit theories are based on observable, large scale properties and we call them *global* limit theories. Furthermore there are hybrid theories such as the local-global convergence of bounded degree graphs [51].

Scaling limits of $0-1$ sequences: As an illustration we start with a rather simple (warm up) limit theory for $0-1$ sequences. Later we will see a different and much more complicated theory for the same objects. For $k \in \mathbb{N}$ let $[k] := \{1, 2, \ldots, k\}$. A $0-1$ sequence of length k is a function $f : [k] \to \{0, 1\}$. Assume that we are given a growing sequence $\{f_n\}_{n=1}^{\infty}$ of $0-1$ sequences. *In what sense can we say that these sequences converge?* A simple and natural approach would be to regard the set $[k]$ as a discretization of the $(0, 1]$ interval. This way, for a $0-1$ sequence s of length k we can define the function $\tilde{s} : [0, 1] \to \{0, 1\}$ by $\tilde{s}(x) := s(\lceil kx \rceil)$ (and $\tilde{s}(0) := 0$). Now we can replace the functions f_n by $\tilde{f}_n$ and use one of the readily available convergence notions for functions on $[0, 1]$ such as L^2 or L^1 convergence. Note that they are equivalent for $0-1$ valued functions. The limit object in L^2 is a Lebesgue measurable function $f : [0, 1] \to \{0, 1\}$ with the property that the measure of $f^{-1}(1) \triangle \tilde{f}_n^{-1}(1)$ converges to 0 as n goes to infinity. A much more interesting and flexible limit concept is given by the weak convergence in $L^2([0, 1])$. For $0-1$ valued functions this is equivalent with the fact that for every interval $I = [a, b] \subseteq [0, 1]$ the measure of $I \cap \tilde{f}_n^{-1}(1)$ converges to some quantity $\mu(I)$ as n goes to infinity. The limit object is a measurable function $f : [0, 1] \to [0, 1]$ with the property that $\mu(I) = \int_I f d\lambda$ where λ is the Lebesgue measure. If $\tilde{f}_n$ is L^2 convergent then its weak limit is the same as the L^2 limit. However many more sequences satisfy weak convergence.

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of $0-1$ sequences. We say that f_n is **scaling convergent** *if $\{\tilde{f}_n\}_{n=1}^{\infty}$ is a weakly convergent sequence of functions in $L^2([0, 1])$. The limit object (scaling limit) is a measurable function of the form $f : [0, 1] \to [0, 1]$.*

Although scaling convergence is a rather simplistic limit notion we can use it as a toy example to illustrate some of the fundamental concepts that appear in other, more interesting limit theories.

- **Compactness:** *Every sequence of $0-1$ sequences has a scaling convergent subsequence*

- **Uniformity norm:** *Scaling convergence can be metrized through norms. An example for such a norm is the "intervall norm" defined by $\|f\|_{\text{in}} := \sup_I \left|\int_I f d\lambda\right|$. where I runs through all intervals in $(0, 1]$. The distance of two $0-1$ sequences f_1 and f_2 (not necessarily of equal lenght) is defined as $\|\tilde{f}_1 - \tilde{f}_2\|_{\text{in}}$.*

- **Quasi randomness:** *A $0-1$ sequence f is ϵ-quasi random with density $p \in [0, 1]$ if $\|\tilde{f} - p\|_{\text{in}} \leq \epsilon$. Note that if f_n is a sequence of $0-1$ sequences such that f_n is ϵ_n quasi random with density p and ϵ_n goes to 0 then f_n converges to the constant p function.*

- **Random objects are quasi random:** *Let f_n be a random $0-1$ sequence of length n in which the probabilty of 1 is p. For an arbitrary $\epsilon > 0$ we have that if n is large enough then with probability arbitrarily close to 1 the function f_n is ϵ quasi random.*

- **Low complexity approximation (regularization):** *For every $\epsilon > 0$ there is some natural number N_ϵ such that for every $0-1$ sequence f there is a function $g : [N_\epsilon] \to [0,1]$ such that $\|\tilde{f} - \tilde{g}\|_{\text{in}} \le \epsilon$. (Note $\tilde{g}$ is defined by the same formula as for $0-1$ sequences and $\tilde{g}$ is a step function on $[0,1]$ with N_ϵ steps.)*

Local limits of $0-1$ sequences: The main problem with scaling convergence is that highly structured sequences such as periodic sequences like $0,1,0,1,0,1,\dots$ are viewed as quasi random. The above limit concept is based on a prescribed geometric correspondence between integer intervals and the continuous $[0,1]$ interval. A different and much more useful limit concept does not assume any prescribed geometry. It is based on the local statistical properties of $0-1$ sequences. For any given $0-1$ sequence h of length k and f of length $n \ge k$ we define $t(h,f)$ to be the probability that randomly chosen k consecutive bits in f are identical to the sequence h (if $n < k$ then we simply define $t(H,f)$ to be 0).

A sequence $\{f_n\}_{n=1}^{\infty}$ of growing $0-1$ sequences is called **locally** *convergent if for every fix $0-1$ sequence h we have that $\lim_{n\to\infty} t(h,f_n)$ exists.*

This definition was first used by Furstenberg [40] in his famous correspondence principle stated in the 70's, a major inspiration for all modern limit theories. In Furstenberg's approach finite $0-1$ sequences are regarded as approximations of subsets in certain dynamical systems called measure preserving systems. A measure preserving system is a probability space $(\Omega, \mathcal{B}, \mu)$ together with a measurable transformation $T : \Omega \to \Omega$ with the property that $\mu(T^{-1}(A)) = \mu(A)$ for every $A \in \mathcal{B}$.

Furstenberg's correspondence principle for $\mathbb{Z}$: *Let f_n be a locally convergent sequence of $0-1$ sequences. Then there is a measure preserving system $(\Omega, \mathcal{B}, \mu, T)$ and a measurable set $S \subseteq \Omega$ such that for every $0-1$ sequence $h : [k] \to \{0,1\}$ the quantity $\lim_{n\to\infty} t(h,f_n)$ is equal to the probability that $(1_S(x), 1_S(x^T), \dots, 1_S(x^{T^{k-1}})) = h$ for a random element $x \in \Omega$.*

Note that originally the correspondence principle was stated in a different and more general form for amenable groups. If the group is $\mathbb{Z}$ then it is basically equivalent with the above statement. A measure preserving system is called **ergodic** if there is no set $A \in \mathcal{A}$ such that $0 < \mu(A) < 1$ and $\mu(A \triangle T^{-1}(A)) = 0$. Every measure preserving system is the combination of ergodic ones and thus ergodic measure preserving systems are the building blocks of this theory.

We give two examples for convergent $0-1$ sequences and their limits. Let α be a fixed irrational number. Then, as n tends to infinity, the sequences $1_{[0,1/2]}(\{\alpha i\}), i = 1,2,\dots,n$ (where $\{x\}$ denotes the fractional part of x) approximate the semicircle in a dynamical system where the circle is rotated by $2\pi\alpha$ degrees. Both the circle and the semicircle appears in the limit. A much more surprising example (in a slightly different form) is given by Host and Kra in [54]. Let us take two $\mathbb{Q}$-independent irrational numbers α, β and let $a_i := 1_{[0,1/2]}(\{[i\beta]i\alpha - i(i-1)\alpha\beta/2\})$ where $[x]$ denotes the integer part of x. In this case the limiting dynamical system is defined on a three dimensional compact manifold called *Heisenberg nilmanifold.*

Topologization and algebraization: At this point it is important to mention that Furstenberg's correspondence principle does not immediately give a "natural" topological representation of the limiting measure preserving system. In fact the proof yields a system in which the ground space is the compact set $\{0,1\}^{\mathbb{Z}}$ with the Borel σ-algebra, T is the shift of coordinates by one and μ

is some shift invariant measure. The notion of isomorphism between systems allows us to switch $\{0,1\}^{\mathbb{Z}}$ to any other standard Borel space. However in certain classes of systems it is possible to define a "nicest" or "most natural" topology . An old example for such a topologization is given by Kronecker systems [41]. Assume that the measure preserving map T is ergodic and it has the property that $L^2(\Omega)$ is generated by the eigenvectors of the induced action of T on $L^2(\Omega)$. It turns out that such systems can be represented as rotations in compact abelian groups (called Kronecker systems). The problem of topologization is a recurring topic in limit theories. It often comes together with some form of "algebraization" in the frame of which the unique nicest topology is used to identify an underlying algebraic structure that is intimately tied to the dynamics. Again this can be demonstrated on Kronecker systems where finding the right topology helps in identifying the Abelian group structure. Note that there is a highly successful and beautiful story of topologization and algebraization in ergodic theory in which certain factor-systems of arbitrary measure preserving systems (called characteristic factors) are identified as inverse limits of geometric objects (called nilmanifolds) arising from nilpotent Lie groups [52],[93]. As this breakthrough was also crucial in the development of higher order Fourier analysis we will give more details in the next paragraph. In many limit theories the following general scheme appears .

discrete objects $\rightarrow$ measurable objects $\rightarrow$ topological objects $\rightarrow$ algebraic objects

The first arrow denotes the limit theory, the second arrow denotes topologization and the third arrow is the algebraization.

Factors: Factor systems play a crucial role in ergodic theory. A factor of a measure preserving system $(\Omega,\mathcal{B},\mu,T)$ is a sub σ-algebra $\mathcal{F}$ in $\mathcal{B}$ that is T invariant (if $B\in\mathcal{F}$ then $T^{-1}(B)\in\mathcal{F}$). Note that if $\mathcal{F}$ is a factor then $(\Omega,\mathcal{F},\mu,T)$ is also a measure preserving system. Often there is a duality between a system of "observable quantities" defined through averages and certain factors, called *characteristic factors.* For example the averages

$$t(f):=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\int_x f(x)f(T^i(x))f(T^{2i}(x))d\mu$$

defined for bounded measurable functions satisfy that $t(f)=t(\mathbb{E}(f|\mathcal{K}))$ where $\mathcal{K}$ is the Kronecker factor of the system (the unique largest factor that is a Kronecker system) and $\mathbb{E}(f|\mathcal{K})$ is the conditional expectation with respect to $\mathcal{K}$. Since conditional expectation is an elementary operation, this means that properties of $t(f)$ can be completely described in terms of Kronecker systems. The ergodic theoretic proof [40] of Roth theorem [73] on 3-term arithmetic progressions is based on this fact and a limiting argument using Furstenberg's correspondence principle. It turns out that in every ergodic measure preserving system there is a sequence of increasing, uniquely defined factors $\mathcal{K}_1\leq\mathcal{K}_2\leq\dots$ which starts with the Kronecker factor. Similarly to Roth's theorem, the study of k-term arithmetic progressions can be reduced to $\mathcal{K}_{k-2}$. The results in [52] and [93] give a complete geometric description for these factors in terms of nilsystems. Let G be a k-step nilpotent Lie group and $\Lambda\leq G$ be a co-compact subgroup. The space $N=\{g\Lambda : g\in G\}$ of left cosets of Λ is a finite dimensional compact manifold on which G acts by left multiplication. It is known that there is a unique G invariant probability measure μ on N. We have that $\{N,\mathcal{B},\mu,g\}$ is a measure preserving system for every $g\in G$ (where $\mathcal{B}$ is the Borel σ-algebra). If g acts in an ergodic way then it is called a k-step nilsystem. It was proved in [52] and [93] that for every k the factor $\mathcal{K}_k$ of an ergodic system is the inverse limits of k-step nilsystems.

Local and global limits of graphs: Although Furstenberg's correspondence principle gives the first example for a local limit theory, a systematic study of similar structural limit theories started much later. The general program of studying structures in the limit became popular in the early 2000's when graph limit theory was born [11],[64],[65],[19],[20],[22]. The motivation to develop an analytic theory for large networks came partially from applied mathematics. The growing access to large networks such as social networks, internet graphs and biological networks like the brain generated a demand for new mathematical tools to understand their approximate structure. Another motivation came from extremal combinatorics where inequalities between subgraph densities are extensively studied. An analytic view of graphs enables the use of powerful methods such as differential calculus to solve extremal problems. Similarly to ergodic theory certain graph sequences approximate infinite structures which can not be perfectly represented by finite objects. It turns out that there are simple extremal problems for graphs which have no precise finite solutions but a nice exact solution appears in the limit. This is somewhat similar to the situation with the inequality $(x^2-2)^2 \geq 0$ which has no precise solution in $\mathbb{Q}$ but it has two solutions in $\mathbb{R}$.

Similarly to $0-1$ sequences graph convergence can be defined through converging sample distributions and thus the convergence notion will depend on the sampling method. Quite surprisingly there are two different natural sampling methods. The first one works well if the graph has a non negligible edge density (such graphs are called *dense*) and the second one is defined only for bounded degree graphs. Note that on n vertices a dense graph has cn^2 edges for some non negligible $c > 0$ whereas a bounded degree graph has cn edges for some bounded c. This means that dense and bounded degree graphs are at the two opposite ends of the density spectrum. If a graph is neither dense nor bounded degree then we call it *intermediate*.

Let $G = (V, E)$ be a finite graph. In the first sampling method we choose k vertices $v_1, v_2, \dots, v_k$ independently and uniformly from V and take the graph G_k spanned on these vertices. We regard G_k as a random graph on $[k]$. For a graph H on the vertex set $[k]$ let $t^0(H,G)$ denote the probability that $G_k = H$. In dense graph limit theory, a graph sequence $\{G_n\}_{n=1}^\infty$ is called convergent if for every fixed graph H the limit $\lim_{i\to\infty} t^0(H,G)$ exists. Another equivalent approach is to define $t(H,G)$ as the probability that a random map from $V(H)$ to $V(G)$ is a *graph homomorphism* i.e. it takes every edge of H to an edge of G. This number is called the homomorphism density of H in G. In a sequence $\{G_n\}_{n=1}^\infty$, the convergence of $t(H,G_n)$ for all graphs H is equivalent with the convergence of $t^0(H,G_n)$ for all graphs H. The advantage of using homomorphism densities is that they have nicer algebraic properties such as multiplicativity and reflection positivity [64].

For the second sampling method let $\mathcal{G}_d$ denote the set of finite graphs with maximum degree at most d. Let furthermore $\mathcal{G}_d^r$ denote the set of graphs of maximum degree at most d with a distinguished vertex o called the *root* such that every other vertex is of distance at most r from o. Now if $G = (V, E)$ is in $\mathcal{G}_d$ then let v be a uniform random vertex in V. Let $N_r(v)$ denote the v-rooted isomorphism class of the radius r-neighborhood of v in G. We have that $N_r(v)$ is an element in $\mathcal{G}_d^r$ and thus the random choice of v imposes a probability distribution $\mu(r,G)$ on $\mathcal{G}_d^r$. A graph sequence $\{G_n\}_{n=1}^\infty$ is called Benjamini-Schramm convergent if $\mu(r,G_n)$ is convergent in distribution for every fixed r. The convergence notion was introduced in the paper [11] to study random walks on planar graphs. Colored and directed versions of this convergence notion can be also introduced in a similar way. Benjamini-Schramm convergence provides a rather general framework for many different problems. Note that it generalizes the local convergence of $0-1$ sequences because one can represent finite $0-1$ sequences by directed paths with 0 and 1 labels on the nodes. Limit objects for Benjamini-Schram convergent sequences are probability distributions on infinite rooted graphs with a certain measure preserving property that generalizes the concept of measure preserving system. Note that Benjamini-Schramm convergence is closely related to group theory. A finitely presented group is called Sofic if its Cayley graph is the limit of finite graphs in which the edges are directed and labeled by the generators of the group. Sofic groups are much better understood than general abstract groups. The study of sofic groups is a fruitful interplay between graph limit theory and group theory.

Global aspects of graph limit theory arise both in the dense and the bounded degree frameworks. In case of dense graph limit theory the local point of view is often not strong enough. Although it turns out that one can represent convergent sequences by so-called graphons [64] i.e. symemtric measurable functions of the form $W : [0,1]^2 \to [0,1]$ a stronger theorem that connects the convergence with Szemerédi's regularity lemma is more useful. Szemerédi's famous regularity lemma is a structure theorem describing the large scale structure of graphs in terms of quasi rendom parts. A basic compactness result in dense graph limit theory [65] (see also Theorem 2.1) connects the local and global point of views. This is used in many applications including property testing [67] and large deviation principles [26].

The Benjamini-Schramm convergence is inherently a local convergence notion and thus it is not strong enough for many applications. For example random d-regular graphs are locally tree-like but they have a highly non-trivial global structure that has not been completely described. The formalize this problem one needs a refinement of Benjamini-Schramm convergence called local-global convergence (see definition 3.1). The concept of local-global convergence was successfully used in the study of eigenvectors of random regular graphs. It was proved by complicated analytic, information theoretic and graph limit methods in [10] that almost eigenvectors of random regular graphs have a near Gaussian entry distribution. This serves as an illustrative example for the fact that deep results in graph theory can be obtained through the limit approach. We give a detailed description of the theory behind local-global graph convergence in this thesis. The corresponding chapter is based on the paper [51]. One of the main results is a characterization of local-global limits in terms of graphins (see theorem 3.3).

We have to mention that the branch of graph limit theory that deals with intermediate graphs (between dense and bounded degree) is rather underdeveloped. There are numerous competing candidates for an intermediate limit theory [16],[15],[83],[59],[71], [38]. A recent one [9] unifies many of these theories. The hope is that at least one of these approaches will become a useful tool to study real life networks such as connections in the brain or social networks. These networks are typically of intermediate type.

Limits in additive combinatorics and higher order Fourier analysis: Let A be a finite Abelian group and S be a subset in A. Many questions in additive combinatorics deal with the approximate structure of S. For example Szemerédi's theorem can be interpreted as a result about the density of arithemtic progressions of subsets in cyclic groups. It turns out that limit approaches are natural in this subject. Let $M \in \mathbb{Z}^{m\times n}$ be an integer matrix such that each element in M is coprime to the order of A. Then we can define the density of M in the pair (A, S) as the probability that $\sum M_{i,j}x_j \in S$ holds for every i with random uniform independent choice of elements $x_1, x_2, \ldots, x_n \in A$. For example the density of 3 term arithmetic progressions in S is the density of the matrix $((1, 0), (1, 1), (1, 2))$ in S. We say that a sequence $\{(A_i, S_i)\}_{i=1}^{\infty}$ is convergence if the density of all coprime matrix M in the elements of the sequence converges. This type of convergence was first investigated in [85] and limit objects were also constructed. The subject is deeply connected to Gowers norms and the subject of higher order Fourier analysis. In this thesis we describe an interesting part of this limit theory which deals with special linear patterns (so-called complexiyty one patterns) including 3 term arithmetic progressions. This special case creates a triple correspondence between graph limits, additive combinatorics and harmonic analysis. Two major results of this part of the thesis is theorem 5.13 and theorem 5.9.

Edge coloring models

Homomorphism densities play a crucial role in graph limit theory. If G is a fixed graph then the numbers $t(H, G)$ satisfy a number of algebraic properties. Quite interestingly it was observed by Freedman, Lovász and Schrijver [36] that there is a dual version of the homomorphism number which satisfies similar algebraic properties. While homomorphism number can be viewed as a summation of certain products over all labelings of the vertices of H by the vertices of G there is similar quantity which comes from summing certain products over all labelings of the edges of H. However in this second case the quantity depends on a multivariate function of the form $t : \mathbb{N}^k \to \mathbb{R}$ called an "edge coloring model" rather than a graph G. In statistical physics they come up as values of certain partition functions and they also come up as evaluations of tensor networks.

Assume that $\mathcal{C} = \{c_1, c_2, \ldots, c_d\}$ is a finite set of colors and $\psi : E(H) \to \mathcal{C}$ is $\mathcal{C}$ coloring of the edges of H. To each vertex v of H we can associate a vector v_ψ in $\mathbb{N}^d$ whos i-th coordinate measures the multiplicity of the color c_i on the edges adjacent to v. Then we denote by $t_\psi(H)$ the product of v_ψ over all vertices of H. Finally we by abusing the notation we define $t(H)$ as the sum of $t_\psi(H)$ over all $\mathcal{C}$ colorings of the edges in H. It is a very natural question (motivated mostly by statistical physics) to ask which graph parameters arise this way. Lovász, Freedman and Schrijver formulated three algebraic properties and they conjectured that these properties give an algebraic characterization of these graph parameters. We present the proof of this conjecture in this thesis (see Theorem 4.2). The corresponding section is based on the paper [80]. The proof uses a collection of algebraic methods including invariant theory of the orthogonal groups and some basic algebraic geometry. Later on these techniques were further developed by various authors and they served as the basis for many similar follow up theorems.

2 Dense graph limit theory

This chapter is very far from being a complete account of dense graph limit theory. We mostly focus on basic facts about the graph limit space summarized in a fundamental theorem (see theorem 2.1) that connects graph limits with Szemerédi's famous regularity lemma. In particular it shows that the regularity lemma (even in stronger forms) can be viewed as a compactness result in analysis. Along these lines we introduce several basic concepts and a general regularity lemma in Hilbert spaces which is useful in other sturctural limit theories.

A *graphon* is a measurable function of the form $W : [0,1]^2 \to [0,1]$ with the property $W(x,y) = W(y,x)$ for every $x, y \in [0,1]$. Let $\mathcal{W}_0$ denote the set of all graphons. If G is a finite graph on the vertex set $[n]$ then its graphon representation W_G is defined by the formula

$$W(x,y) = 1_{E(G)}(\lceil nx \rceil, \lceil ny \rceil).$$

For a graph H on the vertex set $[k]$ let

$$t(H,W) := \int_{x_1,x_2,\dots,x_k \in [0,1]} \prod_{(i,j)\in E(H)} W(x_i,x_j)\ dx_1 dx_2 \dots dx_k.$$

The quantity $t(H,W)$ is an analytic generalization of the so called homomorphism density defined for finite graphs. This is justified by the easy observation that $t(H,G) = t(H,W_G)$. Let $\mathcal{W}$ denote the set of all bounded measurable function on $[0,1]^2$ (up to 0 measure change). We will need the so-called cut norm $\|.\|_\square$ on $\mathcal{W}$. Let $F : [0,1]^2 \to \mathbb{R}$ be a bounded measurable function. Then

$$\|F\|_\square := \sup_{A,B\subseteq[0,1]} \left| \int_{A\times B} F(x,y)\ dxdy \right|$$

where A and B run through all measurable sets in $[0,1]$. Using this norm we can introduce a measure for "similarity" of two graphons U and W by $\|U - W\|_\square$. However this is not the similarity notion that we use for convergence. We need to factor out by graphon ismorphisms. If $\psi : [0,1] \to [0,1]$ is a measure preserving transformation the we define $W^\psi(x,y) := W(\psi(x), \psi(y))$. It is easy to check that this transformation on graphons preserves the homomorphism densities: $t(H,W) = t(H,W^\psi)$ holds for every finite graph H. The next distance was introduced in [65] :

$$\delta_\square(U,W) := \inf_{\phi,\psi:[0,1]\to[0,1]} \|U^\phi - W^\psi\|_\square$$

where ϕ and ψ are measure preserving transformations. It is easy to check that $\delta_\square$ is a pseudometrics i.e. it satisfies all axioms except that $d(x,y) = 0$ does not necessarily imply that $x = y$. In order to get an actual metrics we have to factor out by the equivalence relation $\sim_{\delta_\square}$ defined by $x \sim_{\delta_\square} y \Leftrightarrow d(x,y) = 0$. Let $\mathcal{X}_0 := \mathcal{W}_0 / \sim_{\delta_\square}$. Since $\delta_\square(U,W) = 0$ implies that $t(H,U) = t(H,W)$ holds for every graph H we have that $t(H,-)$ is well defined on $\mathcal{X}_0$. The following result [64],[65] in graph limit theory is fundamental in many applications.

Theorem 2.1 *We have the following statements for the metric space $(\mathcal{X}_0, \delta_\square)$.*

1. *The metric $\delta_\square$ defines a compact, Hausdorff, second countable topology on $\mathcal{X}_0$.*
2. *The function $X \to t(H,X)$ is a continuous function on $\mathcal{X}_0$ for every finite graph H.*

Two important corollaries are the following.

Corollary 2.2 *Assume that $\{G_i\}_{i=1}^\infty$ is a sequence of graphs such that $f(H) := \lim_{i\to\infty} t(H,G_i)$ exists for every finite graph H. Then there is a graphon $W \in \mathcal{W}$ such that $f(H) = t(H,W)$ holds for every H.*

Corollary 2.3 *Szemerédi's regularity lemma [87] (even in stronger forms) follows from Theorem 2.1.*

Note that, although Corollary 2.2 may be deduced from earlier results on exchangeability [1], Theorem 2.1 combines both the local and global aspects of convergence and so it is a stronger statement. In some sense it can be regarded as a common generalization of both Szemerédi's regularity lemma [87] and a result on exchangeability [1].

Topologization of graph limit theory: In the definition of a graphon $W : [0,1]^2 \to [0,1]$ the $[0,1]$ interval on the left hand side is replaceable by any standard probability space (Ω, μ). In general we need that (Ω, μ) is atomless but for certain special graphons even atoms maybe allowed. Note that the values of W represent probabilities and so the $[0,1]$ interval is crucial on the right hand side. Thus the general form of a graphon is a symmetric measurable function $W : \Omega \times \Omega \to [0,1]$. Homomorphism densities $t(H,W)$ are defined for all such general graphons and two of them are equivalent if all homomorphism densities are the same. The folowing question arises: *Given a graphon W. Is there a most natural topological space X and Borel measure μ on X such that W is equivalent with a graphon of the form $W' : X^2 \to [0,1]$?* An answer to this question was given in [66]. For a genral graphon $W : \Omega \times \Omega \to [0,1]$ there is a unique purified version of W on some Polish space X with various useful properties. The language of topologization induced a line of exciting research in extremal combinatorics. Here we give a brief overview on applications of graphons in extremal graph theory.

Extremal graphs and graphons: The study of inequalities between subgraph densities and the structure of extremal graphs is an old topic in extremal combinatorics. A classical example is Mantel's theorem which implies that a triangle free graph H on $2n$ vertices maximizes the number of edges if H is the complete bipartite graph with equal color classes. Another example is given by the Chung-Graham-Wilson theorem [27]. If we wish to minimize the density of the four cycle in a graph H with edge density $1/2$ then H has to be sufficiently quasi random. However the perfect minimum of the problem (that is $1/16$) can not be attained by any finite graph but one can get arbitrarily close to it. Both statements can be conveniently formulated in the framework of dense graph limit theory. In the first one we maximize $t(e,G)$ in a graph G with the restriction that $t(C_3,G) = 0$ (where e is the edge and C_3 is the triangle). In the second one we fix $t(e,G)$ to be $1/2$ and we minimize $t(C_4,G)$. Since the graphon space is the completion of the space of graphs it is very natural to investigate these problems in a way that we replace G by a graphon W. If we fix finite graphs $H_1, H_2, \ldots, H_k$ then all possible inequalities between $t(H_1,W), t(H_2,W), \ldots, t(H_k,W)$ are encoded in the k-dimesional point set

$$\mathcal{L}(H_1, H_2, \ldots, H_k) := \{(t(H_1,W), t(H_2,W), \ldots, t(H_k,W)) : W \in \mathcal{W}\}.$$

Note that this is a closed subset in $[0,1]^k$. As an example let e be a single edge and let P_2 denote the path with two edges. It is easy to prove that $t(P_2,W) \geq t(e,W)^2$. This inequality is encoded encoded in $\mathcal{L}(e,P_2)$ is the form that $\mathcal{L}(e,P_2) \subseteq \{(x,y) : y \geq x^2\}$. We have however that $\mathcal{L}(e,P_2)$ carries much more information. The shape of $\mathcal{L}(H_1, H_2, \ldots, H_k)$ is know in very few instances. It took decades of research to completely describe the two dimensional shape $\mathcal{L}(e,C_3)$ which gives all possible inequalities between $t(e,W)$ and $t(C_3,W)$. The characterization of $\mathcal{L}(e,C_3)$ was completed by Razborov [72] partially using limit methods (a certain differentiation on the graph limit space). Another direction of research investigates the structure of a graphon W with given subgraph densities. *A graphon W is called finitely forcible [68] if there are finitely many graphs $H_1, H_2, \dots, H_k$ such that if $t(H_i, W') = t(H_i, W)$ holds for $i = 1, 2, \dots, k$ for some $W' \in \mathcal{W}$ then W' is equivalent with W.* The motivation to study finitely forcible graphons is that they represent a large family of extremal problems with unique solution. It is very natural to ask how complicated can extremal graph theory get at the structural level. Originally it was conjectured that finitely forcible graphons admit a step function structure which is equivalent with the fact that the topologization of the graphon is a finite space. This was disproved in [68] and various examples were given with more interesting underlying topolgy. However the topology in all of these examples is compact and finite dimensional. It was asked in [68] whether this is always the case. Quite surprisingly both conjectures turned out to be false. Extremal problems with strikingly complicated topologies were constructed in [43], [28]. This gives a very strong justification of graph limit theory in extremal cobinatorics by showing that complicated infinite structures are somehow encoded into finite looking problems. The marriage between extremal graph theory and graph limit theory has turned into a growing subject with surprising results. It brought topology and analysis into graph theory and gave a deep insight into the nature and complexity of extremal structures.

2.1 Strong and Weak Regularity Lemma

We start with stating a standard version of the Lemma. For a graph $G = (V, E)$ and for $X, Y \subseteq V$, let $e_G(X, Y)$ denote the number of edges with one endnode in X and another in Y; edges with both endnodes in $X \cap Y$ are counted twice.

Let G be a bipartite graph G with bipartition $\{U, W\}$. The ratio $d = d_G(U, W) = \frac{e_G(U,W)}{|U|\cdot|W|}$ can be thought of as the density of edges between U and W. On the average, we expect that for $X \subseteq U$ and $Y \subseteq W$,

$$e_G(X, Y) \approx d|X| \cdot |Y|.$$

For two arbitrary subsets of the nodes, $e_G(X, Y)$ may be very far from this "expected value". If G is a random graph, then, however, it will be close; random graphs are very "homogeneous" in this respect. So the following definition captures how "random-like" the bipartite graph G is: We say that G is *ε-regular*, if

$$\left|\frac{e_G(X, Y)}{|X| \cdot |Y|} - d\right| \le \varepsilon$$

holds for all subsets $X \subseteq U$ and $Y \subseteq W$ such that $|X| > \varepsilon|U|$ and $|Y| > \varepsilon|W|$. Notice that we could not require the condition to hold for small X and Y: for example, if both have one element, then the quotient $e_G(X, Y)/(|X| \cdot |Y|)$ is either 0 or 1.

Let $G = (V, E)$ be a graph (not necessarily bipartite) and let S, T be disjoint subsets of V. We denote by $G[S, T]$ the bipartite graph on $S \cup T$ obtained by keeping just those edges of G that connect S and T.

A partition $\{V_1, \dots, V_k\}$ of V is called an *equipartition* if $\lfloor |V|/k \rfloor \le |V_i| \le \lceil |V_i|/k \rceil$ for all $1 \le i \le k$.

With these definitions, the Regularity Lemma can be stated as follows:

Lemma 2.4 (Szemerédi Regularity Lemma, usual form) *For every $\varepsilon > 0$ and $m > 0$ there is a $k = k(\varepsilon, m)$ such that every graph $G = (V, E)$ on at least k nodes has an equipartition $\{V_1, \dots, V_k\}$ $(m \le k \le k(\varepsilon, l))$ such that for all but εk^2 pairs of indices $1 \le i < j \le k$, the bipartite graph $G[V_i, V_j]$ is ε-regular.*

Let us restate the Regularity Lemma in a form that is more suited for our discussions. Consider a graph $G = (V, E)$ and two subsets $U, W \subseteq V$ (not necessarily disjoint). We can measure how non-random the graph between U and W is by its *irregularity*

$$\mathrm{irreg}_G(U, W) = \max_{X \subseteq U, Y \subseteq V} \big| e_G(X, Y) - d|X| \cdot |Y| \big|.$$

(Note that by scaling up by $|X| \cdot |Y|$, we can maximize over all sets $\subseteq U$ and $Y \subseteq W$.) Clearly $\mathrm{irreg}_G(U, W) \le |U| \cdot |W|$.

Lemma 2.5 (Szemerédi Regularity Lemma, second form) *For every $\varepsilon > 0$ there is a $k(\varepsilon) > 0$ such that every graph $G = (V, E)$ has an equipartition $\mathcal{P}$ into $k \le k(\varepsilon)$ classes $V_1, \dots, V_k$ such that*

$$\sum_{1 \le i < j \le k} \mathrm{irreg}_G(V_i, V_j) \le \varepsilon |V|^2.$$

The equivalence of the two forms is easy to prove. One can add further requirements (at the cost of increasing $k(\varepsilon)$), like the requirement that $\{V_1, \dots, V_k\}$ refines a given partition.

We give one more reformulation for further reference. For $u, v \in V$, let $a_G(u, v) = 1$ if $uv \in E$ and $a_G u, v = 0$ otherwise. For a partition $\mathcal{P} = \{S_1, \dots, S_k\}$ of V and $u, v \in V$, let $a_{\mathcal{P}}(u, v) = d_G(S_i, S_j)$ where $u \in S_i$ and $v \in S_j$.

Lemma 2.6 (Szemerédi Regularity Lemma, third form) *For every $\varepsilon > 0$ there is a $k(\varepsilon) > 0$ such that every graph $G = (V, E)$ has an equipartition $\mathcal{P}$ into $k \le k(\varepsilon)$ classes such that*

$$\Big| \sum_{uv \in E(H)} (a_G(u, v) - a_{\mathcal{P}}(u, v) \Big| \le \varepsilon$$

for every graph H on V that is the union of at most k^2 complete bipartite graphs.

To see how this implies the previous form, let $X = X_{ij} \subseteq V_i$ and $Y = X_{ji} \subseteq V_j$ attain the maximum in the definition of $\mathrm{irreg}_G(V_i, V_j)$, and let H_{ij} be a complete bipartite graph between X_{ij} and X_{ji}. Let H be the union of those H_{ij} for which $e_G(X_{ij}, X_{ji}) > d_G(V_i, V_j)$ and let H' be the union of the rest. Applying Lemma 2.6 to both H and H', we obtain Lemma 2.5.

One feature of the Regularity Lemma, which unfortunately forbids practical applications, is that $k(\varepsilon)$ is very large: the best proof gives a tower of height about $1/\varepsilon^2$, and unfortunately this is not far from the truth, as was shown by Gowers [44].

A related result with a more reasonable threshold was proved by Frieze and Kannan [39], but they measure irregularity in a different way. For a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of V, define $d_{ij} = \frac{e_G(V_i, V_j)}{|V_i| \cdot |V_j|}$. For any two sets $S, T \subseteq V(G)$, we expect that the number of edges of G connecting S to T is about

$$e_{\mathcal{P}}(S, T) = \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij} |V_i \cap S| \cdot |V_j \cap T|.$$

So we can measure the irregularity of the partition by $\max_{S,T} |e_G(S, T) - e_{\mathcal{P}}(S, T)|$. The Weak Regularity Lemma [39] says the following.

Lemma 2.7 (Weak Regularity Lemma) *For every $\varepsilon > 0$ and every graph $G = (V, E)$, V has a partition $\mathcal{P}$ into $k \le 2^{2/\varepsilon^2}$ classes $V_1, \dots, V_k$ such that for all $S, T \subseteq V$,*

$$|e_G(S,T) - e_{\mathcal{P}}(S,T)| \le \varepsilon |V|^2.$$

Note that we do not require here that $\mathcal{P}$ is an equipartition; it is not hard to see that this version implies that we could require $\mathcal{P}$ to be an equipartition, at the cost of increasing the bound on k to $2^{c/\varepsilon^2}$ with a larger absolute constant c.

The partition in the weak lemma has substantially weaker properties than the partition in the strong lemma; these properties are sufficient in some, but not all, applications. The bound on the number of partition classes is still rather large (exponential), but at least not a tower. We'll see that the proof obtains the partition as an "overlay" of only $2/\varepsilon^2$ sets, which in some applications can be treated as if there were only about $1/\varepsilon^2$ classes, which makes the weak lemma quite efficient (see e.g. its applications in [2]). We'll come back to the sharpness of the threshold in Section 2.5.

Other versions of the Regularity Lemma strengthen, rather than weaken, the conclusion (of course, at the cost of replacing the tower function by an even more formidable value). Such a "super-strong" Regularity Lemma was proved by Alon, Fisher, Krivelevich and Szegedy [3, 4]. Alon and Shapira [6] used this to obtain very general results in the theory of "Property Testing" in computer science.

It turns out that the Regularity Lemma has reformulations in other branches of mathematics. A probabilistic and information theoretic version was given by Tao [88]. Our goal is to describe three reformulations in analysis.

2.2 The analytic language

In this part we revisit some of the definitions from the beginning of this chapter and we add new details. A two-variable function $W : [0,1]^2 \to \mathbb{R}$ is called *symmetric* if $W(x,y) = W(y,x)$ for all $0 \le x, y \le 1$. Recall that $\mathcal{W}$ denotes the set of all bounded symmetric measurable functions $W : [0,1]^2 \to \mathbb{R}$ and that $\mathcal{W}_0$ denotes the set of symmetric measurable functions $W : [0,1]^2 \to [0,1]$. We call a function $U \in \mathcal{W}$ *a stepfunction with at most m steps* if there is a partition $\{S_1, \dots, S_m\}$ of $[0,1]$ such that U is constant on every $S_i \times S_j$. From the analytic point of view, we think of graphs as $0-1$ valued stepfunctions in $\mathcal{W}_0$ such that the steps S_i have equal sizes. It is clear that every such function represents a graph on the vertex set $\{S_i\}$ and every graph arises this way.

Every $W \in \mathcal{W}$ can be considered as a kernel operator on the Hilbert space $L_2([0,1]^2)$ by

$$(Wf)(x) = \int_0^1 W(x,y) f(y)\, dy.$$

Besides the standard L_2 and L_1 norms, we'll need the following norm on $\mathcal{W}$:

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \left| \int\limits_{S \times T} W(x,y)\, dx\, dy \right|.$$

For the case of matrices, and up to scaling, this norm is called the "cut norm"; various important properties of it were proved by Alon and Naor [5] and by Alon, Fernandez de la Vega, Kannan and

Karpinski [2]. Many of these extend to the infinite case without any change. In particular, $\|W\|_\square$ is within absolute constant factors to the $L_1 \to L_\infty$ norm of W as a kernel operator. Furthermore, the following useful equations and inequalities are easy to verify:

$$\|W\|_\square = \sup_{f,g:[0,1]\to[0,1]} |\langle f, Wg\rangle| \ge \sup_{f:[0,1]\to[0,1]} |\langle f, Wf\rangle|$$
$$\ge \sup_{S\subseteq[0,1]} \left| \int\limits_{S\times S} W(x,y)\,dx\,dy \right| \ge \frac{1}{2}\|W\|_\square. \tag{1}$$

The Weak Regularity Lemma in these terms asserts the following:

Lemma 2.8 (Weak Regularity Lemma, Analytic Form) *For every function $W \in \mathcal{W}_0$ and $\varepsilon > 0$ there is a stepfunction $W' \in \mathcal{W}_0$ with at most $\lceil 2^{2/\varepsilon^2}\rceil$ steps such that $\|W - W'\|_\square \le \varepsilon$.*

For every $W \in \mathcal{W}$ and every partition $\mathcal{P} = \{P_1, \dots, P_k\}$ of $[0,1]$ into measurable sets, let $W_{\mathcal{P}} : [0,1]^2 \to \mathbb{R}$ denote the stepfunction obtained from W by replacing its value at $(x,y) \in P_i \times P_j$ by the average of W over $P_i \times P_j$. (This is not defined when $\lambda(P_i) = 0$ or $\lambda(P_j) = 0$, but this is of measure 0; here λ denotes the Lebesgue measure.)

It was observed in [2] that we can replace the stepfunction W' in Lemma 2.8 by the stepfunction $W_{\mathcal{P}}$, where $\mathcal{P}$ is the partition into the steps of W', at the cost of increasing the error ε by a factor of at most 2. Furthermore, at the cost of replacing the bound $2^{\lceil 2/\varepsilon^2\rceil}$ on the number of steps by $2^{\lceil 20/\varepsilon^2\rceil}$, we could require that the steps have the same measure.

It can also be shown [20] that finite simple graphs (0-1 valued symmetric stepfunctions) are dense in the set $\mathcal{W}_0$ with respect to the $\|.\|_\square$-norm.

The norm $\|.\|_\square$ relates to other norms by the following inequalities. It is trivial that

$$\|W\|_\square \le \|W\|_1. \tag{2}$$

The following inequalities, proved in [20], are still simple but less obvious. For every $W \in \mathcal{W}$, let $W \circ W$ denote its square as a kernel operator, i.e.,

$$(W \circ W)(x,y) = \int_0^1 W(x,t)W(t,y)\,dt.$$

Then

$$\|W\|_\square^4 \le \|W \circ W\|_2^2 \le \|W\|_\square \|W\|_\infty^2 \|W\|_1. \tag{3}$$

So for functions in $\mathcal{W}$,

$$\|W \circ W\|_2^{1/2} \le \|W\|_\square \le \|W \circ W\|_2^2.$$

It can be checked that the left hand side, as a function of W, is a norm. Due to its more explicit form, this is often easier to handle than $\|W\|_\square$.

We conclude this section with formulating an analytic version of the strong Szemerédi Lemma (third version). A *rectangle* in $[0,1]$ is any set of the form $S \times T$, where S and T are measurable subsets of $[0,1]$.

Lemma 2.9 (Strong Regularity Lemma, Analytic Form) *For every $\varepsilon > 0$ there is an integer $k(\varepsilon) > 0$ such that for every function $W \in \mathcal{W}_0$ there is a partition $\mathcal{P} = \{S_1, \dots, S_k\}$ of $[0,1]$ into $k \le k(\varepsilon)$ sets of equal measure with the following property: For every set $R \subseteq [0,1]^2$ that is the union of at most k^2 rectangles, we have*

$$\Big|\int_R (W - W_{\mathcal{P}})\,dx\,dy\Big| \le \varepsilon.$$

2.3 The Regularity Lemma in Hilbert space

The following lemma is an extension of the Regularity lemma to a very general setting of Hilbert spaces.

Lemma 2.10 (Regularity Lemma in Hilbert Space) *Let $\mathcal{K}_1, \mathcal{K}_2, \dots$ be arbitrary nonempty subsets of a Hilbert space $\mathcal{H}$. Then for every $\varepsilon > 0$ and $f \in \mathcal{H}$ there is an $m \le \lceil 1/\varepsilon^2 \rceil$ and there are $f_i \in \mathcal{K}_i$ ($1 \le i \le m$) and $\gamma_1, \gamma_2, \dots, \gamma_m \in \mathbb{R}$ such that for every $g \in \mathcal{K}_{m+1}$*

$$|\langle g, f - (\gamma_1 f_1 + \dots + \gamma_m f_m)\rangle| \le \varepsilon \cdot \|g\| \cdot \|f\|.$$

Before proving this lemma, a little discussion is in order. Assume that the sets $\mathcal{K}_n$ are subspaces. Then a natural choice for the function $\gamma_1 f_1 + \dots + \gamma_m f_m$ is the best approximation of f in the subspace $\mathcal{K}_1 + \dots + \mathcal{K}_m$ (or an approximately best approximation, if the best does not exist), and the error $f - (\gamma_1 f_1 + \dots + \gamma_m f_m)$ is orthogonal (or almost orthogonal) to every $g \in \mathcal{K}_1 + \dots + \mathcal{K}_m$. The main point in this lemma is that it is also almost orthogonal to the *next* set K_{m+1}.

Proof. Let

$$\eta_k = \inf_{\{\gamma_i\},\{f_i\}} \|f - \sum_{i=1}^k \gamma_i f_i\|^2,$$

where the infimum is taken over all $\gamma_1, \dots, \gamma_k \in \mathbb{R}$ and $f_i \in \mathcal{K}_i$. Clearly we have $\|f\|^2 \ge \eta_1 \ge \eta_2 \ge \dots \ge 0$. Hence there is an $m \le \lceil 1/\varepsilon^2 \rceil$ such that $\eta_m < \eta_{m+1} + \varepsilon^2\|f\|^2$. So there are $\gamma_1, \dots, \gamma_m \in \mathbb{R}$ and $f_i \in \mathcal{K}_i$ such that

$$\|f - \sum_{i=1}^m \gamma_i f_i\|^2 \le \eta_{m+1} + \varepsilon^2\|f\|^2.$$

Let $f^* = \sum_i \gamma_i f_i$, and consider any $g \in \mathcal{K}_{m+1}$. By the definition of η_{m+1}, we have for every real t that

$$\|f - (f^* + tg)\|^2 \ge \eta_{m+1} \ge \|f - f^*\|^2 - \varepsilon^2\|f\|^2,$$

or

$$\|g\|^2 t^2 - 2\langle g, f - f^*\rangle t + \varepsilon^2\|f\|^2 \ge 0.$$

The discriminant of this quadratic polynomial must be nonpositive, which proves the lemma. ∎

We derive some consequences of this Lemma. First, let us apply this lemma to the case when the Hilbert space is $L^2([0,1]^2)$, and each $\mathcal{K}_n$ is the set of indicator functions of product sets $S \times S$, where S is a measurable subset of $[0,1]$. Let $f \in \mathcal{W}_0$, then $f^* = \sum_{i=1}^k \gamma_i f_i$ is a stepfunction with at most 2^k steps, and so we get a stepfunction $W^* \in \mathcal{W}$ with at most $2^{\lceil 1/\varepsilon^2 \rceil}$ steps such that for every measurable set $S \subseteq [0,1]$,

$$\Big|\int_{S\times S} (W - W^*)\Big| \le \varepsilon.$$

It is easy to see that the conclusion implies that for any two measurable sets $S, T \subseteq [0,1]$,

$$\Big|\int_{S\times T}(W - W^*)\Big| \le 2\varepsilon,$$

which implies Lemma 2.8 (up to the factor of 2).

We say that a partition $\mathcal{P}$ of $[0,1]$ is a *weak Szemerédi partition* for W with error ε, if

$$\Big|\int_{S\times S}(W - W_{\mathcal{P}})\Big| \le \varepsilon$$

holds for every subset $S \subseteq [0,1]$. So every function has a weak Szemerédi partition with error ε, with at most $2^{2/\varepsilon^2}$ classes.

To derive the graph theoretic form of the (weak) Regularity Lemma from lemma 2.10, we represent the graph G on n nodes by a stepfunction W_G: we consider the adjacency matrix $A = (a_{ij})$ of G, and replace each entry a_{ij} by a square of size $(1/n)\times(1/n)$ with the constant function a_{ij} on this square. Let $\mathcal{A}$ be the algebra of subsets of $[0,1]$ generated by the intervals corresponding to nodes of G. We let $\mathcal{K}_n$ be the set of indicator functions of product sets $S \times S$ ($S \in \mathcal{A}$). Analogously to the proof of Lemma 2.8 above, we get a partition $\mathcal{P} = \{S_1, \dots, S_m\}$ of $[0,1]$ into sets in $\mathcal{A}$ such that

$$\Big|\int_{S\times T}(W_G(x,y) - (W_G)_{\mathcal{P}}(x,y))\,dx\,dy\Big| \le 2\varepsilon$$

for all sets $S, T \in \mathcal{A}$. This translates into the conclusion of Lemma 2.7.

Next we show how to get the strong analytic form of the Regularity Lemma 2.9; the graph theoretic form can be obtained similarly (just there is a little extra trouble because of divisibilities). Let us define a sequence $s(1), s(2), \dots$ of positive integers by $s(1) = 1$ and $s(k+1) = 2^{s(1)^4 \cdots s(k)^4}$. Let us apply Lemma 2.10 to the Hilbert space $L_2([0,1]^2)$ and the function W as before, but choose $\mathcal{K}_n$ to be the set of stepfunctions with at most $s(n)$ steps. Lemma 2.10 gives us a function W^*, which is a stepfunction with at most $m = s(1)s(2)\dots s(k)$ steps; let $S_1, \dots S_m$ be these steps. This stepfunction has the property that for every stepfunction U with at most $s(k+1)$ steps,

$$\Big|\int_{[0,1]^2} U(W - W^*)\,dx\,dy \le \varepsilon.$$

We further partition each S_i into an appropriate number of sets of measure $1/m^2$ (called *good* sets and a "remainder" of measure less than $1/m^2$. We combine these remainders into single set, whose measure is less than $1/m$. We partition this into sets of size $1/m^2$; there will be at most m such sets, which we call *bad* sets. So we get a partition $\mathcal{Q} = \{T_1, \dots, T_{m^2}\}$ of $[0,1]$ into m^2 equal parts, out of which (say) $T_1, \dots, T_{m^2-m}$ are good sets.

Let $R \subseteq [0,1]^2$ be a set that is the union of m^2 rectangles. We claim that

$$\Big|\int_R (W - W_{\mathcal{Q}})\Big| \le 3\varepsilon.$$

We start with removing from R all points in sets $T_i \times T_j$, where either T_i or T_j is bad. The remaining set R' is again the union of at most m^2 rectangles, and since the measure of $R \setminus R'$ is less than $2/m < \varepsilon$, it suffices to prove that

$$\Big|\int_{R'} (W - W_{\mathcal{Q}})\Big| \le 2\varepsilon.$$

Clearly, the indicator function of R' is a stepfunction with at most $2^{m^4} \le s(k+1)$ steps, and hence by the conclusion of Lemma 2.10, we have

$$\Big|\int_{R'} (W - W^*)\Big| \le \varepsilon.$$

So it suffices to verify that

$$\Big|\int_{R'} (W_{\mathcal{Q}} - W^*)\Big| \le \varepsilon. \tag{4}$$

If T_i and T_j are good sets, then both W^* and $W_{\mathcal{Q}}$ are constant on $T_i \times T_j$, so we can either include or exclude the rectangle $T_i \times T_j$ from R', and not decrease the left hand side of (4). Doing so for every pair of good sets, we obtain a set R'', which is the union of certain sets $T_i \times T_j$, where both T_i and T_j are good. Thus by the definition of $W_{\mathcal{Q}}$, we have

$$\Big|\int_{R''} (W_{\mathcal{Q}} - W^*)\Big| = \Big|\int_{R''} (W - W^*)\Big| \le \varepsilon$$

(by the assertion of Lemma 2.10). This concludes the proof of the strong Szemerédi Lemma.

There may be further interesting choices of the Hilbert space $\mathcal{H}$ and subsets $\mathcal{K}_n$. For example, let $\mathcal{H} = L_2[0,1]$, and let $\mathcal{K}_n$ be the set of polynomials of degree at most 2^n. Then we get:

Corollary 2.11 *For every $\varepsilon > 0$ and every function $f \in L_2[0,1]$ there is a polynomial $p \in \mathcal{R}[x]$ of degree $d \le 2^{\lceil 1/\varepsilon^2 \rceil}$ such that*

$$\langle g, f - p\rangle \le \varepsilon \|f\| \cdot \|g\|$$

for every polynomial g of degree at most $2d$.

2.4 The Regularity Lemma as compactness

In this chapter we show that the Regularity Lemma can be formulated as a compactness theorem.

Recall that a map $\phi : [0,1] \to [0,1]$ is *measure preserving* if $\lambda(\phi^{-1}(U)) = \lambda(U)$ for every measurable set $U \subseteq [0,1]$. We say that ϕ is a *measure preserving bijection* if it is bijective and its inverse is also measure preserving.

Let W be a function from $\mathcal{W}$. We define W^ϕ by $W^\phi(x,y) = W(\phi(x),\phi(y))$. We define a "distance" on the space $\mathcal{W}$ by

$$\delta_\square(U,W) = \inf_\phi \|U^\phi - W\|_\square,$$

where ϕ ranges over all measure preserving bijections $[0,1] \to [0,1]$. It is not hard to check that $\delta_\square(U,W) = \delta_\square(W,U)$, and that this distance satisfies the triangle inequality. Furthermore, $\delta_\square(U,W) = \delta_\square(U^\phi,W)$ for every measure preserving bijection ϕ.

The distance of two different functions can be 0; various characterizations of when the $\delta_\square$ distance is 0 are given in [18] and [65].

We construct a metric space $\mathcal{X}$ from $(\mathcal{W},\delta_\square)$ by identifying functions U and W with $\delta_\square(U,W) = 0$. Let $\mathcal{X}_0$ denote the image of $\mathcal{W}_0$ under this identification. Informally speaking, the elements of $\mathcal{X}_0$ are the isomorphism classes of functions in $\mathcal{W}_0$. Clearly the distance $\delta_\square$ is well defined on $\mathcal{X}_0$.

The following fact can be regarded as a topological interpretation of the Regularity Lemma. (We prove it by the methods in [63].)

Theorem 2.12 *The metric space $\mathcal{X}_0$ is compact.*

Proof. Let $W_1, W_2, \dots$ be a sequence of functions in $\mathcal{W}_0$. We want to construct a subsequence that has a limit in $\mathcal{X}_0$.

Using Lemma 2.8 and the remarks after it, for each k and n we construct a partition $\mathcal{P}_{n,k}$ such that these partitions and the corresponding stepfunctions $W_{n,k} = W_{\mathcal{P}_{n,k}} \in \mathcal{W}_0$ satisfy the following.

- $\|W_n - W_{n,k}\|_\square \le 1/k$.
- $|\mathcal{P}_{n,k}| = m_k$ (where m_k depends only on k).
- The partition $\mathcal{P}_{n,k+1}$ refines $\mathcal{P}_{n,k}$ for every k.

We'll only use that $\delta_\square(W_n, W_{n,k}) \le 1/k$, which means that we can rearrange the range of $W_{n,k}$ as we wish; in particular, we may assume that all steps are intervals.

Now we can select a subsequence of the W_n for which the length of the i-th interval of $W_{n,1}$ converges for every i, and also the value of $W_{n,1}$ on the product of the i-th and j-th intervals converges for every i and j (as $n \to \infty$). It follows then that the sequence $W_{n,1}$ converges to a limit U_1 almost everywhere, which itself is a stepfunction with m_1 steps that are intervals.

We repeat this for $k = 2, 3, \dots$, to get subsequences for which $W_{k,n} \to U_k$ almost everywhere, where U_k is a stepfunction with m_k steps that are intervals.

For every $k < l$, the partition into the steps of $W_{n,l}$ is a refinement of the partition into the steps of $W_{n,k}$, and hence it is easy to see that the same relation holds for the partitions into the steps of U_l and U_k. Furthermore, the function $W_{n,k}$ can be obtained from the function $W_{n,l}$ by averaging its value over each step, and it follows that a similar relation holds for U_l and U_k.

Let (X, Y) be a random point in $[0,1]^2$ chosen uniformly, then this property of the functions U_k implies that the sequence $(U_1(X,Y), U_2(X,Y), \dots)$ is a martingale. Since the random variables $U_i(X,Y)$ remain bounded, the Martingale Convergence Theorem (see e.g. [92], Theorem 11.5) implies that this martingale is convergent with probability 1. In other words, the sequence $(U_1, U_2, \dots)$ is convergent almost everywhere. Let U be its limit.

Fix any $\varepsilon > 0$. Then there is a $k > 3/\varepsilon$ such that $\|U - U_k\|_1 < \varepsilon/3$. Fixing this k, there is an n_0 such that $\|U_k - W_{n,k}\|_1 < \varepsilon/3$ for all $n \ge n_0$. Then

$$\begin{aligned}
\delta_\square(U, W_n) &\le \delta_\square(U, U_k) + \delta_\square(U_k, W_{n,k}) + \delta_\square(W_{n,k}, W_n) \\
&\le \|U - U_k\|_1 + \|U_k - W_{n,k}\|_1 + \delta_\square(W_{n,k}, W_n) \\
&\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned}$$

This proves that $W_n \to U$ in the metric space $\mathcal{X}_0$. ■

Note that in the proof above, the explicit bound on the number of partition classes in the Regularity Lemma was not used, only that their number is bounded by a function of ε, independent of the function. This is quite often the case with applications of the Lemma.

Now we show how this compactness statement implies the following strong form of the Regularity Lemma.

Lemma 2.13 (Very Strong Regularity Lemma) *Let $h(\varepsilon,t)>0$, $(\varepsilon>0, t\in\mathbb{N})$ be an arbitrary fixed function. Then for every $\varepsilon>0$ there is a threshold $k(\varepsilon)$ such that for every function $W\in\mathcal{W}_0$ there are two functions $W', U\in\mathcal{W}_0$ such that U is a stepfunction with $l\le k(\varepsilon)$ steps, and*

$$\|W-W'\|_\square \le h(\varepsilon,l), \qquad \|W'-U\|_1\le\varepsilon.$$

The role of the two norms could be interchanged: the function $W''=U-W'+W$ satisfies

$$\|W-W''\|_1\le\varepsilon, \qquad \|W''-U\|_\square \le h(\varepsilon,1).$$

Choosing $h(\varepsilon,t)=\varepsilon$, we get Lemma 2.8. Choosing $h(\varepsilon,t)=\varepsilon/t^2$, it is not hard to see that the strong form of Szemerédi's Lemma follows. Choosing $h(\varepsilon,t)$ appropriately small, we get the "super-strong" Regularity Lemma from [3, 4] mentioned in the introduction. Our Lemma is very closely related to a version of the regularity lemma given by Tao [88].

Proof. We may assume that h is monotone decreasing in its second variable. Let us fix a number $\varepsilon>0$. Every function $U\in\mathcal{W}_0$ is the limit of stepfunctions in the $\|.\|_1$ norm, hence there is a stepfunction $U'\in\mathcal{W}_0$ with $\|U-U'\|_1\le\varepsilon$. Let $f(U)$ denote the minimum number of steps in such a stepfunction U'. For a function $U\in\mathcal{W}_0$, let $B(U)$ denote the open ball $\{W \mid \delta_\square(U,W)<h(\varepsilon,f(U))\}$. Using Theorem 2.12, we obtain that there is a finite set of functions $W_1,W_2,\dots,W_t\in\mathcal{W}_0$ with $\cup_{i=1}^t B(W_i)=\mathcal{W}_0$. This means that for every function $W\in\mathcal{W}_0$ there is a function W_m $(1\le m\le t)$ and a stepfunction $U_0\in\mathcal{W}_0$ with $f(W_m)$ steps such that $\delta_\square(W,W_m)<h(\varepsilon,f(W_m))$ and $\|W_m-U_0\|_1<\varepsilon$.

Set $l=f(W_m)$ and $k(\varepsilon)=\max_{i=1}^t f(W_i)$. There is a measure preserving bijection $\phi:\ [0,1]\mapsto[0,1]$ such that $\|W-W_m^\phi\|_\square<h(\varepsilon,l)$. Then $U=U_0^\phi$ is a stepfunction with l steps, and $W'=W_m^\phi$ satisfies

$$\|W'-U\|_1=\|W_m^\phi-U_0^\phi\|_1=\|W_m-U_0\|_1<\varepsilon$$

and

$$\delta_\square(W,W')=\delta_\square(W,W_m^\phi)=\delta_\square(W,W_m)\le h(\varepsilon,l),$$

which completes the proof. ∎

2.5 The Regularity Lemma and covering by small balls

Every function $W\in\mathcal{W}$ gives rise to a metric on $[0,1]$ by

$$d_W^1(x_1,x_2)=\|W(x_1,.)-W(x_2,.)\|_2=\Big(\int_0^1 (W(x_1,y)-W(x_2,y))^2\,dy\Big)^{1/2}.$$

It turns out that for our purposes, the following distance function is more important: we square W as a kernel operator, and then consider the above distance. More precisely, we define

$$\begin{aligned}d_W(x_1,x_2)&=d^1_{W\circ W}(x_1,x_2)\\&=\Big(\int_0^1\Big(\int_0^1 W(x_1,y)W(y,z)\,dy-\int_0^1 W(x_2,y)W(y,z)\,dy\Big)^2 dz\Big)^{1/2}.\end{aligned}$$

Our goal is to prove that the (weak) Regularity Lemma is equivalent to the assertion that most of the metric space $([0,1],\delta_W)$ can be covered by a bounded number of small balls. More exactly:

Theorem 2.14 *Let $W \in \mathcal{W}_0$ and let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a partition of $[0,1]$ into measurable sets.*

(a) *If $\mathcal{P}$ is a weak Szemerédi partition with error $\varepsilon^2/8$, then there is a set $S \subseteq [0,1]$ with $\lambda(S) \le \varepsilon$ such that for each partition class, $P_i \setminus S$ has diameter at most ε in the d_W metric.*

(b) *If there is a set $S \subseteq [0,1]$ with $\lambda(S) \le (\varepsilon/5)^4$ such that for each partition class, $P_i \setminus S$ has diameter at most $(\varepsilon/5)^2$ in the d_W metric, then $\mathcal{P}$ is a weak Szemerédi partition with error ε.*

Combining this fact with the existence of weak Szemerédi partitions, we get the following:

Corollary 2.15 *For every function $W \in \mathcal{W}$ and every $\varepsilon > 0$ there is a partition $\mathcal{P} = \{P_0, P_1, \dots, P_k\}$ of $[0,1]$ into measurable sets with $k \le 2^{\lceil 64/\varepsilon^4 \rceil}$ such that $\lambda(P_0) \le \varepsilon$ and for $1 \le i \le k$, P_i has diameter at most ε in the d_W metric.*

It is straightforward to formulate this theorem for graphs instead of functions $W \in \mathcal{W}$: We define the distance of two nodes u, v of a graph G by squaring the adjacency matrix, and taking the euclidean distance between the row vectors corresponding to u and v, divided by $n^{3/2}$. Then the statement of the Theorem is analogous, and the proof is the same.

Proof. (a) Suppose that $\mathcal{P}$ is a weak Szemerédi partition with error $\varepsilon^2/8$. Let $R = W - W_{\mathcal{P}}$, then we know that $\|R\|_\square \le \varepsilon^2/8$.

For every $x \in [0,1]$, define

$$F(x) = \int_0^1 \Big(\int_0^1 R(x,s) W(s,z)\, ds \Big)^2 dz.$$

Then we have

$$\int_0^1 F(x)\, dx = \int_0^1 \int_0^1 \int_0^1 \int_0^1 R(x,t) R(x,s) W(s,z) W(t,z)\, dx\, ds\, dt\, dz.$$

Fix z and t, then since $-1 \le R(x,t) \le 1$ and $0 \le W(s,z) \le 1$, we have

$$\int_0^1 \int_0^1 R(x,t) R(x,s) W(s,z)\, dx\, ds \le \varepsilon^2/4,$$

and so

$$\int_0^1 F(x)\, dx \le \varepsilon^2/4.$$

Hence there is a set $S \subseteq [0,1]$ with measure at most ε such that for $x \in [0,1] \setminus S$, we have $F(x) \le \varepsilon/4$.

Let $x, y \in [0,1] \setminus S$ be two points in the same partition class of $\mathcal{P}$. Then $W_{\mathcal{P}}(x,s) = W_{\mathcal{P}}(y,s)$ for every $s \in [0,1]$, and hence

$$\begin{aligned}
d_W(x,y)^2 &= \int_0^1 \Big(\int_0^1 (W(x,s) - W(y,s)) W(s,z)\, ds \Big)^2 dz \\
&= \int_0^1 \Big(\int_0^1 (R(x,s) - R(y,s)) W(s,z)\, ds \Big)^2 dz \\
&= \int_0^1 \Big(\int_0^1 R(x,s) W(s,z)\, ds - \int_0^1 R(y,s)) W(s,z)\, ds \Big)^2 dz \\
&\le 2 \int_0^1 \Big(\int_0^1 R(x,s) W(s,z)\, ds \Big)^2 dz + 2 \int_0^1 \Big(\int_0^1 R(y,s) W(s,z)\, ds \Big)^2 dz \\
&= 2F(x) + 2F(y) \le \varepsilon.
\end{aligned}$$

(b) We want to show that $\|W-W_{\mathcal{P}}\|_\square<\varepsilon$. By (1), it suffices to show that for any 0-1 valued function f,

$$\langle f,(W-W_{\mathcal{P}})f\rangle \le \frac{1}{2}\varepsilon. \tag{5}$$

Let us write $f=f_{\mathcal{P}}+g$, where $f_{\mathcal{P}}(x)$ is obtained by replacing $f(x)$ by the average of f over the class P_i containing x. It is easy to check that we have

$$\langle f,(W-W_{\mathcal{P}})f\rangle = \langle f+f_{\mathcal{P}}, Wg\rangle. \tag{6}$$

By Cauchy-Schwartz,

$$\langle f+f_{\mathcal{P}}, Wg\rangle \le \|f+f_{\mathcal{P}}\|\cdot\|Wg\| \le 2\|Wg\|. \tag{7}$$

We have

$$\|Wg\|^2=\langle g,W^2g\rangle=\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz.$$

For each x, let $\phi(x)$ be an arbitrary, but fixed, element of the class P_i containing x such that $x\notin S$ (if $P_i\subseteq S$ then we define $\phi(x)$ to be 0). Then

$$\begin{aligned}
&\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz\\
&\qquad=\int_{[0,1]^3} g(x)\big(W(x,y)W(y,z)-W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz\\
&\qquad\quad+\int_{[0,1]^3} g(x)W(x,y)W(y,\phi(z))g(z)\,dx\,dy\,dz\,.
\end{aligned}$$

Here the last integral is 0, since the integral of g over each partition class is 0. Furthermore,

$$\begin{aligned}
&\int_{[0,1]^3} g(x)\big(W(x,y)W(y,z)-W(x,y)W(y,\phi(z))\big)g(z)\,dx\,dy\,dz\\
&\qquad\le\left(\int_{[0,1]^3} g(x)^2g(z)^2\,dx\,dy\,dz\right)^{1/2}\\
&\qquad\quad\times\left(\int_{[0,1]^3}\big(W(x,y)W(y,z)-W(x,y)W(y,\phi(z))\big)^2\,dx\,dy\,dz\right)^{1/2}.
\end{aligned}$$

Here the first factor is at most 1, and

$$\begin{aligned}
&\int_{[0,1]^3}\big(W(x,y)W(y,z)-W(x,y)W(y,\phi(z))\big)^2\,dx\,dy\,dz=\int_0^1 d_W(z,\phi(z))^2\,dz\\
&\qquad=\int_{[0,1]/S} d_W(z,\phi(z))^2\,dz+\int_S d_W(z,\phi(z))^2\,dz\le 2(\varepsilon/5)^4.
\end{aligned}$$

Thus

$$\int_{[0,1]^3} g(x)W(x,y)W(y,z)g(z)\,dx\,dy\,dz\le 2^{1/2}(\varepsilon/5)^2,$$

and so $\|Wg\|\le\frac{1}{5}2^{1/4}\varepsilon<\frac{1}{4}\varepsilon$. By (6) and (7), this proves (5), and completes the proof. ∎

2.6 Proof of theorem 2.1

In this section we finish the proof of theorem 2.1. The first part of the theorem is just the statement of theorem 2.12. It remains to show the continuity of functions of the form $W \mapsto t(F, W)$. We use the following observation and lemma. It is easy to see that

$$\|U\|_\square = \sup_{0 \le f, g \le 1} \left| \int_0^1 \int_0^1 U(x, y) f(x) g(y) \right|. \tag{8}$$

Lemma 2.16 *Let $U, W : [0,1]^2 \to [0,1]$ be two symmetric integrable functions. Then for every simple finite graph F,*

$$|t(F, U) - t(F, W)| \le |E(F)| \cdot \|U - W\|_\square.$$

Proof. Let $V(F) = [n]$ and $E(F) = \{e_1, \dots, e_m\}$. Let $e_t = i_t j_t$. Define $E_t = \{e_1, \dots, e_t\}$. Then

$$t(F, U) - t(F, W) = \int_{[0,1]^n} \Big(\prod_{ij \in E(F)} W(x_i, x_j) - \prod_{ij \in E(F)} U(x_i, x_j) \Big)\, dx$$

We can write

$$\prod_{ij \in E(F)} W(x_i, x_j) - \prod_{ij \in E(F)} U(x_i, x_j) = \sum_{t=0}^{m-1} X_t(x_1, \dots, x_n),$$

where

$$X_t(x_1, \dots, x_n) = \Big(\prod_{ij \in E_{t-1}} W(x_i, x_j) \Big) \Big(\prod_{ij \in E(F) \setminus E_t} U(x_i, x_j) \Big) (W(x_{i_t}, x_{j_t}) - U(x_{i_t}, x_{j_t})).$$

To estimate the integral of a given term, let us integrate first the variables x_{i_t} and x_{j_t}; then by (8),

$$\left| \int_0^1 \int_0^1 X_t(x_1, \dots, x_n)\, dx_{i_t}\, dx_{j_t} \right| \le \|U - W\|_\square,$$

and so

$$|t(F, U) - t(F, W)| \le \sum_{t=0}^{m-1} \left| \int_{[0,1]^n} X_t(x_1, \dots, x_n)\, dx \right| \le m \|U - W\|_\square$$

as claimed. ∎

Now observe that since $t(H, W^\phi) = t(H, W)$ holds for every measure preserving function ϕ on $[0,1]$ we have by lemma 2.16 that

$$|t(F, U) - t(F, W)| \le |E(F)| \cdot \|U^\phi - W^\psi\|_\square.$$

holds for every pair of measure preserving functions ϕ and ψ. It follows that

$$|t(F, U) - t(F, W)| \le |E(F)| \cdot \delta_\square(U, W).$$

This completes the proof.

2.7 Two applications

We conclude with two applications of this characterization of weak Szemerédi partitions. First we prove that the exponential dependence of the number of classes on ε in Lemma 2.8 is necessary. Taking an appropriately dense finite subgraph of our construction, one can prove a similar bound on the threshold in the finite version Lemma 2.7.

Let S^d be the d-dimensional sphere, endowed with the uniform probability measure μ (it does not matter which probability space we consider as the domain of W, so we may consider S^d instead of $[0,1]$). For $x, y \in S^d$, let

$$W(x,y) = \begin{cases} 1, & \text{if } x^{\mathsf{T}} y \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

We prove:

Proposition 2.17 *Every weak Szemerédi partition of W with error $\varepsilon \leq 1/(8d+8)$ contains at least 2^{d-1} classes.*

Proof. We may assume that $d \geq 3$. Let $\sphericalangle(x,y)$ denote the angle (spherical distance) between the points $x, y \in S^d$. Then clearly

$$W^{(2)}(x,y) = \frac{1}{2} - \frac{\sphericalangle(x,y)}{\pi}.$$

From this it is routine to verify that for any two points $x, y \in S^d$,

$$d_W(x,y) \geq \frac{2}{\sqrt{d+1}} \sphericalangle(x,y). \tag{9}$$

Let $\mathcal{P} = \{P_1, \dots, P_k\}$ be a weak Szemerédi partition of Ω for the function W, with error ε. Then by Theorem 2.14(a), there is a set $T \subseteq S^d$ with $\lambda(T) \leq (8\varepsilon)^{1/2}$ such that the diameter of $P_i \setminus T$ in the d_W metric is at most $(8\varepsilon)^{1/2}$ for every i. By (9), this implies that the diameter of $P_i \setminus T$ in spherical distance is at most $\sqrt{2\varepsilon(d+1)}$, and hence its measure satisfies $\lambda(P_i \setminus T) \leq (\sqrt{2\varepsilon(d+1)})^d \leq 2^{-d}$. Since the sets $P_i \setminus T$ $(i = 1, \dots, k)$ and T cover S^d, we get

$$k 2^{-d} + (8\varepsilon)^{1/2} \geq 1,$$

which implies that $k \geq 2^{d-1}$. ∎

Thus for a given $\varepsilon > 0$, we get weak Szemerédi partitions with error at most ε with $2^{2/\varepsilon^2}$ classes, and for some functions we need at least $(1/4)2^{1/(8\varepsilon)}$ classes. It is not clear whether the best threshold has $1/\varepsilon$ or $1/\varepsilon^2$ in the exponent.

As a second application of Theorem 2.14, we sketch a (somewhat surprising) algorithm to construct a weak Szemerédi partition.

Let $W \in \mathcal{W}_0$ and $\varepsilon^2/5 > 0$. Set

$$m = \left\lceil \frac{80}{\varepsilon^2} \ln \frac{80}{\varepsilon^2} \right\rceil 2^{\lceil 10^{12}/\varepsilon^{16} \rceil}.$$

Choose independent uniform random points $X_1, \dots, X_m$ from $[0,1]$. Let $S_1, \dots, S_m$ be the Voronoi cells of these points with respect to the metric δ_W; in other words, let $x \in S_i$ if x is closer to X_i

than to any other X_j; if there are more than one points X_j at minimum distance from x, then we assign x to that with smallest subscript. This way get a partition $\mathcal{S}(X_1,\dots,X_m)=\{S_1,\dots,S_m\}$ of $[0,1]$.

Theorem 2.18 *With probability at least* $3/4$*, the partition* $\mathcal{S}(X_1,\dots,X_m)$ *is a weak Szemerédi partition with error at most* ε*.*

We have described this algorithm as applied to a function $W\in\mathcal{W}_0$, but it is straightforward to modify it so that it applies to a graph G. Our algorithm gives a larger number of classes than that of Frieze and Kannan [39], and it is also slower (primarily because of the cost of squaring the adjacency matrix at the beginning). Our purpose with this formulation is to illuminate this geometric connection.

Proof. Let $k=2^{\lceil 10^{12}/\varepsilon^{16}\rceil}$. By Corollary 2.15, there is a partition $\{T_0,T_1,\dots,T_k\}$ of $[0,1]$ into $k+1$ measurable sets such that $\lambda(T_0)\le\varepsilon^2/10$ and for $1\le i\le k$, T_i has diameter at most $\varepsilon^2/10$ in the d_W metric. let $\alpha_i=\lambda(T_i)$. Let I be the set of those indices $i\in\{1,\dots,k\}$ for which T_i contains at least one sample point X_j (we don't care whether T_0 contains a sample point). Then

$$\mathsf{E}\big(\lambda(\cup_{i\notin I}T_i)\big)=\sum_{i=1}^{k}\alpha_i(1-\alpha_i)^m.$$

To estimate this sum, let $c=\varepsilon^2/(80k)$. Then we have

$$\begin{aligned}\mathsf{E}\big(\lambda(\cup_{i\notin I}T_i)\big)&=\sum_{i:\ \alpha_i\le c}\alpha_i(1-\alpha_i)^m+\sum_{i:\ \alpha_i>c}\alpha_i(1-\alpha_i)^m\le ck+(1-c)^m\\&\le\frac{\varepsilon^2}{80}+e^{-\varepsilon^2m/(80k)}\le\frac{\varepsilon^2}{80}+\frac{\varepsilon^2}{80}=\frac{\varepsilon^2}{40}.\end{aligned}$$

So with probability at least $3/4$, we have $\lambda(\cup_{i\notin I}T_i)\le\varepsilon^2/10$. In such a case, the set $S=\cup_{i\in I}T_i\cup T_0$ has measure $\lambda(S)\le\varepsilon^2/5$.

We claim that for $j=1,\dots,m$, the diameter of $S_j\setminus S$ is at most $\varepsilon^2/5$. It suffices to prove that $d_W(x,X_j)\le\varepsilon^2/10$ for every point $x\in S_j\setminus S$. Indeed, there is an $i\in I$ such that $x\in T_i$. The set T_i has diameter at most $\varepsilon^2/10$, and (since $i\in I$) there is a sample point $X_h\in T_i$. Thus $d_W(x,X_h)\le\varepsilon^2/10$, but since x belongs to the Voronoi cell of X_j, it follows that $d_W(x,X_j)\le\varepsilon^2/10$.

We are done by Theorem 2.14(b). ■

3 Local-global limits of bunded degree graphs

The colored neighborhood metric for sparse graphs was introduced by Bollobás and Riordan [14]. The corresponding convergence notion refines a convergence notion introduced by Benjamini and Schramm [11]. We prove that even in this refined sense, the limit of a convergent graph sequence (with uniformly bounded degree) can be represented by a graphing. We study various topics related to this convergence notion such as: Bernoulli graphings, factor of i.i.d processes and hyperfiniteness.

Quite interestingly (or disturbingly) there is no unified theory of graph convergence. Instead there are various convergence notions that work well in different situations. For example the theory of dense graph limits works well if the number of edges is quadratic in the number of vertices but it trivializes for graphs that are sparser than that. On the other hand the Benjamini–Schramm limit [11] is only defined for graphs which have a linear number of edges in terms of the vertices. In the regime between linear and quadratic the situation gets even more complicated.

In this part of the thesis we focus on the very sparse case were graphs have degrees bounded by some fixed number d (which we consider as fixed throughout). According to Benjamini and Schramm, a graph sequence $\{G_n\}_{n=1}^{\infty}$ is convergent if the distribution of the isomorphism types of neighborhoods of radius r (when a vertex is chosen uniformly at random in G_n) converges for every fixed r. This notion of convergence is called *local convergence*, *weak convergence* or *Benjamini–Schramm convergence*.

The following example illustrates why a different, stronger notion of convergence is needed in some cases. For odd n, let G_n be a d-regular expander graph on n nodes. For even n, let G_n be the disjoint union of two d-regular expander graphs on $n/2$ nodes. Assume that the girth of G_n tends to infinity. Then the sequence G_n is locally convergent, but clearly even and odd members of the sequence are quite different, and it would be desirable to distinguish them.

Bollobás and Riordan [14] introduced such a finer convergence notion (i.e., fewer sequences are convergent). A graph sequence $\{G_n\}_{n=1}^{\infty}$ is convergent in this sense if for every $r, k \in \mathbb{N}$ and $\epsilon > 0$ there is an index l such that if $n, m > l$ then for every coloring of the vertices of G_n with k colors the distribution of colored neighborhoods of radius r can be approximated with error at most ϵ by the colored neighborhood statistics of another coloring of the vertices of G_m. This is equivalent to saying that G_n and G_m are close in the colored neighborhood metric introduced in [14]. This finer convergence is sensitive to both local and global properties of the graphs, whereas the Benjamini–Schramm convergence is only sensitive to local properties. For this reason we call this notion *local-global convergence*.

Benjamini and Schramm described a limit object for locally convergent sequences in the form of an involution-invariant distribution on rooted countable graphs with bounded degree. One can also describe this limit object as a *graphing* (Elek [30]), which is a bounded degree graph on a Borel probability space such that the edge set is Borel measurable and it satisfies a certain measure preservation property (we will give a precise definition below). Neighborhood statistics in graphings can be defined by using the probability space structure on the vertex set. Every involution-invariant distribution can be represented by a graphing. We note that graphings are common generalizations of bounded degree graphs and measure preserving systems and so they are also interesting from an ergodic theoretic point of view.

However, the graphing representing the limit object of a locally convergent graph sequence is not unique: quite different graphings can describe the same involution-invariant distribution. In other words, a graphing contains more information that just the limiting neighborhood distribution. This suggests that graphings can be used to represent limit objects for more refined convergence notions. Indeed, we will show that the limit of a local-global convergent sequence can also be represented by a graphing in the sense that the graphs converge to the graphing in the colored neighborhood metric. This means that for every local-global convergent sequence we produce a graphing which contains both local and global information about the graphs.

We highlight the importance of a special family of graphings called *Bernoulli graphings*. We show that with given local statistics the Bernoulli graphings contains the least global information. This means that the global properties of a Bernoulli graphing can be modeled with an arbitrary precision on any other graphing with the same local statistics. Graph sequences that converge to Bernoulli graphings will be called *Bernoulli sequences*. For a graph G, Being close to a Bernoulli graphing in the colored neighborhood metric, means that the local statistics of a coloring on G can be modeled by a randomized process called *local algorithm* or *factor of i.i.d process*.

Roughly speaking a hyperfinite graph sequence is a bounded degree sequence whose members can be cut into small connected components using a small set of vertices (or equivalently edges). We prove that a locally convergent hyperfinite sequence is a local-global convergent Bernoulli sequence. It is an interesting question how to construct non hyper-finite sequences that are Bernoulli. A good candidate is a growing sequence of random d-regular graphs however this is a hard open problem. Describing the local-global limits of random d-regular graphs would give a deep understanding of their structure, and it seems to be one of the most interesting problems in this topic. We describe a few related conjectures.

3.1 Local-Global convergence of bounded degree graphs

A *rooted graph* is a pair (G, o) where o is a vertex of a graph G. The *radius* of a rooted graph is the distance of the farthest vertex in G to o. We denote by U^r the set of all rooted graphs with radius r (and all degrees bounded by d). For an integer $r > 0$, and a vertex v in a graph G, let $N_{G,r}(v)$ denote the subgraph of G rooted at v and induced by the vertices that are at a distance at most r from v. Two rooted graphs (G, o) and (G', o') are said to be *isomorphic* if there is an isomorphism from G to G' that maps o to o'.

Given a finite graph G and a radius $r \geq 1$, we can choose a node $\mathbf{v} \in V(G)$ uniformly and randomly, and consider the distribution of $N_{G,r}(\mathbf{v})$. Let $P_{G,r}$ denote this probability measure on U^r. We say that a sequence (G_n) of finite graph is *locally convergent* (or Benjamini–Schramm convergent), if $P_{G_n,r}$ converges to a limit distribution as $n \to \infty$, for every fixed $r \geq 1$.

Note that since U^r is finite, all the usual distances on $M(U^r)$ are topologically equivalent. Most of the time we work with the *total variation distance* d_{var}, defined (in general, for a space X) by

$$d_{\mathrm{var}}(\mu, \nu) = \sup_{A \subseteq X} |\mu(A) - \nu(A)|$$

where A runs through the Borel measurable sets. Denote the set of probability measures on a Borel space X by $M(X)$.

To define our refinement of local convergence, we consider node colorings. For a finite graph G let $K(k, G)$ denote the set of all vertex colorings with k colors. Fix integers k and r, and let $U^{r,k}$ be the set of all triples (H, o, c) where (H, o) is a rooted graph of radius at most r and c is an arbitrary k-coloring of $V(H)$. Consider a finite graph G together with an $c \in K(k, G)$. Pick a random vertex v from G, then the restriction of the k-coloring to $N_{G,r}(v)$ is an element in $U^{r,k}$, and thus for the graph G, every $c \in K(k, G)$ introduces a probability distribution on $U^{r,k}$ which we denote by $P_{G,r}[c]$. Let

$$Q_{G,r,k} := \big\{ P_{G,r}[c] : \ c \in K(k, G) \big\} \subseteq M(U^{r,k}).$$

These sets are similar to “quotient sets” introduced in [22] for dense graphs, except that there only edges with given coloring were counted, not larger neighborhoods. Notice that the sets $Q_{G,r,k}$ are finite, and they are subsets of the finite dimensional space $\mathbb{R}^{U^{r,k}}$ that is independent of the graph G.

Definition 3.1 A sequence of finite graphs $\{G_n\}_{n=1}^{\infty}$ with all degrees at most d is called *locally-globally convergent* if for every $r, k \geq 1$, the sequence $(Q_{G_n,r,k})_{n=1}^{\infty}$ converges in the Hausdorff distance inside the compact metric space $(M(U^{r,k}), d_{\mathrm{var}})$.

Since compact subsets of a compact metric space form a compact space with respect to the Hausdorff metric, it follows that every infinite sequence of finite graphs contains a locally-globally convergent subsequence.

We could have fixed $k = 1$, to get a metric definition of Benjamini–Schramm convergence. We don't know whether $k = 2$ would give a convergence notion equivalent to local-global convergence.

It is natural to ask if we obtain a different convergence notion if we replace vertex colorings by edge colorings or other locally defined extra structures. However, it turns out that they can all be encoded by vertex colorings and thus they don't lead to different convergence notions. As an example, we show how to encode edge colorings by vertex colorings.

Let G be a graph with all degrees at most d and let $c: E(G) \to [k]$ be an edge coloring of G. First we can create a new edge coloring $c_2: E(G) \to [30d^3k]$ such that $c_2(e) \equiv c(e)$ modulo k for every $e \in E(G)$ and if $c_2(e_1) = c_2(e_2)$ holds then the edges e_1 and e_2 are of distance at least 3 in the edge graph of G. It is clear that c_2 encodes the coloring c in the sense that local statistics of c_2 modulo k give the local statistics of c. Let S denote the set of subsets of $[30d^3k]$ of size at most d. We create the vertex coloring $c_2: V(G) \to S$ in the way that $c(v)$ is the set of colors of the edges incident to v in c_2. Now it is easy to see that c_3 encodes the coloring c_2 in the following way. If $e = (v, w)$ is an edge in G then $c_2(e)$ is the intersection of the sets $c_3(v)$ and $c_3(w)$.

3.2 Involution-invariant measures and graphings

Benjamini and Schramm [11] associated a limit object with every locally convergent graph sequence as follows. Let $\mathfrak{G}$ denote the set of (isomorphism classes of) rooted, connected (possibly infinite) graphs with all degrees at most d. For a rooted graph (B, o) with radius r, we denote by $\mathfrak{G}(B, o)$ the set of all rooted graphs (G, o) such that $N_{G,r}(o) \cong (B, o)$. For a rooted graph (G, o), we define a neighborhood basis as $\mathfrak{G}(N_{G,r}(o))$. These neighborhoods define a topology on $\mathfrak{G}$. It is easy to see that this is a compact separable space.

The Benjamini–Schramm limit of the locally convergent graph sequence (G_n) is a probability measure ν on the Borel sets of $\mathfrak{G}$, such that

$$P_{G,r}(B, o) \to \nu(\mathfrak{G}(B, o))$$

for every $r \geq 1$ and every r-ball (B, o).

Not every probability measure on $\mathfrak{G}$ arises as the limit of a convergent graph sequence. One property that all limits have is called *involution invariance* or *unimodularity*. To define this, let $\tilde{\mathfrak{G}}$ denote the space of graphs in $\mathfrak{G}(C)$ with a distinguished edge incident to the root. We define continuous transformation $\alpha: \tilde{\mathfrak{G}}(C) \to \tilde{\mathfrak{G}}(C)$ such that α moves the root to the other endpoint of the distinguished edge. If μ is any probability measure on $\mathfrak{G}(C)$, then let μ^* be the unique probability measure such that $d\mu^*/d\mu(G)$ is proportional to the degree of the root in G. We define the measure

$\tilde{\mu}$ on $\tilde{\mathfrak{G}}(C)$ such that on a μ^*-random graph we distinguish a random edge incident to the root. The measure μ is called *involution-invariant* if $\tilde{\mu}$ is invariant under α. Involution-invariant measures on $\mathfrak{G}(C)$ form a closed set in the weak topology.

It is easy to see that if μ is a measure on $\mathfrak{G}$ that is the limit of finite graphs, then it is involution-invariant. It is not known whether all involution-invariant measures arise as graph limits.

In the dense setting, the set of the symmetric measurable maps $w : [0,1]^2 \to [0,1]$ were used to generalize the concept of graphs and describe graph limits. For local-global convergence of Definition 3.1, graphings serve this purpose.

Definition 3.2 Let X be a Polish topological space and let ν be a probability measure on the Borel sets in X. A *graphing* is a graph $\mathcal{G}$ on $V(\mathcal{G}) = X$ with Borel measurable edge set $E(\mathcal{G}) \subset X \times X$ in which all degrees are at most d and

$$\int_A e(x,B)d\nu(x) = \int_B e(x,A)d\nu(x) \tag{10}$$

for all measurable sets $A, B \subseteq X$, where $e(x,S)$ is the number of edges from $x \in X$ to $S \subseteq X$.

If condition (10) holds, then $\eta^*(A \times B) = \int_A e(x,B)d\nu(x)$ defines a measure on the Borel sets of $X \times X$ that is concentrated on $E(\mathcal{G})$, symmetric in the two coordinates, and its marginal ν^* satisfies $(d\nu^*/d\nu)(x) = \deg(x)$. Normalizing by $d_0 = \int_X \deg(x)\,dx$, we get a probability distribution η on the set of edges. We can generate a random edge from η by selecting a random point v from ν^* and selecting uniformly a random edge incident with v. Conversely, if we generate a random oriented edge this way, and the distribution that is obtained is invariant under flipping the orientation, then (10) follows by Fubini's Theorem. Note that every finite graph G is a graphing where $X = V(G)$ and $\nu_{\mathcal{G}}$ is the uniform distribution on $V(G)$.

Let $\mathcal{G}$ be a graphing (of degree at most d) on the probability space (X, ν). Then it induces a measure $\mu_{\mathcal{G}}$ on $\mathfrak{G}$: pick a random element $x \in X$ and take its connected component G_x rooted by x. It is easy to see that $\mu_{\mathcal{G}}$ is an involution-invariant measure (in fact, (10) expresses just this property).

Let $\mathcal{G}$ be a graphing as in definition 3.2. A *vertex coloring* of $\mathcal{G}$ with k colors is a measurable function $c :\ X \to [k]$. The set of all colorings with k colors will be denoted by $K(k, \mathcal{G})$. We define $P_{\mathcal{G},r}[c]$ and $Q_{\mathcal{G},r,k}$ in a similar way as in a finite graph. (Notice that it makes sense to talk about a random vertex in $\mathcal{G}$.) The set $Q_{\mathcal{G},r,k}$ is a subset of the finite dimensional space $\mathbb{R}^{U^{r,k}}$, but in general it is infinite and not closed; we will often use its closure $\overline{Q}_{\mathcal{G},r,k}$

Now we are ready to state our main theorem.

Theorem 3.3 *Let $\{G_i\}_{i=1}^{\infty}$ be a local-global convergent sequence of finite graphs with all degrees at most d. Then there exists a graphing $\mathcal{G}$ such that $Q_{G_n,r,k} \to Q_{\mathcal{G},r,k}$ $(n \to \infty)$ in Hausdorff distance for every r and k.*

To what degree is the limit object determined? This question leads to different notions of "isomorphism" between graphings.

Definition 3.4 (a) Two graphings $(\mathcal{G}_1, X_1, \nu_1)$ and $(\mathcal{G}_2, X_2, \nu_2)$ are called *locally equivalent* if for every $r \in \mathbb{N}$ the distribution of $N_{\mathcal{G}_1,r}(x_1)$ is the same as the distribution of $N_{\mathcal{G}_2,r}(x_2)$ for random $x_1 \in X_1$ and $x_2 \in X_2$.

(b) We say that these graphings are *locally-globally equivalent* if $\overline{Q}_{\mathcal{G}_1,r,k} = \overline{Q}_{\mathcal{G}_2,r,k}$ for every $r,k \in \mathbb{N}$.

Local equivalence of two graphings means that they induce the same involution-invariant measure on $\mathfrak{G}$. Local-global equivalence implies local equivalence by setting $k = 1$.

Definition 3.5 (Local-global partial order) Assume that $\mathcal{G}_1$ and $\mathcal{G}_2$ are two graphings of maximal degree at most d. We say that $\mathcal{G}_1 \prec \mathcal{G}_2$ if $\overline{Q}_{\mathcal{G}_1,r,k} \subseteq \overline{Q}_{\mathcal{G}_2,r,k}$ for every $r,k \geq 1$. In particular, $\mathcal{G}_1$ and $\mathcal{G}_2$ are locally-globally equivalent if and only if both $\mathcal{G}_1 \prec \mathcal{G}_2$ and $\mathcal{G}_2 \prec \mathcal{G}_1$ hold.

An easy way to prove a relation $\mathcal{G}_1 \prec \mathcal{G}_2$ between two graphings is the following. We call a measure preserving map $\phi : V(\mathcal{G}_1) \to V(\mathcal{G}_2)$ a *local isomorphism*, if restricted to any connected component of $\mathcal{G}_1$, we get an isomorphism with one of the connected components of $\mathcal{G}_2$. Clearly local isomorphisms can be combined. (However, a local isomorphism may not be invertible!) It is easy to see that the existence of a local isomorphism $\mathcal{G}_1 \to \mathcal{G}_2$ implies that $\mathcal{G}_1$ and $\mathcal{G}_2$ are locally equivalent, and $\mathcal{G}_2 \prec \mathcal{G}_1$.

We are going to study local-global equivalence and local-global partial order in Section 3.6. In particular, we will show that among all graphings in a local equivalence class, there is always a smallest and a largest in this partial order.

We conclude this chapter with a few remarks.

Directed graphings. Let X, ν be as in 3.2 and let $E(\mathcal{G}) \subset X \times X$ be the edge set of a directed Borel graph $\mathcal{G}$ of bounded degree. For two set $A, B \subset X$ let $e(S,T) = |E(\mathcal{G}) \cap A \times B|$ denote the number of directed edges from A to B (This quantity may be infinite). Then $\mathcal{G}$ is called a (directed) graphing if

$$\int_A e(x,B)\, d\nu = \int_B e(A,x)\, d\nu$$

holds for any two measurable sets A, B.

Some examples. The simplest example for a directed graphing comes from ergodic theory. Let $T : X \to X$ be a measure preserving transformation which has a measure preserving inverse. Then the graph $\{(x, T(x))|x \in X\}$ is a directed graphing. More specifically let θ be an irrational number and $T(x) = x+\theta$ on the circle group $\mathbb{R}/\mathbb{Z}$. Then $\{(x, x+\theta)|x \in \mathbb{R}/\mathbb{Z}\}$ is an ergodic directed graphing. A similar but undirected graphing on $\mathbb{R}/\mathbb{Z}$ is given by the edge set $\{(x \pm \theta)|x \in \mathbb{R}/\mathbb{Z}\}$.

Decomposition into maps. The following construction can be used to verify the graphing axiom (10) in some cases. Let (X, ν) be a measure space. A measure preserving equivalence between two measurable sets $A, B \subset X$ is a map $\psi : A \to B$ which is measure preserving and has a measure preserving inverse. A partial measure preserving equivalence between A and B is a measure preserving equivalence between $A' \subset A$ and $B' \subset B$. Let $X = X_1 \cup X_2 \cup \cdots \cup X_n$ be an almost disjoint (intersections have 0 measure) decomposition of X and for each pair $1 \leq i < j \leq n$ let $\psi_{i,j}$ be a partial measure preserving equivalence between X_i and X_j. Then the symmetrized version of $E = \cup_{1 \leq i < j \leq n}\{(x, \psi_{i,j}(x))|x \in X_i\}$ is the edge set of a graphing. In is not hard to prove that each graphing has such a decomposition. In fact any measurable coloring of the vertex set in which vertices of the same color are of distance at least 3 yields such a decomposition. (The existence of such a coloring follows from results of Kechris, Solecki and Todorcevic [56].) This construction gives an upper bound of $O(d^2)$ for the number of maps. With more care (considering the line graph of $\mathcal{G}$), one can reduce this to $2d-1$. It is not known whether $d+1$ maps would suffice. (We come back to this issue later.)

3.3 Local limits of decorated graphs

In this section we extend the formalism behind the Benjamini–Schramm limits for the case when vertices are decorated by elements from a compact space. Let C be a second countable compact Hausdorff space. Let $\mathfrak{G}(C)$ denote the space of (isomorphism classes of) rooted, connected countable graphs with all degrees at most d such that the vertices are decorated by elements from C; so the points of $\mathfrak{G}(C)$ are triples (G, o, w), where G is a connected countable graph, $o \in V(G)$, and $w: V(G) \to C$. If C is the trivial (one point) compact space, then $\mathfrak{G}(C)$ can be identified with the space $\mathfrak{G}$ defined earlier. Two important special cases for us will be when $C = [0,1]$ (weighting by real numbers in $[0,1]$), and $C = [k]$ (coloring by k colors). With a slight abuse of notation, these will be denoted by $\mathfrak{G}[0,1]$ and $\mathfrak{G}[k]$.

We put a compact topology on $\mathfrak{G}(C)$ by specifying a basis of it. Let r be an arbitrary natural number and (H, o) be a finite rooted graph of radius r. Assume furthermore that every vertex v of (H, o) is decorated by an open set U_v in C. Let S be the collection of all $(G, o) \in \mathfrak{G}(C)$ where the neighborhood $N_{G,r}(o)$ is isomorphic to (H, o), and furthermore there is an isomorphism $\alpha: N_{G,r}(o) \to (H, o)$ such that the decoration of v is contained in $U_{\alpha(v)}$ for every $v \in V(G)$. It is easy to see that $\mathfrak{G}(C)$ with this topology is a compact, second countable, Hausdorff space. As a consequence, probability measures on $\mathfrak{G}(C)$ form a compact space in the weak topology.

If G is a finite graph with all degrees at most d in which the vertices are C-labeled, then we can construct a probability measure μ_G on $\mathfrak{G}(C)$ by putting a root o on a randomly chosen vertex $v \in V(G)$ and keeping only the connected component of the root. If $\{G_i\}_{i=1}^{\infty}$ is a sequence of C-labeled graphs then we say that they are *locally convergent* if the corresponding measures $\{\mu_{G_i}\}_{i=1}^{\infty}$ converge in the weak topology to some measure μ. The measure μ is the limit object of the sequence.

We define involution-invariance completely analogously to the undecorated case, simply replacing $\mathfrak{G}$ by $\mathfrak{G}(C)$ everywhere. Involution-invariant measures on $\mathfrak{G}(C)$ form a closed set in the weak topology. It follows that if μ is a measure on $\mathfrak{G}(C)$ that is the limit of finite colored graphs, then it is involution-invariant.

Similarly as in the undecorated case, every C-decorated graphing defines an involution-invariant distribution. Let $\mathcal{G}$ be a graphing and $c: V(\mathcal{G}) \to C$, a Borel function with values in the compact space C. Then we can create a measure $\mu_{\mathcal{G},c}$ on $\mathfrak{G}(C)$ by picking a random element $x \in V(\mathcal{G})$ and then taking its connected component G_x rooted by x together with the vertex labels given by the restriction of c to $V(G_x)$. It is easy to see that $\mu_{\mathcal{G},c}$ is an involution-invariant measure.

We can define a Borel graph $\mathcal{G}(C)$ on $\mathfrak{G}(C)$. The edge set of this graph consists of pairs $((G, o_1), (G, o_2)) \in \mathfrak{G}(C) \times \mathfrak{G}(C)$ such that (o_1, o_2) is an edge in G. Note that loop edges can arise in this graph. For example if there is an automorphism of G which takes o_1 to o_2 then (G, o_1) is identified with (G, o_2) in $\mathfrak{G}(C)$. Every involution-invariant measure ν on $\mathfrak{G}(C)$ is a probability measure on the vertex set of $\mathcal{G}(C)$. This graph is not a graphing in general, because of the problem with automorphisms, which also lead to loops. It is not hard to show, however, that if with probability 1 a ν-random connected component has no automorphisms, then we get a graphing. One important role of appropriate decorations is to break symmetries and make this graph a graphing.

Let $f:\ \mathfrak{G}(C)\to\mathbb{R}$ be any function. We call f *local*, if there is a positive integer r such that $f(G,o)$ depends (measurably) on the decorated r-neighborhood $N_{G,r}(o)$. Clearly every local function is measurable. Local functions correspond to those functions on finite graphs that can be computed by a constant-time local algorithm.

3.4 A regularization lemma

The following lemma is the main ingredient in proving Theorem 3.3. It serves as a "regularity lemma" in our framework for bounded degree graphs.

Lemma 3.6 (Regularization) *For positive integers r,k and real number $\epsilon>0$, there exists an integer $t_{r,k,\epsilon}$ such that the following holds. For every graph G with all degrees at most d there exists a $t_{r,k,\epsilon}$-vertex coloring q of G such that*

- *If $q(v)=q(w)$, then either $v=w$ or the distance of v and w in G is at least $r+1$,*
- *For every $g\in K(k,G)$, there exists $\alpha:[t_{r,k,\epsilon}]\to[k]$ such that*

$$d_{\text{var}}(P_{G,r}[g],P_{G,r}[\alpha\circ q])\le\epsilon.$$

Proof. The space $M(U^{r,k})$ is a bounded dimensional compact set with the topology generated by d_{var}. Let N be an $\epsilon/2$-net in $M(U^{r,k})$ in d_{var}. Let N_G be the subset of points in N that are at most $\epsilon/2$ far from a point of the form $P_{G,r}[g]$ for some $g\in K(k,G)$. For each $n\in N_G$ we choose a representative $x_n=P(G,r,k)[g_n]$ such that $d_{\text{var}}(n,x_n)\le\epsilon/2$. It is clear that for every $g\in K(k,G)$ there is a point x_n such that $d_{\text{var}}(P(G,r,k)[g],x_n)\le\epsilon$. Let f be the common refinement of all the partitions $\{g_n\}_{n\in N_G}$. Clearly f has a bounded number of partition sets in terms of r,k and d and it satisfies the third condition.

Now we further refine f to satisfy the first condition. Let f' be a proper coloring of the graph G with $(d+1)^r$ colors in which every two vertices in distance at most r receive different colors. The common refinement q of f, f' satisfies the first condition. ∎

3.5 Proof of the main theorem

Now we introduce the space X which will serve as a universal Borel space for the limit graphings of sequences of finite graphs with all degrees at most d. Let $C=\prod_{k,r,n}[t_{r,k,1/n}]$ be the compact space with the product topology. We denote by X the compact space $\mathfrak{G}(C)$ and by $E\subset X\times X$ the edge set $\mathfrak{E}(C)$. Let $q:X\to C$ be the function such that $q((G,o))$ is the color of the root o. Furthermore for $r,k,n\in\mathbb{N}$ we define the coloring $q_{r,k,n}:X\to[t_{r,k,1/n}]$ as the composition of q with the projection to the coordinate (r,k,n) in C.

Let $\{G_i\}_{i=1}^{\infty}$ be a local-global convergent sequence of graphs with all degrees at most d. For each G_i and triple $(r,k,n)\in\mathbb{N}^3$ we chose a coloring $q^i_{r,k,n}:V(G_i)\to[t_{r,k,1/n}]$ guaranteed by lemma 3.6. Let $q_i:\ V(G_i)\to C$ be defined as $\prod_{r,k,n}\{q^i_{r,k,n}(v)\}\in C$. As described in chapter 3.3, each graph

G_i together with the coloring q_i defines a probability measure μ_i on X by putting the root on a random vertex of G_i and keeping only the connected component of the root.

By choosing a subsequence from $\{G_i\}_{i=1}^{\infty}$ we can assume that the sequence $\{\mu_i\}_{i=1}^{\infty}$ weakly converges to a probability distribution μ on X. Our goal is to show that the Borel graph (X, E) with the measure μ is a graphing, which represents the local global limit of $\{G_i\}_{i=1}^{\infty}$.

Let us first observe that for a μ-random element (G, o) in (X, μ) with probability one the vertex labels $V(G) \to C$ are all different. This follows from the fact that the colorings $q^i_{r,k,n}$ separate points in G_i that are closer than $r+1$, and this property is preserved by the Benjamini–Schramm limit. This means that if $v, w \in V(G)$ are of distance r then their colors projected to the coordinate (r, k, n) (where k, n are arbitrary) are different.

Lemma 3.7 *The measurable graph (X, E, μ) is a graphing.*

Proof. Let us introduce the measures $\{\eta_i^*\}_{i=1}^{\infty}$, similarly as in section 3.2, by

$$\eta_i^*(A \times B) = \int_A e(x, B) d\mu_i(x),$$

(where $A, B \subseteq X$ are measurable, and $e(x, B)$ is the number of edges $(x, y) \in E$ with $y \in B$). We define η^* analogously. Assume that $A, B \subset X$ are open-closed sets. The weak convergence of $\{\mu_i\}_{i=1}^{\infty}$ implies that $\lim_{i\to\infty} \eta_i^*(A \times B) = \eta^*(A \times B)$ and $\lim_{i\to\infty} \eta_i^*(B \times A) = \eta^*(B \times A)$. On the other hand it is clear that for every i we have $\eta_i^*(A \times B) = \eta_i^*(B \times A)$ since both are equal (up to normalization by $|V(G)|$) to the number of edges between the sets $\{v|(G_i, v) \in A\}$ and $\{v|(G_i, v) \in B\}$. (Here we use that the vertex labels in G_i are all different.) We obtain that $\eta^*(B \times A) = \eta^*(A \times B)$. Since such product sets generate the whole σ-algebra on $X \times X$, the proof is complete. ∎

Lemma 3.8 *The probability distributions $P_{G_i,r}[q^i_{r,k,n}]$ converge to $P_{\mathcal{G},r}[q_{r,k,n}]$ for every fixed triple $r, k, n \in \mathbb{N}$ where $k' = t_{r,k,1/n}$.*

Proof. For every point $G = (G, o) \in X$ we can associate another rooted graph G' which is the connected component of G in the graphing $\mathcal{G}$ rooted by G. There is a natural vertex coloring on G' which is the restriction of the function q to the vertices of G' and so we can regard G' as an element in X. We claim that G is isomorphic (in a root and label preserving way) to G' with probability one. We use that with probability one all the vertex labels of G are different. For such graphs G, the map given by $v \mapsto (G, v)$ defines a decoration-preserving isomorphism between G and G'. (The fact that the vertex labels in G are all different guarantees that the map is one to one.)

Now the claim implies that the probability distribution $P_{\mathcal{G},r}[q_{r,k,n}]$ is the same as the distribution of $N_{G,r}(o)$ where the vertex labels are projected to the coordinate (r, k, n) and (G, o) is a μ random element in X. Using the local convergence of $\{\mu_i\}_{i=1}^{\infty}$ the proof is complete. ∎

Lemma 3.9 *For every $r, k, n \in \mathbb{N}$ there is an index j such that for every $i \geq j$ and $c \in K(k, G_i)$ there is a coloring c' of X such that $d_{\mathrm{var}}(P_{G_i,r}[c] - P_{\mathcal{G},r}[c']) \leq 1/n$.*

Proof. Let $k' = t_{r,k,1/(2n)}$. By lemma 3.8 there is an index j such that

$$d_{\mathrm{var}}(P_{G_i,r,k'}[q^i_{r,k,2n}], P_{\mathcal{G},r,k'}[q_{r,k,2n}]) \leq 1/(2n) \tag{11}$$

for every index $i \geq j$. Let $i \geq j$ be arbitrary and let $c \in K(k, G_i)$ be an arbitrary coloring. Then by lemma 3.6 there is a map $\alpha : [t_{r,k,1/(2n)}] \to [k]$ such that

$$d_{\mathrm{var}}(P_{G_i,r}[\alpha \circ q^i_{r,k,2n}], P_{G_i,r}[c]) \leq 1/(2n).$$

On the other hand the definition of the total variation distance and (11) guarantee that

$$d_{\mathrm{var}}(P_{G_i,r}[\alpha \circ q^i_{r,k,2n}], P_{\mathcal{G},r}[\alpha \circ q_{r,k,2n}]) \leq 1/(2n).$$

It shows that $\alpha \circ q_{r,k,2n}$ is a good choice for c'. ∎

Lemma 3.10 *For every coloring $c \in K(k, \mathcal{G})$ and natural number n there is an index j such that for every $i \geq j$ there is a coloring $c' \in K(k, G_i)$ such that $d_{\mathrm{var}}(P_{G_i,r}[c'] - P_{\mathcal{G},r}[c]) \leq 1/n$.*

Proof. Let $c : X \to [k]$ be a Borel coloring. Then for every $\epsilon > 0$ there is another coloring $c_\epsilon : X \to [k]$ which is continuous and $|\mu(c^{-1}(i)) - \mu(c_\epsilon^{-1}(i))| \leq \epsilon$ holds for all $1 \leq i \leq k$. There is an index j_ϵ such that if $a \geq j_\epsilon$ then $|\mu_a(c_\epsilon^{-1}(i)) - \mu(c^{-1}(i))| \leq \epsilon$ holds for all $1 \leq i \leq k$. Each coloring c_ϵ induces a coloring f^i_ϵ on G_i such that the color of a vertex $v \in V(G_i)$ is the c_ϵ color of the rooted graph $(G_i, v) \in X$. It is easy to see that $P_{G_i,r}[f^\epsilon_i]$,$P_{\mathcal{G},r}[c]$ and $P_{\mathcal{G},r}[c_\epsilon]$ are arbitrarily close to each other if ϵ is small enough and $i \geq j_\epsilon$. This completes the proof. ∎

3.6 Bernoulli graphings and Bernoulli graph sequences.

Probably the most fundamental graphing construction is the Bernoulli graphing corresponding to an involution invariant measure. These graphings are closely related to factor of i.i.d processes and local algorithms. In this chapter we explain their role in local-global convergence.

Definition 3.11 (Bernoulli graphings) Let μ be an involution invariant measure on $\mathfrak{G}$. Let ν be the probability measure on $\mathfrak{G}[0,1]$ produced by putting independent random elements from $[0,1]$ on the nodes of a μ-random graph. The graph $(\mathfrak{G}[0,1], \mathfrak{E}[0,1])$ with the measure ν will be called the *Bernoulli graphing* corresponding to μ, and denoted by $\mathcal{G}_\mu$.

It is not hard to see that $\mathcal{G}_\mu$ is a graphing and it represents the involution-invariant distribution μ (Elek [30]).

Remark 3.12 Perhaps it would be more natural to decorate the nodes of the μ-random graph by independent bits, or more generally, by colors from $[k]$ for some fixed $k \geq 2$. This would yield an involution-invariant distribution on $\mathfrak{G}[k]$, but the graph $\mathfrak{G}[k]$, together with this distribution, would not necessarily form a graphing.

We define the Bernoulli graphing $\mathcal{G}_B$ corresponding to an arbitrary graphing $\mathcal{G}$ as the Bernoulli graphing defined by the involution-invariant distribution induced by $\mathcal{G}$. Clearly $\mathcal{G}$ and $\mathcal{G}_B$ are locally equivalent.

The simplest example for a Bernoulli graphing is provided by the involution invariant measure which is concentrated on a single d-regular rooted tree. Let T denote the rooted d-regular tree and let (X, ν) be the probability space in which we put independent random labels from $[0,1]$ on the vertices of T. Two points of X are connected in $\mathcal{G}$ if they can be obtained from each other by replacing the root to a neighboring vertex. It seems to be an interesting problem to decide if the sets $Q_{\mathcal{G},r}$ are all closed.

The following is a related construction. For every graphing $\mathcal{G}$ on the probability space (X,ν), we define its *Bernoulli lift* $\mathcal{G}^+$ as follows. The underlying set X^+ of $\mathcal{G}^+$ will be pairs (x,ξ), where $x \in X$ and $\xi :\ V(\mathcal{G}_x) \to [0,1]$. We connect (x,ξ) to (y,υ) if y is a neighbor of x and $\xi = \upsilon$ (note that if y is a neighbor of x, then $\mathcal{G}_x = \mathcal{G}_y$). We can generate a random element of Ω by picking a ν-random point $x \in X$, and then assigning independent random weights $\xi(u)$ to the nodes u of $\mathcal{G}_x$.

We define two maps $\phi :\ V(\mathcal{G}^+) \to V(\mathcal{G})$ and $\psi :\ V(\mathcal{G}^+) \to V(\mathcal{G}_B)$ by $\phi(x,\xi) = x$ and $\psi(x,\xi) = \big(\mathcal{G}_x,\xi\big)$. It is easy to check that the maps ϕ and ψ are local isomorphisms. This implies that graphing $\mathcal{G}$ is locally equivalent to its Bernoulli lift $\mathcal{G}^+$ as well as its Bernoulli graphing $\mathcal{G}_B$.

Our main goal in this section is to describe the relationship between $\mathcal{G}$, $\mathcal{G}_B$ and $\mathcal{G}^+$ from the point of view of local-global equivalence.

Theorem 3.13 *Every graphing is local-global equivalent to its Bernoulli lift.*

As a corollary of this theorem, we get that Bernoulli graphings are minimal elements in their local equivalence class:

Corollary 3.14 (Minimality of Bernoulli graphings) *For any graphing $\mathcal{G}$, we have $\mathcal{G}_B \prec \mathcal{G}$.*

Indeed, the relation $\mathcal{G}_B \prec \mathcal{G}^+$ follows immediately from the construction of the map $\psi :\ V(\mathcal{G}^+) \to V(\mathcal{G}_B)$ above.

To prove Theorem 3.13 we need a definition and a couple of lemmas.

Definition 3.15 (Quasirandom colorings) Let $\mathcal{G}$ be a graphing on the space (X,ν) and let $c :\ X \to [k]$ be a measurable coloring. Let $\mu_{r,k}$ be the probability distribution on $U^{r,k}$ obtained from ν by considering the r-neighborhood of a random element $x \in X$ and decorating its vertices by random independent elements from $[k]$. We say that c is (r,ϵ)-*quasirandom*, if $d_{\mathrm{var}}(P_{\mathcal{G},r}[c],\mu_{r,k}) \le \epsilon$.

Lemma 3.16 (Existence of quasirandom colorings) *Let $\mathcal{G}$ be a graphing on the space (X,ν). Then for every $k,r \in \mathbb{N},\epsilon > 0$ there is an (r,ϵ)-quasirandom coloring $c :\ X \to [k]$.*

Proof. Let $C = \{0,1\}^{\mathbb{N}}$ be the Cantor set with the uniform measure and let $\psi : X \to C$ be a measurable equivalence between the probability spaces X and C. Let $\pi_i : C \to \{0,1\}^i$ be the projection to the first i coordinates. The map π_i is measure preserving if we consider the uniform measure on $\{0,1\}^i$. Fix $k,r \in \mathbb{N}$, and let $g_i : \{0,1\}^i \to [k]$ be a uniform random coloring of $\{0,1\}^i$ with k colors. Our goal is to show that if i is big enough then with a large probability $g_i \circ \pi_i$ is (r,ϵ)-quasirandom.

Claim 1 *For every $\epsilon_1 > 0$ and $n \in \mathbb{N}$ there is an index j such that if $x_1,\dots,x_n \in X$ are independent ν-random points, then with probability $1-\epsilon_1$, the map π_j separates all the points in $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$.*

It is easy to see that $\pi = (\pi_1,\pi_2,\dots)$ separates the points of $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$ with probability one on X^n. Let Y_j denote the set of points $(x_1,x_2,\dots,x_n)$ in X^n for which π_j separates the points in

$\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$. We have that Y_j is an increasing chain of measurable sets such that $\nu(\cup_{i=1}^\infty Y_i) = 1$. This shows that for some index j we have $\nu(Y_j) > 1 - \epsilon_1$.

Let $x = (x_1, \ldots, x_n) \in X^n$ and let g be a k-coloring $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$. Let us choose a random $t \in [n]$ uniformly. Let as say that x is *representative* if the distribution of $N_{\mathcal{G},r}(x_t)$ for a random $t \in [n]$ is $\varepsilon/6$-close to the distribution μ_r. Let as say that (x, g) is *representative* if the distribution of the colored neighborhood $(N_{\mathcal{G},r}(x_t), g)$ is $\varepsilon/3$-close to the distribution $\mu_{r,k}$.

Let $x = (x_1, \ldots, x_n) \in X^n$ be chosen randomly and independently from the distribution ν. We note that with probability 1, the neighborhoods $N_{\mathcal{G},r}(x_i)$ are disjoint. If n is large enough, then (just by the Law of Large Numbers)

$$\mathsf{P}_x(x \text{ representative}) \geq 1 - \frac{\varepsilon}{6}.$$

Hence if g is a uniform random k-coloring of $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$, and n is large enough, then (by the Law of Large Numbers again), we have

$$\mathsf{P}_{x,g}((x, g) \text{ representative}) \geq 1 - \frac{\varepsilon}{3}.$$

Let us fix n so that this holds.

Next, using Claim I, we fix j so that (for a random x) π_j separates all the points in $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$ is probability at least $1 - \varepsilon/3$. Whenever this happens, the restriction of $\pi_j \circ g_j$ to $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$ is a uniform random k-coloring. In other words, we can generate a uniform random k-coloring of $\cup_{i=1}^n N_{\mathcal{G},r}(x_i)$ by restricting $\pi_j \circ g_j$ to it if π_j separates and randomly k-coloring it otherwise. Thus

$$\mathsf{P}_{x,g_j}((x, \pi_j \circ g_j) \text{ representative}) \geq \mathsf{P}_{x,g}((x, g) \text{ representative}) - \frac{\varepsilon}{3} \geq 1 - \frac{2\varepsilon}{3}.$$

It follows that there is at least one k-coloring g_j for which

$$\mathsf{P}_x((x, \pi_j \circ g_j) \text{ representative}) \geq 1 - \frac{2\varepsilon}{3}.$$

Let us fix such a g_j, then $c = \pi_j \circ g_j$ is an (r, ε)-quasirandom k-coloring of X. In fact, we can generate a random point of x by first generating n independent random points $x_1, \ldots, x_n$ and choosing one of them, x_t, uniformly at random. Then with probability at least $1 - 2\varepsilon/3$, $(x, \pi_j \circ g_j)$ is representative, and whenever this happens, then distribution of the colored neighborhood $(N_{\mathcal{G},r}(x_t), \pi_j \circ g_j)$ is $\varepsilon/3$-close to the distribution $\mu_{r,k}$. It follows that the total variation distance of $(N_{\mathcal{G},r}(x_t), \pi_j \circ g_j)$ from $\mu_{r,k}$, when x is also randomly chosen, is at most ε. ■

Our next lemma shows that we can approximate any measurable k-coloring of $\mathcal{G}^+$ by a k-coloring that is local and depends only on a discrete approximation of the nodeweights. To be precise, we define the (m, s)-*discretization* ($m, s \in \mathbb{N}$) as the map $\xi_{m,s} : X^+ \to U^{s,m}$, where $\xi_{m,s}(x)$ is obtained by considering the neighborhood $N_{\mathcal{G}^+,s}(x)$, and replacing every nodeweight $w(v)$ by $\lceil mw(v) \rceil$.

Lemma 3.17 *For every $r \geq 1$ and $\varepsilon > 0$, and every measurable k-coloring c of $\mathcal{G}^+$, there are positive integers s and m and a map $f : U^{s,m} \to [k]$ such that*

$$d_{\text{var}}(P_{\mathcal{G}^+,r}[c], P_{\mathcal{G}^+,r}[f \circ \xi_{s,m}]) \leq \varepsilon.$$

Proof. Let (X^+, ν^+) be the underlying space of $\mathcal{G}^+$. Let $\mathcal{K}$ denote the set of all bounded functions on X^+ that factor through an (m,s)-discretization for some $m, s \in \mathbb{N}$. These functions form a vector space, and the sets $\xi_{m,s}^{-1}(y)$ $(y \in U^{s,m})$ generate the Borel sets of X^+. Hence by the Monotone Class Theorem, the closure of $\mathcal{K}$ under pointwise convergence contains every measurable function on X^+.

In particular, there are positive integers s and m and a function $g : U^{s,m} \to [k]$ for which

$$\nu^+\Big\{x \in X^+ :\ |c(x) - g(\xi_{s,m}(x))| > \frac{1}{2}\Big\} < \frac{\varepsilon}{d^{r+1}}.$$

Rounding the values of g to the next integer, we get a function f for which

$$\nu\big\{x \in X^+ :\ c(x) \neq f(\xi_{s,m}(x))\big\} < \frac{\varepsilon}{d^{r+1}}.$$

For a random point $x \in X^+$, the probability that the colorings c and $f \circ \xi_{s,m}$ differ on any node in its r-neighborhood is less than ε. This implies the Lemma. ■

Now we are able to prove the main theorem in this section.

Proof of Theorem 3.13. Our goal is to approximate every element in $Q_{\mathcal{G}^+,r,k}$ by an element in $Q_{\mathcal{G},r,k}$ with arbitrary precision $\varepsilon > 0$. In other words, we want to construct, for every measurable k-coloring of $\mathcal{G}^+$, a measurable k-coloring of $\mathcal{G}$ that defines a similar distribution of colored neighborhoods.

We invoke Lemma 3.17 with $\varepsilon/2$ in place of ε to get two integers $s, m \geq 1$ and a map $f : U^{s,m} \to [k]$. Let q be an $(s, \varepsilon/2)$-quasirandom m-coloring of $V(\mathcal{G})$ guaranteed by lemma 3.16, and let $\mathcal{G}' = (\mathcal{G}, q) \in \mathfrak{G}[m]$. Consider the k-coloring $h = f \circ N_{\mathcal{G}',s}$. We claim that h has similar statistics as c:

$$d_{\rm var}(P_{\mathcal{G}^+,r}[c], P_{\mathcal{G},r}[h]) \leq \varepsilon$$

By the choice of f, it suffices to prove that

$$d_{\rm var}(P_{\mathcal{G}^+,r}[f \circ \xi_{s,m}], P_{\mathcal{G},r}[f \circ N_{\mathcal{G}',s}]) \leq \frac{\varepsilon}{2}.$$

This follows if we prove that the distributions of $\xi_{s,m}(y)$ (where y is a random point of $\mathcal{G}^+$) and $N_{\mathcal{G}',s}(x)$ (where x is a random point of $\mathcal{G}$) are close. But the distribution of $\xi_{s,m}(y)$ is just $\mu_{s,m}$, and the distribution of $N_{\mathcal{G}',s}(x)$ is $\varepsilon/2$-close to this by the quasirandomness of q. This completes the proof. □

The following fact shows another connection between a graphing and the associated Bernoulli graphing. We say that two graphings are *bi-locally isomorphic*, if there exists a third graphing that has local isomorphisms into both. The construction of the Bernoulli lift implies that every graphing is bi-locally isomorphic with its Bernoulli graphing. Since by the definition of the Bernoulli graphing, two graphings are weakly equivalent if and only if they have the same Bernoulli graphing, we get the following more explicit characterization:

Proposition 3.18 *Two graphings are locally equivalent if and only if they are bi-locally isomorphic.*

To prove this proposition, it suffices to show that bi-local isomorphism is a transitive relation. This takes some work, which we don't discuss here; for the details, we refer to [62].

Let us turn to graph sequences. Theorem 3.13 motivates the next definition.

Definition 3.19 (Bernoulli graph sequences) A graph sequence is called Bernoulli if it converges to a Bernoulli graphing in the local-global sense.

Note that the Bernoulli graphing to which a given Bernoulli sequence converges is fully determined by the local limit of the sequence. Bernoulli sequences are basically those which have the least possible global structure among sequences with the same local limit. The following provocative conjecture was popularized by the third author in the past few years. The considerable effort put into the topic shows that the solution may require a substantial novel idea.

Conjecture 3.20 (Limits of random d-regular graphs) Let d be a fixed natural number and let G_i be a random d regular graph on i vertices. Then $\{G_i\}_{i=1}^{\infty}$ is a Bernoulli sequence with probability one.

The motivation for this conjecture is that randomness destroys as much global structure as possible. Notice that if conjecture 3.20 is true then the limit object is the Bernoulli graphing produced from the d-regular tree. Even the next two weaker conjectures are unsolved.

Conjecture 3.21 A growing sequence of random d-regular graphs is local-global convergent with probability one.

Conjecture 3.22 For every d there is a Bernoulli graph sequence $\{G_i\}_{i=1}^{\infty}$ whose local limit is the d-regular tree.

3.7 Non-standard graphings

Let $\{G_i\}_{i=1}^{\infty}$ be an arbitrary graph sequence of maximum degree at most d. Let ω be a non-principal ultrafilter on $\mathbb{N}$. Let $\mathbf{G}$ denote the ultraproduct of the graph sequence. The vertex set $\mathbf{V}$ of $\mathbf{G}$ is the ultraproduct of the vertex sets V_i of G_i and the edge set $\mathbf{E} \subset \mathbf{V} \times \mathbf{V}$ is the ultra product of the edge sets $E_i \subset V_i \times V_i$ of G_i. It is clear that G_i has maximum degree at most d, since this property is expressible by a first order formula. We can also construct a σ-algebra $\mathcal{A}$ on $\mathbf{V}$ and a probability measure μ on $\mathbf{V}$ which is the ultralimit of the uniform distributions on the sets V_i. It is not hard to check that $\mathbf{G}$ satisfies the graphing axiom (10).

If $\{G_i\}_{i=1}^{\infty}$ is a locally convergent graph sequence, then $\mathbf{G}$ has neighborhood frequencies that are the limits of the neighborhood frequencies of the graphs G_i. If $\{G_i\}_{i=1}^{\infty}$ is locally-globally convergent, then $Q_{\mathbf{G},r,k}$ is the Hausdorff limit of the sets $Q_{G_i,r,k}$.

However, this does not directly prove Theorem 3.3, since $(\mathbf{V}, \mu)$ is not a separable probability space. One can complete the proof by choosing an appropriate separable sub-sigma-algebra of $\mathbf{G}$ which preserves the graphing structure. We omit the details here.

An attractive feature of ultralimit graphings is that the sets $Q_{\mathbf{G},r,k}$ are all closed. It is not clear if there is a standard graphing representation of the limit of a convergent sequence with this stronger property.

3.8 Hyperfinite graphs and graphings

For a graph G, we define $\tau_q(G)$ as the smallest t such that deleting t appropriate nodes, every connected component of the remaining graph has at most q nodes. We say that a sequence $(G_n)_{n=1}^{\infty}$ of finite graphs is (q,ε)-*hyperfinite*, if $\liminf_n \tau_q(G_n)/|V(G_n)| \le \varepsilon$. We say that $(G_n)_{n=1}^\infty$ is *hyperfinite*, if for every $\varepsilon > 0$ there is a $q \ge 1$ such that $(G_n)_{n=1}^\infty$ is (q,ε)-hyperfinite. We can define hyperfiniteness of a graphing $\mathcal{G}$ on underlying space Ω similarly: let $\tau_q(\mathcal{G})$ denote the infimum of numbers $a \ge 0$ such that we can delete a Borel set $S \subseteq$ with measure a so that every connected component of the remaining graphing has at most q nodes. We say that a graphing $\mathcal{G}$ is (q,ε)-*hyperfinite*, if $\tau_q(\mathcal{G}) \le \varepsilon$, and we say that $\mathcal{G}$ is *hyperfinite*, if for every $\varepsilon > 0$ there is a $q \ge 1$ such that $\mathcal{G}$ is (q,ε)-hyperfinite. Since we are talking about graphs with bounded degree, we could replace deleting nodes by deleting edges in the definitions above.

Hyperfiniteness in different settings was introduced by different people, see Kechris and Miller [55], Elek [31], Schramm [78]. Schramm proved that a weakly convergent sequence of graphs is hyperfinite if and only if its limit is hyperfinite. This does not hold for (q,ε)-hyperfiniteness for a fixed pair q and ε. (As an easy example, a sequence of random d-regular graphs tend to a limiting involution invariant distribution (concentrated on the infinite d-regular tree) that is $(1,1/2)$-hyperfinite, but the sequence is not.) On the other hand, a local-global convergent sequence of graphs behaves nicer:

Proposition 3.23 *Let a sequence $(G_n)_{n=1}^\infty$ of finite graphs converge to a graphing $\mathcal{G}$ in the local-global sense. Then $(G_n)_{n=1}^\infty$ is (q,ε)-hyperfinite if and only if $\mathcal{G}$ is (q,ε)-hyperfinite.*

Proof. A finite graph G satisfies $\tau_q(G) \le \varepsilon|V(G_n)|$ if and only if it has a 2-coloring c such that $P_{G,k,r}[c](c(\text{root}) = 1) \le \varepsilon$ and $P_{G,k,r}[c](B) = 0$ for every colored r-ball B that contains a connected all-blue subgraph with $k+1$ nodes. A graph $\mathcal{G}$ satisfies $\tau_q(\mathcal{G}) \le \varepsilon$ if and only if for every $\varepsilon' > \varepsilon$ it has a 2-coloring c such that $P_{G,k,r}[c](c(\text{root}) = 1) \le \varepsilon'$ and $P_{G,k,r}[c](B) = 0$ for every colored r-ball B that contains a connected all-blue subgraph with $k+1$ nodes. The proposition follows by the definition of local-global convergence to a graphing. ∎

The following important property of hyperfiniteness is closely related to the results of Schramm [78] and Benjamini, Schramm and Shapira [12]. It can be derived using the graph partitioning algorithm of Hassidim, Kelner, Nguyen and Onak [50]; a direct proof is given in [62].

Proposition 3.24 *Hyperfiniteness is invariant under local equivalence.*

Together with Proposition 3.23, the result of Schramm follows easily. We note that (q,ε)-hyperfiniteness for a fixed q and ε is not invariant under local equivalence, which is shown by the local-global limits of random d-regular graphs and of random d-regular bipartite graphs.

Our main result about hyperfinite graphings is a strengthening of Corollary 3.14.

Theorem 3.25 *Every hyperfinite graphing $\mathcal{G}$ is local-global equivalent to its Bernoulli graphing.*

Proof. By Proposition 3.24, the Bernoulli graphing $\mathcal{G}_B$ of a hyperfinite graphing $\mathcal{G}$ is also hyperfinite. By Corollary 3.14, $\mathcal{G}_B \prec \mathcal{G}$. It remains to show that $\mathcal{G} \prec \mathcal{G}_B$. In other words, for every coloring of $\mathcal{G}$ we have to find a coloring of $\mathcal{G}_B$ having almost the same local statistics.

Let (X,ν) be the underlying space of $\mathcal{G}$, let $c:\ X \to [k]$ be a measurable coloring, and let us fix a radius $r \in \mathbb{N}$ and an $\varepsilon > 0$. Let ν_B denote the measure of $\mathcal{G}_B$, and set $\varepsilon_1 = \varepsilon/(8(d+1)^r)$. Let $S \subset \mathfrak{G}[0,1]$ be a subset such that $\nu_B(S) \le \varepsilon_1$ and every connected component of $\mathfrak{G}[0,1]\setminus S$ has at most

n nodes. Let $m \in \mathbb{N}$, and define the coloring $b: \mathfrak{G}[0,1] \to [m] \times \{0,1\}$ by $b(x) = (\lceil mw(o) \rceil, \mathbf{1}_S(x))$ where $x = (G, o, w)$. Choosing m large enough, we may assume that the set of point x for $N_{\mathcal{G}_B,r}(x)$ contains another point with the same color has measure at most ε_1.

By Corollary 3.14, we have $\mathcal{G}_B \prec \mathcal{G}$, which implies that there is a coloring $b^*: X \to [m] \times \{0,1\}$ such that

$$d_{\mathrm{var}}(P_{\mathcal{G},n}[b^*], P_{\mathcal{G}_B,n}[b]) \leq \varepsilon_1. \tag{12}$$

It follows from (12) that there are subsets $T \subseteq \mathfrak{G}[0,1]$ and $T' \subseteq X$ with $\nu_B(T) = \nu(T') \leq 4\varepsilon_1$ such that all points of $\mathcal{G}_B \setminus T$ are contained in connected components that have at most n vertices and whose nodes are colored differently by b, and the same holds for $\mathcal{G}$, T' and b^*. Furthermore, for every $([m] \times \{0,1\})$-colored connected graph H with at most n vertices, the measure of points in components isomorphic to H (as colored graphs) is the same in $\mathcal{G}_B \setminus T$ and $\mathcal{G} \setminus T'$. Let V_H and V'_H be these two sets.

For every connected component C of $\mathcal{G} \setminus T'$ we can specify a "rule": a function $f_C: [m] \times \{0,1\} \to [k]$ such that $c = f_C \circ b^*$ on the nodes of C. This splits every set V'_H into at most k^{2m} measurable sets $V'_{H,f}$, that are unions of components of $\mathcal{G} \setminus T'$. We can split V_H into sets $V_{H,f}$ that are unions of components of $\mathcal{G}_B \setminus T$, so that $\nu_B(V_{H,f}) = \nu(V'_{H,f})$. Applying the rule f on the components in $V_{H,f}$, and coloring the points of T with one of the colors, we get a measurable k-coloring a of $\mathcal{G}_B$, for which

$$d_{\mathrm{var}}(P_{\mathcal{G},r}[c], P_{\mathcal{G}_B,r}[a]) \leq (d+1)^r \varepsilon_1 \leq \varepsilon.$$

This proves the theorem. ∎

Now we are ready to state and prove our main theorem about convergence of hyperfinite graph sequences.

Theorem 3.26 *Every locally convergent hyperfinite graph sequence is a local-global convergent Bernoulli sequence.*

Proof. Let $\{G_i\}_{i=1}^\infty$ be a locally convergent hyperfinite sequence and let μ be the involution invariant measure that is the local limit of the sequence. Proposition 3.24 implies that the Bernoulli graphing $\mathcal{B}$ corresponding to μ is hyperfinite. Assume by contradiction that $\{G_i\}_{i=1}^\infty$ does not converge in the local-global way to $\mathcal{B}$. Then it has a local-global convergent subsequence whose limit graphing $\mathcal{G}$ is not local-global equivalent to $\mathcal{G}_B = \mathcal{B}$. This however contradicts Theorem 3.25. ∎

Corollary 3.27 *Local-global convergence is equivalent to local convergence when restricted to hyperfinite graph sequences.*

3.9 Graphings as operators and expander graphings

Let $\mathcal{G}$ be a Borel graph on the probability space (X, μ). If $f: X \to \mathbb{C}$ is a measurable function then we define $\mathcal{G}f$ by

$$\mathcal{G}f(x) = \sum_{(x,v) \in E(\mathcal{G})} f(v).$$

It takes a short calculation to show that if $\mathcal{G}$ is a graphing then it acts on the Hilbert space $L^2(X,\nu)$ as a bounded self-adjoint operator. Let $f : X \to \mathbb{C}$ be an arbitrary function in $\in L^2(X,\nu)$. Then

$$\int_x |\mathcal{G}f(x)|^2 \, d\nu \leq \int_x d \sum_{(x,v)\in E(\mathcal{G})} |f(v)|^2 \, d\nu = d \int |f(x)|^2 \mathrm{deg}(x) \, d\nu \leq d^2 \|f\|_2^2.$$

The equation in the above calculation uses the fact that $\mathcal{G}$ satisfies (10). It is clear that (10) is equivalent to the statement that the action of $\mathcal{G}$ is self adjoint in the sense that $(\mathcal{G}f, g) = (f, \mathcal{G}g)$ holds for every pair f, g of bounded measurable functions. This implies that the action of $\mathcal{G}$ is also self adjoint on $L^2(X,\nu)$. The Laplace operator corresponding to a graphing is defined as $L = D - \mathcal{G}$ where $Df(x) = f(x)\mathrm{deg}(x)$. It is easy to check that

$$(Lf, f) = \int_{(v,w)\in E(\mathcal{G})} (f(v) - f(w))^2 \, d\eta^* \tag{13}$$

holds in $L^2(X,\nu)$ and thus L is positive semi definite. It follows from (13) that the multiplicity of the eigenvalue 0 in $L^2(X,\nu)$ is 1 if and only if $\mathcal{G}$ is ergodic. This is the analogue of a well known theorem from ergodic theory about the Koopman representation.

Let us restrict our attention to d-regular graphs and graphings. We say that a graphing $\mathcal{G}$ is a *c-expander*, if for every Borel set $S \subseteq X$ with $0 < \nu(S) \leq 1/2$, we have $\nu(N_1(S)) \geq (1+c)\nu(S)$. (Here $N_1(S) = \sup_{x\in S} N_{\mathcal{G},1}(x)$.)

Let $(G_n)_{n=1}^{\infty}$ be a sequence of d-regular graphs that are expanders with expansion $c > 0$. Let us select a local-global convergent subsequence, then its limit is a d-regular graphing that is also a c-expander.

The fact that we can define spectra of graphings allows us to generalize spectral conditions for expanders to graphings.

The theory of graphings is closely related to the theory of measure preserving systems (In a sense, it generalizes ergodic theory). In particular, one can define the notion of ergodicity. A graphing $\mathbb{G}$ is ergodic if there is no measurable partition of the vertex set X into positive measure sets X_1, X_2 such that there is no edge between X_1 and X_2 or equivalently such that X_1 is a union of connected components of $\mathcal{G}$. Note that graphings, when defined on an uncountable set, are never connected as graphs and so the notion of ergodicity is a good replacement for the notion of connectivity.

However, expander graphings show that graphings offer new phenomena. Ergodicity is equivalent to saying that $\nu(N_1(S)) \geq \nu(S)$ for every set S with $0 < \nu(S) \leq 1/2$. Positive expansion is a natural strengthening of this condition (which never holds for dynamical systems).

3.10 Graphings and local algorithms

Elek and Lippner [32] formulate a correspondence between graphings and local algorithms. We can make this more precise using the notions of Bernoulli graphings and factor of i.i.d processes:

Measurable graph theoretic statements for Bernoulli graphings correspond to randomized local algorithms for finite graphs.

Let us start with an example. Let T be the d-regular tree with a distinguished root and let Ω be the compact space $[0,1]^{V(T)}$. Let $f : \Omega \to [k]$ be any measurable function which depends only on the isomorphism class of the labeled rooted tree. In other words f is invariant under the action of the root preserving automorphism group of T. Using the function f we create a random model of k colorings of T in the following way. First we produce a random element $\omega \in \Omega$ by putting independent random elements from $[0,1]$ on the vertices of T and then for every $v \in V(T)$ we define the color $c(v)$ as the value of f on the labeled tree obtained from ω by placing the root to v. We say that f is the *rule* of the coloring process c. Such processes on the tree are called *factor of i.i.d processes.* We say that the rule f has radius r if it depends only on the labels on vertices in T that are of distance at most r from the root.

The following rule (of radius one), in the finite case, is a classical method to construct an independent set of nodes in a graph (see Alon and Spencer [7], Section ***). Let $f: \Omega \to \{0,1\}$ be the function which returns 1 if and only if the label on the root is smaller then the labels on all the neighboring vertices. It is clear that the corresponding random coloring c is the characteristic function on some independent set on T with probability one. We can view c as a randomized algorithm which produces an independent set of points of density $1/(d+1)$. The rule f can also be applied to a finite d-regular graph G, since it has radius one. Let us put random labels from $[0,1]$ on the vertices of G and then let us evaluate the rule f at each vertex using only the neighborhood of radius 1. We get a random $\{0,1\}$ coloring of $V(G)$ such that 1's form an independent set. Such algorithms (corresponding to a rule of bounded radius) are called local algorithms. On the other hand we can view f as the characteristic function of a single (non random) independent set in the Bernoulli graphing corresponding to the tree T. Indeed, let $\mathcal{G}$ be the Bernoulli graphing corresponding to T. The vertex set on $\mathcal{G}$ is $\mathfrak{G}[0,1]$ however in $\mathcal{G}$ almost every vertex is represented by a version of $[0,1]^{V(T)}$ and so we van evaluate or function f for almost every point. It is clear now that $f^{-1}(1)$ is an independent measurable set in $\mathcal{G}$.

A general definition of factor of i.i.d processes can be obtained through Bernoulli graphings. Let μ be an involution invariant measure on $\mathfrak{G}$ and let $\mathcal{G}_\mu$ be the corresponding Bernoulli graphing on $\mathfrak{G}[0,1]$. Let $f: \mathfrak{G}[0,1] \to [k]$ be a Borel function. Then the involution-invariant measure $\mu_{\mathcal{B},f}$ on $\mathfrak{G}[k]$ has the property that it projects to μ when the labels on the vertices are forgotten. In other words $\mu_{\mathcal{B},f}$ puts a k-coloring process on the graphs generated by μ. The measure $\mu_{\mathcal{B},f}$ is called a factor of i.i.d process on μ. The rule of the process is the function f. We say that the rule f has radius r if $f(G_1) = f(G_2)$ whenever the balls of radius r in G_1 and G_2 are isomorphic as rooted labeled graphs.

We can approximate the rule f with an arbitrary precision ϵ with another rule f' of finite radius r (which depends on ϵ) in the sense that $\nu(x|f(x) \neq f'(x)) \leq \epsilon$. An advantage of the finite radius approximation is that it can be used for local algorithms on finite graphs. Let G be a finite graph of maximal degree at most d, and let us put random labels from $[0,1]$ on the vertices in G. Then f' defines a new coloring of G such that the color of a vertex v is computed using f' for the labeled neighborhood of radius r of v.

Nondeterministic property testing. The connection of the two convergence notions can be illuminated by the following algorithmic considerations. Given a (very large) graph G with bounded degree, we use the following sampling method to gain information: we select randomly and uniformly a node of G, and explore its neighborhood with radius r. We can repeat this t times. There are a number of algorithmic tasks (parameter estimation, property testing) that can be studied in this framework; we only sketch a simple version of property testing, and its connection with local-global convergence.

It will be convenient to introduce the edit distance for graphs with bounded degree. For two graphs on the same node set $V(G) = V(G')$, we define

$$d_1(G,G') = \frac{1}{n}|E(G)\triangle E(G')|.$$

For a graph property $\mathcal{P}$, let $\mathcal{P}_{-\varepsilon} = \{G \in \mathcal{G} :\ d_1(G,\mathcal{P}) > \varepsilon\}$.

We say that the graph property $\mathcal{P}$ is *testable*, if for every $\varepsilon > 0$ there are integers $r, t \geq 1$ such that given any graph G that is large enough, taking t samples of radius r as described above, we can guess whether the graph has property $\mathcal{P}$: if $G \in \mathcal{P}$, then our guess should be "YES" with probability at least $2/3$; if $G \in \mathcal{P}_{-\varepsilon}$, then the answer should be "NO" with probability at least $2/3$. A locally convergent graph sequence cannot contain infinitely many graphs from both $\mathcal{P}$ and $\mathcal{P}_{-\varepsilon}$.

Now let us say that $\mathcal{P}$ is *nondeterministically testable*, if there is an integer $k \geq 1$, and a testable property $\mathcal{Q}$ of k-colored graphs with bounded degree, such that $G \in \mathcal{P}$ if and only if there is a k-coloring c such that $(G, c) \in \mathcal{Q}$. This k-coloring is a "witness" for our conclusion. As an example, the property "G is the disjoint union of two graphs with at least $|V(G)|/1000$ nodes" is not testable, but it is nondeterministically testable (a witness is a 2-coloring with no edge between the 2 colors). A local-global convergent graph sequence cannot contain infinitely many graphs from both $\mathcal{P}$ and $\mathcal{P}_{-\varepsilon}$.

3.11 Concluding remarks

Even finer limit notions. Limit graphings can represent even finer information than local-global convergence. Consider the following examples. Let $0 < a < 1$ be an irrational number, and consider the following three graphings: (a) $\mathcal{C}_a$ is obtained by connecting every point $x \in [0,1]$ to the two points $x \pm a \pmod 1$; (b) $\mathcal{C}'_a$ consists of two disjoint copies of $\mathcal{C}_a$ (both with weight $1/2$); (c) $\mathcal{C}''_a$ is obtained by taking two copies of $[0,1]$ (call them upper and lower), each with mass $1/2$, and connecting every lower point $x \in [0,1]$ to the two upper points $x \pm a \pmod 1$.

These three graphings are weakly isomorphic, and either one of them represents the local-global limit of the sequence of cycles. But they are "different": there is no measure preserving isomorphism between them, and this has combinatorial reasons. The graphing $\mathcal{C}'_a$ is "disconnected" (non-ergodic), while $\mathcal{C}''_a$ is "bipartite": it has a partition into two sets with positive measure such that every edge connects the two classes. The graphing $\mathcal{C}_a$ does not have any partition with either one of these properties (even if we allow an exceptional subset of measure 0). This follows from basic ergodic theory.

It seems that the graphing $\mathcal{C}_a$ should represent the limit of odd cycles, $\mathcal{C}'_a$ should represent the limit of graphs consisting of a pair of odd cycles, while $\mathcal{C}''_a$ should represent the limit of even cycles. A theory of convergence that would explain this example has not been worked out, however.

We know [17] that local convergence is equivalent to right-convergence where the target graph is in a small neighborhood of the looped complete graph with all edgeweights 1. Can local-global convergence be characterized by some stronger form of right convergence?

4 Edge coloring models and reflection positivity

The motivation of this chapter comes from statistical physics as well as from combinatorics and topology. The general setup in statistical mechanics can be outlined as follows. Let G be a graph and let $\mathcal{C}$ be a finite set of "states" or "colors". We think of G as a crystal in which either the edges or the vertices are regarded as "sites" which can have states from $\mathcal{C}$. In the first case we speak about **edge coloring models** and in the second case about **vertex coloring models**. A configuration of the whole system is a function which associates a state with each site. The states are interacting with each other at the vertices in edge coloring models and along edges in vertex coloring models. A weight is associated with each such interaction which is a real (or complex) number depending on the interacting states (in vertex coloring models there are additional weights associated the the states). A concrete model is usually given by these numbers. The partition function can be interpreted as a graph parameter which is computed by summing the products of the weights over all possible configurations of the system represented by G. It proves to be useful to extend this graph parameter linearly to the vector space of formal linear combinations of graphs. The elements of this vector space are called quantum graphs. Quantum graphs that can be obtained by gluing together a quantum graph with its reflected version (using the distributive law) are called reflection symmetric. However there are two different reasonable definitions of gluing. In the first one we glue along unfinished edges and in the second one along vertices. Correspondingly we get the notions **edge reflection symmetric** and **vertex reflection symmetric** quantum graphs. A graph parameter is called **edge reflection positive** (resp. **vertex reflection positive**) if it takes non negative values on edge-reflection symmetric (resp. vertex reflection symmetric) quantum graphs. It is a simple fact that the partition function in edge coloring models is edge reflection positive and is vertex-reflection positive in vertex coloring models. A surprising result proved by M. H. Freedman, L. Lovász and A. Schrijver (see [36]) says that vertex reflection positivity is almost enough to characterize the partition functions of vertex coloring models. The extra condition that they need is that the ranks of certain matrices (which describe the gluing operation and are called connection matrices) are growing at most exponentially. They conjectured that similar characterization can be given for edge reflection positive graph parameters. The main result of this chapter (theorem 4.2) is the proof of this conjecture in a strong version where we replace the condition on the rank growth by a week and natural condition namely that the graph parameter is multiplicative for taking disjoint union of graphs.

The major difficulty of the proof is coming from a fact which is interesting on its own right: In contrast with vertex coloring models, partition functions of edge coloring models don't determine the weights. There is an action of the orthogonal group on different edge coloring models, which leaves the corresponding partition function invariant. This phenomenon explains why it is difficult to reconstruct an edge coloring model from its partition function. In contrast with vertex coloring models we are searching for an orbit of the orthogonal group rather than one specific object. Our main tools to handle this difficulty are commutative algebra and the theory of invariants of the orthogonal group.

A topological version of the above described reflection symmetry and reflection positivity arises in topological quantum field theory (see [8] and [35]) where the gluing operation is defined on the formal linear space of manifolds with a fixed boundary.

We should also emphasize that the subject has a close connection to pure combinatorics. The partition function of a vertex coloring model can be interpreted as the number of graph homomorphisms into a fixed graph. This shows that the number of proper colorings and many related important graph parameters are coming from vertex coloring models (see [36]). In many other cases where we count certain structures in a graph (perfect matchings, fully packed loop configurations etc...) it turns out that this number is the value of the partition function of an edge coloring model. The orthogonal invariance of edge coloring models generates interesting equations between such numbers (A simple example is shown in Chapter 4.8). Another peculiar fact, that we show, is that vertex coloring models can be represented by complex valued edge coloring models such that the values of the two partition functions are identical. In some special cases the representing edge coloring model is also real valued and in this case the corresponding graph parameter is both vertex and edge reflection positive. We show that the Ising model is such an example. Finally We mention that a version of vertex coloring models with an infinite number of states is worked out and characterized in [63]. In such a model the states are elements of a measure space on which the weights are given by a measurable function. From the combinatorial point of view, these vertex coloring models can be regarded as limits of sequences of finite graphs and such objects are relevant to extremal combinatorics. In Chapter 4.7 we point out that some of these infinite models can be represented by edge coloring models with finitely many states.

4.1 Circles and Quantum graphs

Throughout this part of the thesis it will be convenient to extend the concept of graphs by introducing edges that are not incident to any vertex. We call such edges *circle* edges and picture them as topological circles. Formally, a circle is an element of the edge set which has no endpoints. Let $\mathcal{G}$ denote the set of isomorphism classes of graphs, in which loops, circles and multiple edges are allowed. We denote by $\emptyset \in \mathcal{G}$ the empty graph whose vertex and edge sets are both empty. If $G_1, G_2 \in \mathcal{G}$ are two graphs then their disjoint union $G_1 \cup G_2$ is defined to be a graph whose vertex set and edge set is the disjoint union of those of G_1 and G_2. Every element of $\mathcal{G}$ is the disjoint union of an ordinary (circle free) graph and a finite number of circles.

Let R be an arbitrary commutative ring with 1. An R-valued graph parameter is a map $f : \mathcal{G} \to R$. We say that f is *multiplicative* if $f(G_1 \cup G_2) = f(G_1)f(G_2)$ for any two graphs $G_1, G_2 \in \mathcal{G}$ and $f(\emptyset) = 1$.

Let F be a field and let $\mathcal{Q}(F)$ denote the vectorspace of finite $F-$linear combinations of elements of $\mathcal{G}$. The elements of $\mathcal{Q}(F)$ are called *quantum graphs*. The operation of taking disjoint union can be extended to quantum graphs by using the distributive law. It is easy to see that $\mathcal{Q}(F)$ becomes an F-algebra with 1 if we introduce disjoint union as multiplication. If the ring R is an F-algebra then any multiplicative graph parameter $f : \mathcal{G} \to R$ extends uniquely to an algebra homomorphism $f : \mathcal{Q}(F) \to R$. As a consequence we have that the image of $\mathcal{Q}(F)$ is a subalgebra of R.

The usual setting is that F is the field of real numbers and R is either the field of real numbers or a polynomial ring over the reals. For this reason we use the shorthand notation $\mathcal{Q}$ instead of $\mathcal{Q}(\mathbb{R})$.

4.2 Edge reflection positivity

In this section we will need the notion of graphs with outgoing (open) edges. An outgoing edge can be pictured as an edge which goes out from the graph but is not finished. It can also happen that such an edge goes out in both directions and so it is not incident to any of the verteces. However the behavior of these edges is different from circles because we want to maintain the possibility of finishing them. To define this concept precisely we need to introduce the set of "open ends" $O(G)$ of a graph G. A graph G with outgoing edges is a triple $(V(G), E(G), O(G))$ where $(O(G) \cup V(G), E(G))$ is a graph in $\mathcal{G}$ with the property that the degrees of open ends are exactly 1.

We define $\mathcal{G}_k$ to be the set of all graphs with exactly k open ends which are labeled by the numbers $1, 2, \ldots, k$. There is a natural operation

$$g : \mathcal{G}_k \times \mathcal{G}_k \to \mathcal{G}$$

which is called *gluing* and defined in the following way. Let G_1 and G_2 be two graphs in $\mathcal{G}_k$. Let us take the disjoint union of them and identify their open ends which have the same label. This way we obtain a graph in $\mathcal{G}$ in which there are k labeled vertices of degree 2. Finally we eliminate these vertices (and their two incident edges) by introducing a new edge which connects their two neighbors directly. It is easy to see that the resulting graph does not depend on the order in which we eliminate the labeled vertices. Note also that the resulting graph may contain circles even if the original two graphs did not have any. This explains the importance of circles. The notation of gluing is also defined for graphs with no outgoing edges, but in this case gluing is the same as taking disjoint union.

Let $\mathcal{Q}_k$ denote the vectorspace of formal $\mathbb{R}$-linear combinations of elements of $\mathcal{G}_k$. Now the gluing operation extends uniquely to a symmetric bilinear form

$$g : \mathcal{Q}_k \times \mathcal{Q}_k \to \mathcal{Q}.$$

We say that a quantum graph $Q \in \mathcal{Q}$ is *edge reflection symmetric* if $Q = g(H, H)$ for some quantum graph $H \in \mathcal{Q}_k$ with $k \geq 0$. A graph parameter $f : \mathcal{G} \to \mathbb{R}$ is called *edge reflection positive* if its linear extension $f : \mathcal{Q} \to \mathbb{R}$ takes non-negative values on all edge reflection symmetric quantum graphs. In other words f is edge reflection positive if and only if the bilinear forms

$$f \circ g : \mathcal{Q}_k \times \mathcal{Q}_k \to \mathbb{R}$$

are positive semi-definite for all $k \geq 0$. One can write up the matrices of these scalar products in the natural basis $\mathcal{G}_k$ and obtain the so-called *connection matrices* $M(k, f)$. These are infinite matrices whose rows and columns are indexed by the elements of $\mathcal{G}_k$ and the entry in the intersection of the row corresponding to G_1 and the column corresponding to G_2 is $f(g(G_1, G_2))$.

4.3 Edge coloring models and the characterization theorem

Let R be a commutative $\mathbb{R}$-algebra with 1. (Usually $R = \mathbb{R}$ or a polynomial ring over $\mathbb{R}$.) Let $\mathcal{C} = \{c_1, c_2, \ldots, c_d\}$ be a finite set of size d whose elements will be referred as colors. An R-valued *edge coloring model* is given by a function $t : \mathbb{N}^d \to R$ where 0 is considered to be a natural number. For every edge coloring model we are going to define an associated graph parameter $t : \mathcal{G} \to R$. Let

$v \in V(G)$ be a vertex and let $\psi : E(G) \to \mathcal{C}$ be a coloring of the edge set of a graph G. We denote by $v_\psi \in \mathbb{N}^d$ the vector whose i-th coordinate is the number of edges with color c_i incident to vertex v. It is important that loop edges are counted twice. Now we define $t_\psi(G)$ by

$$t_\psi(G) = \prod_{v \in V(G)} t(v_\psi)$$

and $t(G)$ by

$$t(G) = \sum_{\psi : E(G) \to \mathcal{C}} t_\psi(G).$$

It is clear that t is a multiplicative graph parameter if we define the empty product to be 1 and moreover the value of t on a single circle is the number of colors (which is d).

Let $k \geq 0$ be a natural number and let $G \in \mathcal{G}_k$ be a graph with k labeled outgoing edges. We say that a coloring $\psi : E(G) \to \mathcal{C}$ is an extension of a coloring $\chi : O(G) \to \mathcal{C}$ of the open ends if each open end $o \in O(G)$ has the same color as the unique edge incident to o. We denote this relation by $\psi > \chi$. For a coloring $\chi : O(G) \to \mathcal{C}$ we introduce $t_\chi(G)$ by

$$t_\chi(G) = \sum_{\psi : E(G) \to \mathcal{C},\ \psi > \chi} t_\psi(G)$$

Now let G_1 and G_2 be two graphs in $\mathcal{G}_k$. Since the open ends in both G_1 and G_2 are labeled by numbers $1, 2, \ldots, k$ we can say, by abusing the notation, that any coloring $\chi : \{1, 2, \ldots, k\} \to \mathcal{C}$ is also a coloring of $O(G_1)$ and $O(G_2)$. Assume that $\psi_1 > \chi$ in G_1 and $\psi_2 > \chi$ in G_2 for the same coloring χ. Then there is a coloring $\psi = g(\psi_1, \psi_2)$ of the edges of $G = g(G_1, G_2) \in \mathcal{G}$ which is obtained by gluing together ψ_1 and ψ_2. This coloring has the property that $v_\psi = v_{\psi_1}$ if $v \in V(G_1)$ and $v_\psi = v_{\psi_2}$ if $v \in V(G_2)$. It follows that

$$t_\psi(G) = t_{\psi_1}(G_1) t_{\psi_2}(G_2)$$

and that

$$t(G) = \sum_{\chi : \{1,2,\ldots,k\} \to \mathcal{C}} t_\chi(G_1) t_\chi(G_2). \tag{14}$$

It is clear that the previous equality also holds for $G_1, G_2 \in \mathcal{Q}_k$ and $G = g(G_1, G_2) \in \mathcal{Q}$ if we extend the invariants t and t_χ linearly to quantum graphs from $\mathcal{Q}$ and $\mathcal{Q}_k$. As a consequence we get that real valued edge coloring models give rise to edge reflection positive graph parameters:

Proposition 4.1 *Let $t : \mathbb{N}^d \to \mathbb{R}$ be a real valued edge coloring model. Then the graph parameter $t : \mathcal{G} \to \mathbb{R}$ is edge reflection positive.*

Proof. Let $k \geq 0$ ba a natural number and $Q = g(H, H)$ for some $H \in \mathcal{G}_k$. Using equation (14) we have that

$$t(Q) = \sum_{\chi : \{1,2,\ldots,k\} \to \mathcal{C}} t_\chi(H)^2 \geq 0.$$

∎

Our main theorem is the converse of the previous statement.

Theorem 4.2 *Let $f : \mathcal{G} \to \mathbb{R}$ be an edge reflection positive and multiplicative graph parameter. Then there is an edge coloring model $t : \mathbb{N}^d \to \mathbb{R}$ such that the corresponding graph parameter equals to f.*

The subsequent chapters will lead to the proof of this theorem.

4.4 Universal edge coloring models

Let us fix a natural number d and let us introduce algebraically independent variables x_v for each vector $v \in \mathbb{N}^d$. Let P_d denote the polynomial ring $\mathbb{R}[\{x_v \mid v \in \mathbb{N}^d\}]$. The *universal edge coloring model* t_d corresponding to d is a P_d valued edge coloring model which is given by the function $t_d(v) = x_v$. An important property of these models is that real valued edge coloring models t with d colors are in one to one correspondence with homomorphisms $\varrho : P_d \to \mathbb{R}$ where the correspondence is given by the equation $\varrho(x_v) = t(v)$. Note that if t and ϱ correspond to each other then $t(Q) = \varrho(t_d(Q))$ for all $Q \in \mathcal{Q}$.

Let us introduce

$$I_d = \{t_d(Q) \mid Q \in \mathcal{Q}\}.$$

Since t_d is multiplicative we have that I_d is a subring of P_d. We will prove later that I_d is the set of all elements in P_d which are invariant under a certain "natural" action of the orthogonal group $O_d(\mathbb{R})$.

4.5 Action of the orthogonal group on edge coloring models

Let d be a natural number and let V be the vectorspace consisting of the formal $\mathbb{R}$-linear combinations of the colors $c_1, c_2, \ldots, c_d$. We say that V is the *color space* and the elements of V will be called *quantum colors*. The space V is endowed with an euclidean scalar product for which $c_1, c_2, \ldots, c_d$ is an orthonormal basis. Let us fix an edge coloring model $t : \mathbb{N}^d \to R$. For every natural number n we define a symmetric n-linear form l_n on V by

$$l_n(c_{i_1}, c_{i_2}, \ldots, c_{i_n}) = t(m_1, m_2, ..., m_d)$$

where m_i denotes the number of occurrence of the color c_i on the list $c_{i_1}, c_{i_2}, \ldots, c_{i_n}$.

Let α be an orthogonal transformation of V. We denote by u^α the image of a vector $u \in V$ under the action of α. We define a new edge coloring model t^α by

$$t^\alpha(i_1, i_2, \ldots, i_d) = l_n(c^\alpha_{j_1}, c^\alpha_{j_2}, \ldots, c^\alpha_{j_n})$$

where $n = i_1 + i_2 + \cdots + i_d$ and $d \geq j_1, j_2, \ldots, j_n \geq 1$ is an arbitrary sequence of integers such that $|\{k \mid j_k = m\}| = i_m$. The goal of this section is to prove the following.

Proposition 4.3 *Let $G \in \mathcal{G}$ be an arbitrary graph. Then $t(G) = t^\alpha(G)$.*

Let $G \in \mathcal{G}$ be a fixed circle free graph. A *half edge* in G is an ordered pair (v, e) of a vertex v and an edge e such that v and e are forming an incident pair. For each edge $e \in E(G)$ we introduce two half edges $h(e, 1) = (v_1, e)$ and $h(e, 2) = (v_2, e)$ where v_1 and v_2 are the two endpoints of e. In case

e is a loop, we think of $h(e,1)$ and $h(e,2)$ as different objects although the corresponding ordered pairs are the same. We denote by $H(G) = \{h(e,i) \mid e \in E(G),\ i \in \{1,2\}\}$ the set of half edges in G.

For each half edge $h(e,i)$, we introduce an isomorphic copy of V which we denote by $V_{e,i}$. In each such space $V_{e,i}$ there is a natural basis whose elements correspond to the colors $c_1, c_2, \dots, c_d$. We denote the elements of this basis by $c_{1,e,i}, c_{2,e,i}, \dots, c_{d,e,i}$. Let

$$W = \bigotimes_{e\in E(G),\ i\in\{1,2\}} V_{e,i}$$

be the tensor product of all these spaces. For the edge coloring model $t : \mathbb{N}^d \to R$ we define a linear form $m : W \to R$ by

$$m\Big(\bigotimes_{e\in E(G)\ ,\ i\in\{1,2\}} u_{e,i}\Big) = \prod_{v\in V(G)} l_{d(v)}(u(v,1), u(v,2), \dots, u(v,d(v)))$$

where $u_{e,i} \in V_{e,i}$ are arbitrary elements, $d(v)$ is the degree of the vertex v and $u(v,1), u(v,2), \dots, u(v,d(v))$ is the list of those $u_{e,i}$-s for which the half edge $h(e,i)$ is incident to v. Since every half edge is incident to exactly one vertex we have that the right hand side is multi linear in the vectors $u_{e,i}$ and thus by the universal property of the tensor product there is a unique m which satisfies the equation.

Let us consider the spaces $W_e = V_{e,1} \otimes V_{e,2}$ associated to the edges of G. A basis of W_e is formed by the elements $c_{i,e,1} \otimes c_{j,e,2}$ where $1 \le i,j \le d$. Thus the elements of W_e can be represented as matrices whose rows and columns are indexed by the elements of $\mathcal{C}$. Let $J_e = \sum_{i=1}^d c_{i,e,1} \otimes c_{i,e,2}$ be the element of W_e which correspond to the identity matrix and let

$$J = \bigotimes_{e\in E(G)} J_e \in \bigotimes_{e\in E(G)} W_e = W.$$

We have that

$$J = \sum_{\psi: E(G)\to\{1,2,\dots,d\}} \ \bigotimes_{e\in E(G)\ ,\ i\in\{1,2\}} c_{\psi(e),e,i}\ .$$

Since the terms of this sum correspond to the colorings of the edges in G it follows that

$$t(G) = m(J).$$

Let J_0 denote the element $\sum_{i=1}^d c_i \otimes c_i$ in $V \otimes V$.

Lemma 4.4 *If $b_1, b_2, \dots, b_d$ is an orthonormal basis in V then $J_0 = \sum_{i=1}^d b_i \otimes b_i$.*

Proof. Assume that $b_j = \sum_{i=1}^d a_{j,i} c_i$ for some real numbers $a_{j,i}$. Then the matrix $A = \{a_{i,j}\}$ is orthogonal and thus $AA^T = 1$. Now

$$\sum_{j=1}^d b_j \otimes b_j = \sum_{j=1}^d \sum_{i=1}^d \sum_{k=1}^d a(j,i)a(j,k) c_i \otimes c_k = \sum_{i=1}^d \sum_{k=1}^d \delta_{i,k} c_i \otimes c_k = J_0.$$

■

Recall that α was an orthogonal transformation of V and thus $c_1^\alpha, c_2^\alpha, \dots, c_d^\alpha$ is an orthonormal basis in V. By lemma 4.4 we obtain that $\sum_{i=1}^d c_{i,e,1}^\alpha \otimes c_{i,e,2}^\alpha = J_e$ and so

$$J = \sum_{\psi:E(G)\to\{1,2,\dots,d\}} \bigotimes_{e\in E(G)\ ,\ i\in\{1,2\}} c^{\alpha}_{\psi(e),e,i} = J^{\alpha}$$

We obtain that

$$t(G) = m(J) = m(J^{\alpha}) = t^{\alpha}(G)$$

for all circle free graphs G.

Now let $H \in \mathcal{G}$ be an arbitrary graph which is the disjoint union of a circle free graph G and n circles. The equation

$$t(H) = t(G)d^n = t^{\alpha}(G)d^n = t^{\alpha}(H)$$

completes the proof of proposition 4.3.

4.6 Action of the orthogonal group on the polynomial ring P_d

Recall that P_d is the polynomial ring $\mathbb{R}[\{x_v \mid v \in \mathbb{N}^d\}]$ and the universal edge coloring model t_d is given by the map $t_d : v \to x_v$. For a fixed orthogonal transformation α of the color space V we have a new edge coloring model $t^{\alpha}_d : \mathbb{N}^d \to P_d$. Using that P_d is a free commutative $\mathbb{R}$-algebra with free generators $\{x_v \mid v \in \mathbb{N}^d\}$ we get that the map

$$x_v \to t^{\alpha}_d(v)$$

extends to an algebra endomorphism $\alpha : R_d \to R_d$. Since α^{-1} induces another endomorphism which is both right and left inverse for α it turns out that α is an automorphism of R_d. Proposition 4.3 implies that

Corollary 4.5 *The elements of the subring $I_d < R_d$ are invariant under the action of α for all orthogonal transformations α.*

We define the hight $h(x_v)$ of a variable $x_v \in P_d$ to be the sum of the components of v. The hight of a monomial $x_{v_1}x_{v_2}\dots x_{v_r}$ is defined to be the multiset $\{h(x_{v_1}), h(x_{v_2}), \dots, h(x_{v_r})\}$. We denote by W_S the linear subspace generated by all the monomials in P_d of hight S. It is clear that P_d is the direct sum of the spaces W_S where S runs over all possible finite multisets of the non negative integers. We show that these subspaces are invariant under the action of the orthogonal group. To see this let us fix an element $\alpha \in O(V)$. Since $(x_{v_1}x_{v_2}\dots x_{v_r})^{\alpha} = x^{\alpha}_{v_1}x^{\alpha}_{v_2}\dots x^{\alpha}_{v_r}$ it is enough to prove that x^{α}_v is a linear combination of variables of hight $n = h(x_v)$ for all v. Let us represent v by a multiset of colors $\{c_{i_1}, c_{i_2}, \dots, c_{i_n}\}$. We have that

$$x^{\alpha}_v = t^{\alpha}_d(v) = l_n(c^{\alpha}_{i_1}, c^{\alpha}_{i_2}, \dots, c^{\alpha}_{i_n}).$$

By the multilinearity of l_n the right hand side of the above equation can be written as a linear combination of some monomials of the form $l_n(c_{k_1}, c_{k_2}, \dots, c_{k_n})$ which are all variables of hight n.

4.7 vertex coloring models as edge coloring models

A vertex coloring model (see [36]) is given by a finite weighted graph H with real edge weights $\beta_H(i,j)$ and positive vertex weights $\alpha_H(i)$. If G is a simple graph then the homomorphism function (or partition function) $\hom(G,H)$ is defined by

$$\hom(G,H) = \sum_{\phi:V(G)\to V(H)} \prod_{v\in V(G)} \alpha_H(\phi(v)) \prod_{uv\in E(G)} \beta_H(\phi(u),\phi(v)).$$

We show that the graph parameter $\hom(G,H)$ can be represented by the partition function of an edge coloring model where the number of colors is the rank of the matrix of the edge weights in H. Let B be the symmetric matrix of the edge weights $\beta_H(i,j)$. From elementary linear algebra we know that

$$B = \lambda_1 u_1 {u_1}^T + \lambda_2 u_2 {u_2}^T + \cdots + \lambda_r u_r {u_r}^T$$

for some real column vectors u_i and numbers $\lambda_i \in \{1,-1\}$ where r is the rank of B. Let t be the edge coloring model given by

$$t(s_1,s_2,\dots,s_r) = \sum_{j=1}^{|V(H)|} \alpha_H(j) \prod_{i=1}^{r} (u_i(j)\sqrt{\lambda_i})^{s_i}.$$

Using that $\beta_H(i,j) = \sum_{k=1}^{r} u_k(i)u_k(j)\lambda_k$ we get that $t(G) = \hom(G,H)$ for an arbitrary simple graph G. It is worth mentioning that if B is positive semidefinite then the numbers λ_i are all 1 and the representing edge coloring model is real valued. However there are cases when the edge coloring model is real valued without this condition. A simple example is the Ising model which can be represented by a weighted graph on 2 vertices with $\beta(1,1) = \beta(2,2) = a \geq 0$, $\beta(1,2) = \beta(2,1) = b \geq 0$ and $\alpha(1) = \alpha(2) = 1$. It is easy to compute that the corresponding edge coloring model is given by

$$t(s_1,s_2) = 2\Big(\frac{a+b}{2}\Big)^{\frac{s_1}{2}}\Big(\frac{a-b}{2}\Big)^{\frac{s_2}{2}} \quad \text{if } s_2 \text{ is even}$$
$$t(s_1,s_2) = 0 \quad \text{if } s_2 \text{ is odd.}$$

It is an interesting phenomenon that the number of colors needed to represent a vertex coloring model by an edge coloring model depends only on the rank of the adjacency matrix of the weighted graph. This leads to a family of infinite vertex coloring models which are still representable by ordinary edge coloring models. Let $w : [0,a]^2 \to \mathbb{R}$ be a bounded symmetric measurable function such that

$$w(x,y) = \sum_{i=1}^{r} \lambda_i f_i(x) f_i(y)$$

for some bounded measurable functions f_i and numbers $\lambda_i \in \{1,-1\}$. Regarding the function w as an infinite adjacency matrix one can define an analogy of the homomorphism function by

$$t_w(G) = \int_{x_1,x_2,\dots,x_m} \prod_{(i,j)\in E(G)\ ,\ i<j} w(x_i,x_j)\ dx_1\ dx_2 \dots dx_m$$

where G is an arbitrary graph with $|V(G)| = m$ such that the vertices of G are indexed by the numbers $\{1,2,\dots,m\}$. Note that t_w is a vertex reflection positive and multiplicative graph parameter.

Let us introduce the following edge coloring model

$$t(s_1, s_2, \dots, s_r) = \int_{x\in[0,a]} \prod_{i=1}^{r} (f_i(x)\sqrt{\lambda_i})^{s_i}\ dx.$$

It can be calculated easily that $t_w(G) = t(G)$ and that if $\lambda_i = 1$ for all i then t gives rise to a real valued edge coloring model.

4.8 Graph parameters from combinatorics

Many interesting graph parameters can be obtained from the following special family of edge coloring models. Let S be a subset of $\mathbb{N}^r$ and let $t_S : \mathbb{N}^r \to \mathbb{R}$ be the function with $t_S(v) = 1$ if $s \in S$ and $t_S(v) = 0$ if $s \notin S$. Let G be a simple graph. The next table lists a few examples.

S	combinatorial meaning of $t_S(G)$
$\{1\} \times \mathbb{N}$	number of perfect machings
$\{2\} \times \mathbb{N}$	number of fully packed loop configurations
$\{0,1\} \times \mathbb{N}$	number of matchins
$\{0,2\} \times \mathbb{N}$	number of loop configurations
$\{0,d\} \times \mathbb{N}$	number of d-regular subgraphs
$\{0,1\}^d$	number of proper edge colorings with d colors
$\{(b,c,d)\|b+d \equiv c+d \equiv 0\ (2)\}$	number of nowhere zero 4-flows
$\{(2,0,0),(0,2,0),(0,0,1)\} \times \mathbb{N}$	permanent of the adjacency matrix

Using the orthogonal invariance of partition functions one can create peculiar equations. For example by rotating the firs example on the list with 45-degree we get the edge coloring model

$$t(a,b) = \sqrt{2}^{\,-(a+b)}(a-b)$$

whose partition function is again the number of perfect matchings. In other words, the partition function of the model $t(a,b) = a-b$ is $2^{|E(G)|}$ times the number of perfect matchings in G.

4.9 Open questions

Let f be an edge coloring model with d colors and let $M(k,f)$ denote its k-th connection matrix (see chapter 4.2). It is not hard to see [37] that $\mathrm{rk}(M(k,f)) \le d^k$.

Question 4.6 What are the possible sequences $\mathrm{rk}(M(k,f))$, $k = 1,2,3,\dots$?

The analogy of this question for vertex coloring models was answered by L. Lovász in [61].

The next question is motivated by chapter 4.7.

Question 4.7 Which are those vertex coloring models whose partition functions are edge reflection positive.

We know only two examples: The Ising model and the vertex coloring models with positive semidefinite adjacency matrices.

4.10 The value of a circle

Let k be a natural number and let $\mathcal{M}_k$ denote the set of those graphs G in $\mathcal{G}_k$ which are circle free and $V(G) = \emptyset$. In particular the Edge set of G is a perfect matching on the k open ends. It follows that if k is an odd number then $\mathcal{M}_k$ is empty. We denote by $\mathcal{QM}_k$ the subspace generated by $\mathcal{M}_k$ in $\mathcal{Q}_k$.

Assume that $k = 2n$ for some natural number n and let A_k denote the subset of all elements G of $\mathcal{M}_k$ with the property that each edge of G connects an open end with label $\leq n$ and another open end with label $> n$. The elements of A_k can be represented by permutations of the set $\{1, 2, \ldots, n\}$ in the way that a permutation π correspond to a matching $a_\pi \in A_k$ where the open end i is connected with $\pi(i) + n$ for all $1 \leq i \leq n$. Now the definition of gluing implies that $g(a_\pi, a_\varrho)$ is a graph which is the disjoint union of $c(\pi\varrho^{-1})$ circles where $c(\sigma)$ denotes the number of cycles in a permutation σ. Recall d be the value of f on a single circle. Using the multiplicativity of f we have that

$$f(g(a_\pi, a_\varrho)) = d^{\ c(\pi\varrho^{-1})}.$$

Let M_n be a matrix whose rows and columns are indexed by permutations from the symmetric group S_n and the entry in the intersection if the row corresponding to π and column corresponding to ϱ is $d^{\ c(\pi\varrho^{-1})}$. The assumption that f is reflection positive implies that M_n must be a positive semidefinite matrix for every n. We will prove that this is only possible if d is a non negative integer.

The matrix M_n is acting on the space of formal linear combinations of the group elements of S_n which is the group algebra $\mathbb{R}[S_n]$. Let

$$w = \sum_{\pi \in S_n} \operatorname{sgn}(\pi)\pi.$$

We have that

$$wM_k = \sum_{\pi,\varrho \in S_n} \operatorname{sgn}(\pi) d^{\ c(\pi\varrho^{-1})} \varrho = \sum_{\pi,\varrho \in S_n} \operatorname{sgn}(\pi\varrho^{-1}) d^{\ c(\pi\varrho^{-1})} \operatorname{sgn}(\varrho)\varrho = w\Big(\sum_{\pi \in S_n} \operatorname{sgn}(\pi) d^{c(\pi)}\Big)$$

This means that w is an eigenvector of M_n with eigenvalue

$$\sum_{\pi \in S_n} \operatorname{sgn}(\pi) d^{c(\pi)} = d(d-1)(d-2)\ldots(d-n+1).$$

The positive semidefinitness of M_n implies that $d(d-1)\ldots(d-n+1)$ must be a non negative number for every natural number n and so d is a non negative integer.

As the next lemma shows, the fact that the circle value is a non negative integer is the first step towards the existence of an edge coloring model representing f.

Lemma 4.8 *Let t be an arbitrary edge coloring model with d colors. Then $f(g(H, K)) = t(g(H, K))$ for every pair $H, K \in \mathcal{QM}_k$.*

Proof. The quantum graph $g(H, K)$ is the linear combination of graphs consisting only of circles. The multiplicativity of f shows that the value of t and f must be the same on such a graph. ∎

Using the terminology of section 4.3, we have that

Lemma 4.9 *Let t be any edge coloring model with d colors and let $H \in \mathcal{QM}_k$ then $f(g(H, H)) = t(g(H, H)) = 0$ if and only if $t_\chi(H) = 0$ for all colorings $\chi : \{1, 2, \ldots, k\} \to \mathcal{C}$.*

4.11 Lifting to the universal edge coloring model

Recall that d is the circle value of f and t_d is the universal edge coloring model with d colors. In this section we prove that

Lemma 4.10 *If $Q \in \mathcal{Q}$ is an arbitrary quantum graph and $t_d(Q) = 0$ then $f(Q) = 0$.*

Proof. For a graph $G \in \mathcal{G}$ we define its hight $h(G)$ to be the multiset of the degrees of the vertices in G. From the definition of the universal edge coloring model it follows that $t_d(G) \in W_{h(G)}$. Every quantum graph $Q \in \mathcal{Q}$ can be written in the form $\sum_S Q_S$ where Q_S is a quantum graph which is a linear combination of graphs of hight S. We have that $t_d(Q) = \sum_S t_d(Q_S)$ and $t_d(Q_S) \in W_S$ for every multiset S. It follow that $t_d(Q) = 0$ implies that $t_d(Q_S) = 0$ for all multiset S. Thus we can assume that Q is homogeneous in the sense that each graph component of Q has the same hight S.

Let $G \in \mathcal{G}$ be a graph which is the disjoint union of a circle free graph H and n circles. Both f and t_d vanish on the quantum graph $G - d^n H$. This means that one can eliminate circles in a quantum graph without changing the value of f and t_d on it. Thus we can assume that Q is a combination of circle free graphs.

Assume that S consists of n numbers an their sum is k. Let G_S be a graph in $\mathcal{G}_k$ with the following properties:

1. $|V(G_S)| = n$
2. $|E(G_S)| = |O(G)| = k$
3. Each edge $e \in E(G)$ connects an open end with a vertex
4. The multiset of the degrees of the verteces is S.

It is clear that G_S is unique up to a relabeling of the open ends. It is also clear that for every graph G of hight S there is a matching $M \in \mathcal{M}_k$ such that $G = g(G_S, M)$. This implies that our quantum graph Q can be written in the form $g(G_S, M)$ where M is in $\mathcal{QM}_k$.

Let $\mathcal{P}_v \subseteq O(G)$ denote the set of those open ends which are connected to the vertex $v \in V(G)$ in G_S and let $\mathcal{P} = \{\mathcal{P}_v | v \in V(G)\}$ be the partition formed by these sets. We denote by $K \leq S_k$ the automorphism group of $\mathcal{P}$. It is clear that $g(G_S, M)$ is isomorphic to $g(G_S, M^\sigma)$ for all $\sigma \in K$. Let

$$\hat{M} = \frac{1}{|K|} \sum_{\sigma \in K} M^\sigma.$$

We have that

$$f(Q) = f(g(G_S, M)) = f(g(G_S, \hat{M}))$$

and

$$0 = t_d(Q) = t_d(g(G_S, M)) = t_d(g(G_S, \hat{M}))$$

.

Using equation (14) from chapter 4.3 we get that

$$0 = t_d(Q) = \sum_{\chi:\{1,2,\ldots,k\} \to \mathcal{C}} t_{d_\chi}(G_S) t_{d_\chi}(\hat{M}).$$

The group K is acting on the colorings $\chi : \{1, 2, \ldots, k\} \to \mathcal{C}$. An orbit of this action can be described as a multiset X_χ of multisets such that the elements of X_χ are multisets of colors describing the color distributions in different partition sets of $\mathcal{P}$. The value of $t_{d_\chi}(G_S)$ depends only on the orbit of χ because G_S and G_S^σ are isomorphic for every $\sigma \in K$. Moreover $t_{d_\chi}(G_S)$ is a monomial of hight S of the form $x_{v_1}x_{v_2}\ldots x_{v_n}$ where the vectors v_i describe multisets of colors which can be seen at different vertices and the list $v_1, v_2, \ldots, x_n$ describes X_χ. On the other hand $t_{d_\chi}(\hat{M})$ is a real number which depend also only the orbit of χ because $\hat{M}$ is K symmetric. By using the fact that different monomials are linearly independent over $\mathbb{R}$ we obtain that $t_{d_\chi}(\hat{M})$ must be 0 for all colorings χ. This implies by lemma 4.9 that $t_d(g(\hat{M}, \hat{M})) = 0$ and so by lemma 4.8 we get that $f(g(\hat{M}, \hat{M})) = 0$.

Since $f(g(-, -))$ is a positive semidefinite form it follows that $f(g(Y, \hat{M})) = 0$ for all $Y \in \mathcal{G}_k$. In particular

$$f(Q) = f(g(G_S, \hat{M})) = 0.$$

■

Corollary 4.11 *There exists a homomorphism $\hat{f} : I_d \to \mathbb{R}$ such that $f(Q) = \hat{f}(t_d(Q))$ for every quantum graph $Q \in \mathcal{Q}$.*

Proof. Recall that $\mathcal{Q}$ has an $\mathbb{R}$-algebra structure and that I_d is the image of $\mathcal{Q}$ under the algebra homomorphism $t_d : \mathcal{Q} \to P_d$. On the other hand $f : \mathcal{Q} \to \mathbb{R}$ is an algebra homomorphism because f is multiplicative. According to the main result of this section we have that

$$\ker(t_d) \subseteq \ker(f) \subseteq \mathcal{Q}$$

and this completes the proof. ■

4.12 Representing invariants of the orthogonal group with quantum graphs

The main result of this section is the following.

Lemma 4.12 *If $p \in P_d$ is invariant under the action of the orthogonal group $O_d(\mathbb{R})$ then there is a quantum graph $Q \in \mathcal{Q}$ such that $t_d(Q) = p$.*

Before we start proving lemma 4.12 we describe a construction which will be useful in this and in later sections.

tensor construction: Let X be an arbitrary finite set with a partition $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ on its elements. Let V_x be an isomorphic copy of the color-space $V = \langle c_1, c_2, \ldots, c_d\rangle_{\mathbb{R}}$ for each element $x \in X$ and let

$$T(X, \mathcal{Y}) = \bigotimes_{x \in X} V_x.$$

Let l_n denote the symmetric n-linear form from chapter 4.5 associated with the universal edge coloring model t_d. For each partition set Y_i we define a multilinear form $\hat{m}_i$ by applying $l_{|Y_i|}$ for the spaces $\{V_x | x \in Y_i\}$. The product $\prod_{i=1}^k \hat{m}_i$ defines a multilinear form in the spaces $\{V_x | x \in X\}$. By factoring $\hat{m}$ trough the tensor product $T(X, \mathcal{Y})$ we get an $\mathbb{R}$-linear form $m : T(X, \mathcal{Y}) \to P_d$.

The space $T(X,\mathcal{Y})$ admits a euclidean scalar product which comes from the euclidean structure on V. An orthonormal basis for this scalar product is formed by the different tensor products of the color vectors. This basis will be called the *color basis.* The orthogonal group $O_d(\mathbb{R})$ which preserves the scalar product on V is also acting on $T(X,\mathcal{Y})$ by taking the tensor product of the actions on V_x. This action has the property that $m(t^\alpha)=m(t)^\alpha$ where $t\in T(X,\mathcal{Y})$ and $\alpha\in O_d(\mathbb{R})$. Let $S=\{|Y_1|,|Y_2|,\dots,|Y_n|\}$ be the multiset of the sizes of the partition sets. If we substitute color vectors into the multilinear form $\hat{m}$ we get all the monomial of hight S in P_d. It follows that m maps $T(X,\mathcal{Y})$ to W_s surjectively.

Proof of lemma 4.12 We know from chapter 4.6 that P_d is the direct sum (as a vectorspace) of the spaces W_S where each W_S is invariant (as a subspace) under the action of $O_d(\mathbb{R})$. It follows that if p is an invariant element of $O_d(\mathbb{R})$ then each W_S component p_S of p must be invariant too. Since p is the sum of its W_S components it is enough to find quantum graphs Q_S with $t_d(Q_S)=p_S$ for each multiset S.

Let $S=\{s_1,s_2,\dots,s_n\}$ be a fixed multiset of natural numbers and let $k=\sum_i s_i$. Let $\mathcal{D}=\{D_1,D_2,\dots,D_n\}$ be a partition of the index set $\{1,2,\dots,k\}$ such that $|D_i|=s_i$ for all $1\le i\le n$. Let $W=T(\{1,2,\dots,k\},\mathcal{D})$ and let W^0 be the kernel of the map m. We have that the space W^0 is invariant under $O_d(\mathbb{R})$ and W/W^0 is isomorphic to W_S in a way that the induced action of $O_d(\mathbb{R})$ on W/W^0 commutes with this isomorphism. By abusing the notation we identify W_S with W/W^0.

Let p_1 be a preimage of p_S under the homomorphism $W\to W/W^0$ and let

$$\bar{p}=\int_{\alpha\in O_d(\mathbb{R})} p_1^\alpha\ d\nu$$

where ν is the normalized Haar measure on the orthogonal group $O_d(\mathbb{R})$. Since p_S is invariant in $W_S=W/W^0$ it follows that $\bar{p}$ is also a preimage of p_S under the map $W\to W/W^0$. Furthermore we have that $\bar{p}$ is an invariant of $O_d(\mathbb{R})$.

The first fundamental theorem of Weyl [91] describes the space of invariant elements in W by determining a generating system for it. The elements of this generating system correspond to partitions of the set $\{1,2,\dots,k\}$ into two element subsets. In particular if k is an odd number then the only invariant is the zero vector. Assume that k is even and let $\mathcal{E}=\{E_1,E_2,\dots,E_{k/2}\}$ be such a partition. Let $\chi:\mathcal{E}\to\mathcal{C}$ be a coloring of the partition sets. The function χ induces a coloring $\hat{\chi}:\{1,2,\dots,k\}\to\mathcal{C}$ such that $\hat{\chi}(j)=\chi(E_{i_j})$ for $1\le j\le k$ where i_j denotes the number for which $j\in E_{i_j}$. We define g_χ to be the tensor product

$$\bigotimes_{i\in\{1,2,\dots,k\}}\hat{\chi}(i)$$

where $\hat{\chi}(i)\in V_i$. The invariant which correspond to $\mathcal{E}$ is

$$g=\sum_{\chi:\mathcal{E}\to\mathcal{C}} g_\chi.$$

We define a graph $G\in\mathcal{G}$ associated to the invariant g. Let $V(G)=\{v_1,v_2,\dots,v_n\}$ and $E(G)=\{e_1,e_2,\dots,e_{k/2}\}$. The edge e_i connects the vertices v_{i_1} and v_{i_2} where one element of E_i is in the partition set D_{i_1} and the other element of E_i is in the partition set D_{i_2}. In other words, the vertices of G correspond to the partition sets in $\mathcal{D}$, the edges correspond to the partition sets in $\mathcal{E}$ and the edge corresponding to E_i is incident to the vertex corresponding to D_j if and only if $E_1 \cap D_j \neq \emptyset$.

Now the spaces V_i are in a one to one correspondence with the half edges in G and the form m coincides with the one defined in chapter 4.5. It follows that $m(g) = t_d(G)$.

Using Weyl's theorem we have that $\bar{p} = \sum_{i=1}^{r} \lambda_i g_i$ for some real numbers λ_i and invariants g_i where for each g_i there is a graph $G_i \in \mathcal{G}$ with $m(g_1) = t_d(G_i)$. It follows that

$$p_S = m(\bar{p}) = \sum_{i=1}^{r} \lambda_i t_d(G_i) = t_d\Big(\sum_{i=1}^{r} \lambda_i G_i\Big).$$

4.13 Projection to subalgebras of the matrix algebra

Let A be a subalgebra of the full matrix algebra $\mathbb{M}_n(\mathbb{R})$ such that $A = \{M^T | M \in A\}$. The bilinear function $(M, K) = \mathrm{tr}(MK^T)$ defines a euclidean scalar product on $\mathbb{M}_n(\mathbb{R})$. Let $\mathcal{P}_A$ denote the orthogonal projection to A.

Lemma 4.13 *If M is a symmetric positive semidefinite matrix then $\mathcal{P}_A(M) = K^2$ for some symmetric matrix $K \in A$.*

Proof. Since A is invariant under transposing we have that $\mathcal{P}_A(M)$ is a symmetric matrix in A. First we prove that the eigenvalues of $\mathcal{P}_A(M)$ are all nonnegative. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the set of the positive eigenvalues of $\mathcal{P}_A(M)$, let $p(x) = (x - \lambda_1)(x - \lambda_2) \ldots (x - \lambda_k)$ and let $H = p(\mathcal{P}_A(M))$. Using that $\mathcal{P}_A$ is self adjoint and that $H^T = H \in A$ have that

$$\mathrm{tr}(\mathcal{P}_A(M)H^2) = (\mathcal{P}_A(M), H^2) = (M, \mathcal{P}_A(H^2)) = (M, H^2) = \mathrm{tr}(MH^2) = \mathrm{tr}(HMH) \geq 0.$$

Since $\mathrm{tr}(\mathcal{P}_A(M)H^2)$ is a positive linear combination of the negative eigenvalues of $\mathcal{P}_A(M)$ it follows that the eigenvalues of $\mathcal{P}_A(M)$ must be all nonnegative.

Let $g \in \mathbb{R}[x]$ be a polynomial such that $g(\lambda_i) = \sqrt{\lambda_i}$ for $1 \leq i \leq k$. Now $K = g(\mathcal{P}_A(M))$ satisfy both $K^2 = \mathcal{P}_A(M)$ and $K \in A$. ■

4.14 Genaralized Brauer algebras

Let $\mathcal{S}$ denote the set of finite multisets of the positive integers. Let $\mu(S)$ denote the sum of the elements of a multiset $S \in \mathcal{S}$. For each multiset $S \in \mathcal{S}$ we introduce a set $O(S)$ of size $\mu(S)$ and we define a partition $P(S)$ on the elements of $O(S)$ such that the multiset of the sizes of the partition sets in $P(S)$ is S. The algebra A_d consists of the formal linear combinations of triples

$$a(S_1, S_2, M)$$

where $S_1, S_2 \in \mathcal{S}$ and M is a perfect matching on the set $O(S_1) \cup O(S_2)$. The product

$$a(S_1, S_2, M_1)a(S_3, S_4, M_2)$$

is defined to be 0 if $S_2 \neq S_3$. If $S_2 = S_3$ then $M_1 \cup M_2$ is the edge set of a graph G with node set $O(S_1) \cup O(S_2) \cup O(S_4)$ such that nodes in $O(S_2)$ have degree 2 and nodes in $O(S_1) \cup O(S_4)$ have degree 1. This means that G is the union of node disjoint pathes and cycles. Replacing each path by a single edge we get a matching M_3 on $O(S_1) \cup O(S_4)$. Assume that the number of cycles in G is n. The product $a(S_1, S_2, M_1)a(S_2, S_4, M_2)$ is defined to be $d^n a(S_1, S_4, M_3)$.

We introduce the transpose map on A_d as the unique linear extension of the map

$$a(S_1, S_2, M)^T = a(S_2, S_1, M).$$

Let $A_d(S_1, S_2)$ denote the space spanned by the elements $a(S_1, S_2, M)$ where M runs through all perfect matchings of $O(S_1) \cup O(S_2)$. We have that

$$A_d = \bigoplus_{S_1, S_2 \in \mathcal{S}} A_d(S_1, S_2).$$

Let

$$A_d(S) = \bigoplus_{S_1 \in \mathcal{S}} A_d(S_1, S)$$

and

$$A_d(S)^T = \bigoplus_{S_1 \in \mathcal{S}} A_d(S, S_1).$$

For an arbitrary basis element $a(S_1, S_2, M)$ we define $\tau(a(S_1, S_2, M)) \in \mathcal{G}$, $\tau_1(a(S_1, S_2, M)) \in \mathcal{G}_{\mu(S_2)}$ and $\tau_2(a(S_1, S_2, M)) \in \mathcal{G}_{\mu(S_1)}$ in the following way. By identifying nodes in $O(S_1) \cup O(S_2)$ belonging to the same partition set of $P(S_1) \cup P(S_2)$ we get $\tau(a(S_1, S_2, M))$. By identifying nodes in $O(S_1)$ (resp $O(S_2)$) belonging to the same partition set of $P(S_1)$ (resp. $P(S_2)$) and defining $O(S_2)$ (resp. $O(S_1)$) to be the set of open edges we get $\tau_1(a(S_1, S_2, M))$ (resp. $\tau_2(a(S_1, S_2, M))$. The map τ extends linearly to a map $\tau : A_d \to \mathcal{Q}$ and the maps τ_1, τ_2 extend to maps

$$\tau_1 : A_d(S) \to \mathcal{Q}_{\mu(S)} \ , \ \tau_2 : A_d(S)^T \to \mathcal{Q}_{\mu(S)}.$$

Lemma 4.14 *If $b \in A_d$ then $f(\tau(bb^T)) \geq 0$.*

Proof. We use that

$$b = \sum_{S \in \mathcal{S}, \mu(S) \leq m} b_S$$

where $b_S \in A_d(S)$ for all S and m is a large-enough natural number. Since

$$bb^T = \sum_{S \in \mathcal{S}, \mu(S) \leq m} b_S b_S^T$$

it suffices to show that $f(\tau(b_S b_S^T)) \geq 0$ for all S. Using that $\tau_1(b_S) = \tau_2(b_S^T)$ we obtain that $Q = g(\tau_1(b_S), \tau_2(b_S^T))$ is a reflection symmetric quantum graph. The graph $\tau(b_S b_S^T)$ can be obtained from Q by a process where in each step we delete a circle from a graph component and multiply it by d. Using that the circle value of f is d, and that f is multiplicative and edge reflection positive we have that

$$f(\tau(b_S b_S^T)) = f(Q) \geq 0.$$

■

Now we describe a matrix representation of the algebra A_d which will be of crucial importance in the next section. Let us introduce the notation $\mathbb{M}(X,Y)$ for the space of real matrices whose rows are indexed by the set X and whose columns are indexed by the set Y where X and Y are finite sets. Recall that $\mathcal{C}=\{c_1,c_2,\dots,c_d\}$ is a set with d colors. Let $S_1,S_2\in\mathcal{S}$ be two multisets and let M be a perfect matching on $O(S_1)\cup O(S_2)$. We say that a coloring of $O(S_1)\cup O(S_2)$ is compatible with M if the two endpoints of each matching edge have the same color. First we represent $a(S_1,S_2,M)$ by a matrix whose rows are indexed by colorings $O(S_1)\to\mathcal{C}$ and whose columns are indexed by colorings $O(S_2)\to\mathcal{C}$. The entry in the intersection of the row χ and column ψ is 1 if the coloring $\chi\times\psi : O(S_1)\cup O(S_2)\to\mathcal{C}$ is compatible with M and is 0 otherwise. By extending this representation linearly to $A_d(S_1,S_2)$ we obtain a map

$$\omega : A_d(S_1,S_2)\to\mathbb{M}(\mathcal{C}^{O(S_1)}\ ,\ \mathcal{C}^{O(S_2)}).$$

The reader can check easily that the map ω satisfies the identity

$$\omega(a(S_1,S_2,M_1)a(S_2,S_3,M_2))=\omega(a(S_1,S_2,M_1))\omega(a(S_2,S_3,M_2)).$$

It follows that $\omega(b_1b_2)=\omega(b_1)\omega(b_2)$ if $b_1\in A_d(S_1,S_2)$ and $b_2\in A_d(S_2,S_3)$. Let

$$\hat{A}_d(S_1,S_2)=\mathbb{M}(\mathcal{C}^{O(S_1)}\ ,\ \mathcal{C}^{O(S_2)})$$

and let

$$\hat{A}_d=\bigoplus_{S_1,S_2\in\mathcal{S}}\hat{A}_d(S_1,S_2).$$

The space $\hat{A}_d$ is endowed with a natural algebra structure in the following way. Assume that

$$b_1\in\hat{A}_d(S_1,S_2)\ ,\ b_2\in\hat{A}_d(S_3,S_4).$$

If $S_2=S_3$ then b_1b_2 is the usual matrix product and if $S_2\neq S_3$ then b_1b_2 is defined to be 0. This multiplication rule defines a multiplication on the whole space $\hat{A}_d$. It is clear that the map ω extends to an algebra homomorphism $\omega : A_d\to\hat{A}_d$. Let

$$A_{d,r}=\bigoplus_{S_1,S_2\in\mathcal{S},\ \mu(S_1),\mu(S_2)\leq r}A_d(S_1,S_2)$$

and let

$$\hat{A}_{d,r}=\bigoplus_{S_1,S_2\in\mathcal{S},\ \mu(S_1),\mu(S_2)\leq r}\hat{A}_d(S_1,S_2)$$

The space $A_{d,r}$ is a subalgebra of A_d and the space $\hat{A}_{d,r}$ is a subalgebra of $\hat{A}_d$. Moreover, ω maps $A_{d,r}$ into $\hat{A}_{d,r}$. It is easy to see that $\hat{A}_{d,m}$ is the full matrix algebra

$$\mathbb{M}\Big(\bigcup_{S\in\mathcal{S},\ \mu(S)\leq r}\mathcal{C}^{O(S)}\ ,\ \bigcup_{S\in\mathcal{S},\ \mu(S)\leq r}\mathcal{C}^{O(S)}\Big)$$

and that $\omega(b^T)=\omega(b)^T$ for all $b\in A_{d,r}$. This implies in particular that $\omega(A_{d,r})$ is a subalgebra of the matrix algebra $\hat{A}_{d,r}$ which is closed under taking transpose. Let us define the euclidean scalar product $(b_1, b_2) = \mathrm{tr}(b_1 b_2)$ on $\hat{A}_{d,r}$. The spaces $\hat{A}_d(S_1, S_2)$ are orthogonal to each other in $\hat{A}_{d,r}$ for different pairs (S_1, S_2). Since $\omega(A_d(S_1, S_2))$ is contained in $\hat{A}_d(S_1, S_2)$ we have that

$$\omega(A_{d,r}) = \bigoplus_{S_1, S_2 \in \mathcal{S},\ \mu(S_1), \mu(S_2) \leq r} \omega(A_d(S_1, S_2))$$

where all the direct summands are orthogonal to each other. Let $\mathcal{P}_{d,r}$ denote the orthogonal projection of $\hat{A}_{d,r}$ to $\omega(A_{d,r})$ and let $\mathcal{P}_{d,r}(S_1, S_2)$ denote the orthogonal projection of $\hat{A}_{d,r}$ to $\omega(A_d(S_1, S_2))$ where $\mu(S_1), \mu(S_2) \leq r$. The above properties imply that the restriction of $\mathcal{P}_{d,r}$ to the space $\hat{A}_d(S_1, S_2)$ is $\mathcal{P}_{d,r}(S_1, S_2)$.

Now we use the tensor construction from chapter 4.12. Let us observe that the elements of the color basis in the space $T(O(S_1) \cup O(S_2), P(S_1) \cup P(S_2))$ correspond to colorings $O(S_1) \cup O(S_2) \to \mathcal{C}$ and the elementary matrices in $\hat{A}_d(S_1, S_2)$ also correspond to such colorings. This gives an isometry between the euclidean spaces $T(O(S_1) \cup O(S_2), P(S_1) \cup P(S_2))$ and $\hat{A}_d(S_1, S_2)$. By abusing the notation, we will identify the these spaces. Now the tensor construction defines maps $m : \hat{A}_d(S_1, S_2) \to P_d$ for all $S_1, S_2 \in \mathcal{S}$. These maps have a unique common linear extension $m : \hat{A}_d \to P_d$.

Let us observe that

$$T(O(S_1) \cup O(S_2), P(S_1) \cup P(S_2)) = T(O(S_1), P(S_1)) \otimes T(O(S_2), P(S_2))$$

and that for $v_1 \in T(O(S_1), P(S_2)),\ v_2 \in T(O(S_2), P(S_2))$ we have that

$$m(v_1)m(v_2) = m(v_1 \otimes v_2).$$

Let

$$\hat{B}_{d,m} = \bigoplus_{S \in \mathcal{S},\ \mu(S) \leq m} T(O(S), P(S)).$$

Is is clear that

$$\hat{B}_{d,m} \otimes \hat{B}_{d,m} = \hat{A}_{d,m}$$

and that

$$m(v_1 \otimes v_2) = m(v_1)m(v_2)$$

for an arbitrary pair $v_1, v_2 \in \hat{B}_{d,m}$.

4.15 The averaging operator

We define the *averaging operator* $\xi : P_d \to I_d$ by

$$\xi(g) = \int_{\alpha \in O_d(\mathbb{R})} g^{\alpha}\, d\nu$$

where ν is the normalized Haar measure on the orthogonal group $O_d(\mathbb{R})$.

Lemma 4.15 *The following diagram is commutative:*

$$\begin{tikzcd}
 & \hat{A}_{d,r} \arrow[r, "m"] \arrow[d, "\mathcal{P}_{d,r}"] & P_d \arrow[d, "\xi"] \\
A_{d,r} \arrow[r, "\omega"] \arrow[d, "\tau"] & \omega(A_{d,r}) \arrow[r, "m"] & I_d \arrow[d, "\hat{f}"] \\
\mathcal{Q} \arrow[rr, "f"] \arrow[urr, "t_d"] & & \mathbb{R}
\end{tikzcd}$$

Proof. Note that each map on the diagram is $\mathbb{R}$-linear and so it is enough to check the commutativity for an appropriately chosen generating system of the spaces.

First we prove that $\xi \circ m = m \circ \mathcal{P}_{d,m}$ by checking it for the spaces $\hat{A}_d(S_1, S_2)$. Recall that $\hat{A}_d(S_1, S_2)$ is identified with the euclidean space $T = T(O(S_1) \cup O(S_2), P(S_1) \cup P(S_2))$ and that the orthogonal group $O_d(\mathbb{R})$ is acting on T by taking the tensor product of the actions on V. This action commutes with the map m and so we have that

$$m\Big(\int_{\alpha \in O_d(\mathbb{R})} t^\alpha \, d\nu\Big) = \xi(m(t))$$

for all $t \in T$. Since the action of $O_d(\mathbb{R})$ preserves the scalar product on T one gets that $\int_{\alpha \in O_d(\mathbb{R})} t^\alpha \, d\nu$ is the orthogonal projection of t to the space of invariant elements. Therefore it suffices to prove that $\omega(A_{d,r})$ is the space of invariants. This follows from Weyl's first fundamental theorem as described in chapter 4.12.

One gets $m \circ \omega = t_d \circ \tau$ by showing that

$$m(\omega(a(S_1, S_2, M))) = t_d(\tau(a(S_1, S_2, M)))$$

for all triples S_1, S_2, M. This follows immediately from the definitions.

The statement $\hat{f} \circ t_d = f$ is proved in corollary 4.11. ∎

Lemma 4.16 *If $g \in P_d$ then for a sufficiently large natural number r there is a symmetric positive semi-definite matrix M in $\hat{A}_{d,r}$ such that $m(M) = g^2$.*

Proof. If r is a large-enough natural number then

$$g = \sum_{S \in \mathcal{S}, \mu(S) \leq r} g_S$$

where g_S is an element of W_S. Let us represent each g_S by an element t_S in the space $T(O(S), P(S))$ such that $m(t_S) = g_S$. This is possible because m is a surjective map to W_S. Setting

$$t = \sum_{S \in \mathcal{S},\ \mu(S) \leq m} t_S \in \hat{B}_{d,m}$$

we have that $m(t) = g$ and that $m(t \otimes t) = g^2$. On the other hand $t \otimes t$ is represented as a rank 1 positive semi-definite matrix M in $\hat{A}_{d,r}$. ∎

Lemma 4.17 *If $g \in P_d$ then $\hat{f}(\xi(g^2)) \geq 0$.*

Proof. Using lemma 4.16 we get that for a sufficiently large r there is a symmetric positive semi-definite matrix $M \in \hat{A}_{d,r}$ such that $m(M) = g^2$. From lemma 4.13 we obtain that $\mathcal{P}_{d,r}(M) = K^2$ where K is a symmetric matrix from $\omega(A_{d,r})$. Let $\bar{K}$ be a preimage of K under the map ω. We have that $\omega(\bar{K}\bar{K}^T) = K^2$. By lemma 4.14 it follows that $f(\tau(\bar{K}\bar{K}^T)) \geq 0$. Lemma 4.15 implies that

$$f(\tau(\bar{K}\bar{K}^T)) = \hat{f}(m(\omega(\bar{K}\bar{K}^T))) = \hat{f}(m(K^2)).$$

It follows that $\hat{f}(m(\mathcal{P}_{d,r}(M))) \geq 0$. Using lemma 4.15 again we obtain that $\hat{f}(\xi(m(M))) \geq 0$ which completes the proof. ■

4.16 Extension of $\hat{f}$ to P_d

In this section we finish the proof of our main theorem by showing that $\hat{f} : I_d \to \mathbb{R}$ extends to a homomorphism $\bar{f} : P_d \to \mathbb{R}$. This is clearly enough because the edge coloring model defined by

$$t(v) = \bar{f}(x_v), \quad v \in \mathbb{N}^d$$

is a real valued edge coloring model which represents the graph parameter f. We will need the following well known consequence of the so-called Positivestellensatz (see: [13]).

Theorem 4.18 *Let $g \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ be a polynomial such that it has no root in $\mathbb{R}^n$. Then there exist polynomials $p, f_1, f_2, \ldots, f_h$ for some natural number h such that*

$$pg = 1 + f_1^2 + f_2^2 + \cdots + f_h^2.$$

Let $P_{d,r}$ be the subring of P_d which is generated by the variables $\{x_v \mid h(v) \leq r,\ v \in \mathbb{N}^d\}$. Since $P_{d,r}$ is the direct sum of the spaces W_S where S is a multiset of $\{0, 1, \ldots, r\}$ we have that $P_{d,r}$ is invariant under the action of $O_d(\mathbb{R})$. Lemma 4.12 shows that $I_{d,r} = I_d \cap P_{d,r}$ is the set of invariant polynomials in $P_{d,r}$. Let $N_{d,r}$ be the kernel of the homomorphism $\hat{f} : I_{d,r} \to \mathbb{R}$.

Lemma 4.19 *The homomorphism $\hat{f} : I_{d,r} \to \mathbb{R}$ extends to a homomorphism $\bar{f}_r : P_{d,r} \to \mathbb{R}$.*

Proof. First of all note that $P_{d,r}$ is a polynomial ring with $t_r = \sum_{i=0}^{r} \binom{i+d-1}{d-1}$ variables. It suffices to prove that there is a point x in $\mathbb{R}^{t_r}$ which is a common root for all the polynomials in $N_{d,r}$ because the substitution of x into polynomials from P_d would yield a homomorphism of the required form. Let M be the ideal generated by $N_{d,r}$ in $P_{d,r}$. Since $P_{d,r}$ is Noetherian we have that there are finitely many polynomials $g_1, g_2, \ldots, g_k \in N_{d,r}$ which generate M as an ideal. We prove by contradiction that $g_1, g_1, \ldots, g_k$ have a common root. Assume that it is not true. Then $s = \sum_{i=1}^{k} g_i^2$ is a polynomial in $N_{d,r}$ which is positive everywhere in $\mathbb{R}^{t_r}$. Using theorem 4.18 we get that there is a polynomial $p \in P_{d,r}$ such that

$$ps = 1 + f_1^2 + f_2^2 + \cdots + f_h^2$$

for some natural number h. Applying the averaging operator ξ for both sides we get that

$$\xi(p)s = 1 + \xi(f_1^2) + \xi(f_2^2) + \cdots + \xi(f_h^2)$$

because s is invariant under the action of $O_d(\mathbb{R})$. The left side is an element of $N_{d,r}$ since $N_{d,r}$ is an ideal in $I_{d,r}$ and $\xi(p)$ is an element of $I_{d,r}$. This is a contradiction because lemma 4.17 shows that

$$\hat{f}(1+\xi(f_1^2)+\xi(f_2)^2+\cdots+\xi(f_h^2))\geq 1$$

which means that the right side in not an element of $N_{d,r}$. ∎

Lemma 4.20 *The map $\hat{f}: I_d \to \mathbb{R}$ extends to a homomorphism $\bar{f}: P_d \to \mathbb{R}$.*

Proof. Let

$$g_s = \sum_{h(v)=s,\ v\in\mathbb{N}^d} g_v^2.$$

It is easy to see that $g_s = t_d(G_s)$ where G_s is the graph with two nodes which are connected by s edges. This implies that g_s is an element of I_d. Let $\bar{f}_r$ be a map described by lemma 4.19. If $r \geq s$ then g_s is in $P_{d,r}$ and

$$f(G_s) = \hat{f}(g_s) = \sum_{h(v)=s,\ v\in\mathbb{N}^d} \bar{f}_r(g_v)^2.$$

It follows that $|\bar{f}_r(g_v)| \leq \sqrt{f(G_s)}$ for all $v \in \mathbb{N}^d$. Using these inequalities we have that there is an infinite sequence $r_1 < r_2 < \dots$ of natural numbers such that $\hat{f}_{r_i}(x_v)$ is convergent for all fixed vector $v \in \mathbb{N}^d$. This means that $\hat{f}_{r_i}(p)$ is convergent for all polynomial $p \in P_d$ and the limit is a homomorphism which is an extension of $\hat{f}$ to P_d. ∎

5 Limits of functions on groups

The so-called graph limit theory (see [63], [65], [20], [61]) gives an analytic approach to a large class of problems in graph theory. A very active field of applications is extremal graph theory where, roughly speaking, the goal is to find the maximal (or minimal) possible value of a graph parameter in a given family of graphs and to study the structure of graphs attaining the extremal value. A classical example is Mantel's theorem which implies that a triangle free graph H on $2n$ vertices maximizes the number of edges if H is the complete bipartite graph with equal color classes. Another example is given by the Chung-Graham-Wilson theorem [27]. If we wish to minimize the density of the four cycles in a graph H with edge density $1/2$ then H has to be sufficiently quasi random. However the perfect minimum of the problem (that is $1/16$) can not be attained by any finite graph but one can get arbitrarily close to it. Such problems justify graph limit theory where in an appropriate completion of the set of graphs the optimum can always be attained if the extremal problem satisfies a certain continuity property. Furthermore one can use variational principles at the exact maximum or minimum bringing the tools of differential calculus into graph theory.

Extremal graph (and hypergraph) theory has a close connection to additive combinatorics. It is well known that the triangle removal lemma by Szemerédi and Ruzsa implies the qualitative version of Roth's theorem on three term arithmetic progressions. The proof relies on an encoding of an integer sequence (or a subset in an abelian group) by a graph that is rather similar to a Cayley graph. Such representations of additive problems in graph theory hint at a limit theory for subsets in abelian groups that is closely connected to graph limit theory. This new limit theory, that is actually a limit theory for functions on abelian groups, was initiated by the author in [81], [84] and [85] in a rather general form.

Motivated by Szemerédi's theorem on arithmetic progressions Gowers initiated a theory of higher order Fourier analysis in [45] (for a textbook on the topic see [89]). He introduced a sequence of norms $\|.\|_{U_k}$ (called uniformity norms) for functions on finite abelian groups. Roughly speaking, in k-th order Fourier analysis functions with small U_{k+1} norm are considered to be "random like". Seperation of noise and structure is a central topic in higher order Fourier analysis. The bigger k is the more functions are considered to be structured and their description gets increasingly hard. Correspondingly, there is a hierarchy of increasingly fine limit notions related to k-th order Fourier analysis as k goes to infinity and the limit objects get increasingly complex. The focus of this part of the thesis is the linear case $k = 1$ that was called "harmonic analytic limit" in [81]. This case is interesting on its own right, covers numerous important questions and is illustrative for the more general limit concept.

We introduce metric, convergence and limit objects for subsets in abelian groups. More generally, since subsets can be represented by their characteristic functions, we study the convergence of functions on abelian groups. This extends the range of possible applications of our approach to problems outside additive combinatorics.

In the first part of the chapter we study a metric $\hat{d}$ and related convergence notion for l^2 functions on discrete (not necessarily commutative) groups. It is important that the metric $\hat{d}$ allows us to compare two functions defined on different groups. In chapter 5.2 we introduce a distance d for measurable functions $f \in L^2(A_1), g \in L^2(A_2)$ defined on compact abelian groups A_1, A_2 such that $d(f, g) := \hat{d}(\hat{f}, \hat{g})$ where $\hat{f}$ and $\hat{g}$ denote the Fourier transforms of f and g. In additive combinatorics, we can use the distance d to compare subsets in finite abelian groups in the following way. If $S_1 \subseteq A_1$ and $S_2 \subseteq A_2$ are subsets in finite abelian groups A_1 and A_2 then their distance is $d(1_{S_1}, 1_{S_2})$. This allows us to talk about convergent sequences of subsets in a sequence of abelian groups.

A crucial property of the metric d (see theorem 5.9) is that it puts a compact topology on the set of all pairs (f, A) where A is a compact abelian group and f is a measurable function on A with values in a fixed compact convex set $K \subset \mathbb{C}$. As a consequence we have that any sequence of subsets $\{S_i \subseteq A_i\}_{i=1}^{\infty}$ in finite abelian groups A_i has a convergent sub-sequence with limit object which is a measurable function of the form $f : A \to [0, 1]$ where A is some compact abelian group. This result is analogous to graph limit theory where graph sequences always have convergent subsequences with limit object which is a symmetric measurable function of the form $W : [0, 1]^2 \to [0, 1]$.

The success of a limit theory depends on how many interesting parameters are continuous with respect to the convergence notion. The parameters that are most interesting in additive combinatorics are densities of linear configurations. A linear configuration is given by a finite set of linear forms i.e. homogeneous linear multivariate polynomials over $\mathbb{Z}$. For example a 3 term arithmetic progression is given by the linear forms $a,\ a + b,\ a + 2b$. If f is a bounded measurable function on a compact abelian group A then we can compute the density of 3-term arithmetic progressions in f as the expected value $\mathbb{E}_{a,b\in A}(f(a)f(a + b)f(a + 2b))$ according to the normalized Haar measure on A. This density concept can be generalized to an arbitrary linear configuration $\mathcal{L} = \{L_1, L_2, \dots, L_k\}$ and the density of $\mathcal{L}$ in f is denoted by $t(\mathcal{L}, f)$ (see formula (15) and the following sentence.). Gowers and Wolf introduced a complexity notion [47] for linear configurations called *true complexity* (see definition 5.12). A useful upper bound for the true complexity is the so-called Cauchy-Schwarz complexity developed by Green and Tao in [48].

We prove the following fact (for precise formulation see theorem 5.13).

Theorem: *If $\mathcal{L}$ has true complexity at most 1 then the density function of $\mathcal{L}$ is continuous in the metric d.*

Examples for linear configurations of complexity 1 include the 3-term arithmetic progression (this was shown in [48]), the parallelogram $a,\ a+b,\ a+c,\ a+b+c$, and the system $\mathcal{L}_H := \{x_i + x_j\ :\ (i,j) \in E(H)\}$ where H is an arbitrary finite graph on $\{1, 2, \ldots, n\}$. The last example gives a close connection with graph limit theory. The density of $\mathcal{L}_H$ in $f \in L^\infty(A)$ is equal to the density of the graph H in the symmetric kernel $W : A \times A \to \mathbb{C}$ defined by $W(x, y) = f(x+y)$. Note that if f has values in $[0, 1]$ then W is a graphon in the graph limit language. We will elaborate on this connection in chapter 5.9.

Let $\mathcal{L}$ be an arbitrary linear configuration. For $0 \leq \delta \leq 1$ and $n \in \mathbb{N}$ let $\rho(\delta, n, \mathcal{L})$ denote the minimal possible density of $\mathcal{L}$ in subsets of $\mathbb{Z}_n$ of size at least δn. Let $\rho(\delta, \mathcal{L}) := \liminf_{p\to\infty} \rho(\delta, p, \mathcal{L})$ where p runs through the prime numbers. A result by Candela and Sisask [25] implies that the $\liminf$ can be relaced by $\lim$ in the definition of $\rho(\delta, \mathcal{L})$. Note that the qualitative version of Roth's theorem is equivalent with the fact that $\rho(\delta, \mathcal{L}) > 0$ if $\delta > 0$ and $\mathcal{L} = \{a,\ a+b,\ a+2b\}$.

Theorem 5.1 *Let $\mathcal{L}$ be a linear configuration of true complexity at most 1. For every $0 \leq \delta \leq 1$ we have that*

$$\rho(\delta, \mathcal{L}) = \min_{A,f}(t(\mathcal{L}, f))$$

where f runs through all measurable functions of the form $f : A \to [0, 1]$ with $\mathbb{E}(f) = \delta$ on compact abelian groups A with torsion-free Pontrjagin dual groups.

We emphasize that in theorem 5.1 we obtain $\rho(\delta, \mathcal{L})$ as an actual minimum and thus there is some function $f_{\delta,\mathcal{L}}$ realizing the value $\rho(\delta)$. If for example $\mathcal{L} = \{a,\ a+b,\ a+2b\}$ then it is not hard to deduce the qualitative version of Roth's theorem from theorem 5.1 using Lebesgue's density theorem. We sketch the proof at the end of chapter 5.9. It would be very interesting to find the explicit form of a minimizer $f_{\delta,\mathcal{L}}$ for every δ or even to obtain any information on $f_{\delta,\mathcal{L}}$ like on which abelian group it is defined?

It is important to mention that our convergence notion behaves quite differently from usual convergence notions in functional analysis. There is an example for a convergent sequence of functions, all of them defined on the circle (complex unit circle with multiplication or equivalently the quotient group $\mathbb{R}/\mathbb{Z}$), but the limit object exists only on the torus (see the example at the end of chapter 5.2).

In the proofs we will extensively use ultralimit methods. Ultralimit methods in graph and hypergraph regularization and limit theory were first introduced in [33]. There are two different reasons to use these methods. One is that they seem to help to get rid of a great deal of technical difficulties and provide cleaner proofs for most of our statements. The other reason is that they point to an interesting connection between ergodic theory and our limit theory. The ultraproduct $\mathbf{A}$

of compact abelian groups $\{A_i\}_{i=1}^{\infty}$ behaves as a measure preserving system. Our limit concept can easily be explained through a factor $\mathcal{F}(\mathbf{A})$ of $\mathbf{A}$ which is a variant of the so called Kronecker factor.

5.1 A limit notion for functions on discrete groups

For an arbitrary group G we denote by $l^2(G)$ the Hilbert space of all functions $f : G \to \mathbb{C}$ such that $\|f\|_2^2 = \sum_{g\in G} |f(g)|^2 < \infty$. If $f \in l^2(G)$ and $\epsilon \geq 0$ then we denote by $\mathrm{supp}_\epsilon(f)$ the set $\{g : g \in G, |f(g)| > \epsilon\}|$ In particular, $\mathrm{supp}(f) := \mathrm{supp}_0(f)$ is the support of f. Not that if $\epsilon > 0$ then $|\mathrm{supp}_\epsilon(f)| \leq \|f\|_2^2/\epsilon^2$ and $\mathrm{supp}(f) = \cup_{n=1}^{\infty}\mathrm{supp}_{1/n}(f)$ is a countable (potentially finite) set. We denote by $\langle f\rangle$ the subgroup of G generated by $\mathrm{supp}(f)$. It is clear that $\langle f\rangle$ is a countable (potentially finite) group.

Two functions $f_1 \in l^2(G_1)$ and $f_2 \in l^2(G_2)$ are called isomorphic if there is a group isomorphism $\alpha : \langle f_1\rangle \to \langle f_2\rangle$ such that $f_1 = f_2 \circ \alpha$. Let us denote by $\mathcal{M}$ the isomorphism classes of l^2 functions on groups. Our goal is to define a metric space structure on $\mathcal{M}$. We will need the next definition.

Definition 5.2 Let G_1 and G_2 be groups. A partial isomorphism of weight n is a bijection $\phi : S_1 \to S_2$ between two subsets $S_1 \subseteq G_1, S_2 \subseteq G_2$ such that $g_1^{\alpha_1} g_2^{\alpha_2} \dots g_n^{\alpha_n} = 1$ holds if and only if $\phi(g_1)^{\alpha_1}\phi(g_2)^{\alpha_2}\dots\phi(g_n)^{\alpha_n} = 1$ for every sequence $g_i \in S_1, \alpha_i \in \{-1, 0, 1\}$ with $1 \leq i \leq n$.

Definition 5.3 Let $f_1 \in l^2(G_1)$ and $f_2 \in l^2(G_2)$. An ϵ-isomorphism between f_1 and f_2 is a partial isomorphism $\phi : S_1 \to S_2$ of weight $\lceil 1/\epsilon\rceil$ between sets with $\mathrm{supp}_\epsilon(f_1) \subseteq S_1 \subseteq G_1$ and $\mathrm{supp}_\epsilon(f_2) \subseteq S_2 \subseteq G_2$ such that $|f_1(g) - f_2(\phi(g))| \leq \epsilon$ holds for every $g \in S_1$. We define $\hat{d}(f_1, f_2)$ as the infimum of all ϵ's such that there is an ϵ-isomorphism between f_1 and f_2.

Note that both partial isomorphism and ϵ-isomorphism are symmetric notions in the sense that if ϕ is a partial isomorphism (resp. ϵ-isomorphism) the ϕ^{-1} is also a partial isomorphism (resp. ϵ-isomorphism).

Proposition 5.4 *The function $\hat{d}$ is a metric on $\mathcal{M}$.*

Proof. First we show that $\hat{d}(f_1, f_2) = 0$ if and only if f_1 and f_2 are isomorphic. If f_1 is isomorphic to f_2 then it is clear that $d(f_1, f_2) = 0$. For the other direction assume w.l.o.g. that $\|f_2\|_2 \leq \|f_1\|_2$. Let $\alpha_n : S_{1,n} \to S_{2,n}$ be an $1/n$-isomorphism between f_1 to f_2 for every n. Clearly, for every element $g \in \mathrm{supp}(f_1)$ there are finitely many possible elements in the sequence $\{\alpha_n(g)\}_{n=1}^{\infty}$ since $\lim_{n\to\infty} f_2(\alpha_n(g)) = f_1(g)$ and there are finitely many elements h in G_2 on which $|f_2(h)| > |f_1(g)|/2$. Using that the support of f_1 is countable we obtain that there is a subsequence $\{\beta_n\}$ of $\{\alpha_n\}$ such that the sequences $\{\beta_n(g)\}$ stabilize (become constant) after finitely many steps for every g with $|f_1(g)| > 0$. This defines a map $\beta = \lim \beta_n$ from $\mathrm{supp}(f_1)$ to $\mathrm{supp}(f_2)$. It is clear that β extend to an injective homomorphism from $\langle f_1\rangle$ to $\langle f_2\rangle$ and it satisfies $f_2(\beta(g)) = f_1(g)$ for every $g \in \langle f_1\rangle$. Using $\|f_2\|_2 \leq \|f_1\|_2$ it follows that every element in $\mathrm{supp}(f_2)$ is in the image of β and so β is a value preserving isomorphism between $\langle f_1\rangle$ and $\langle f_2\rangle$.

It remains to check the triangle inequality for the metric d. Assume that $\alpha : S_1 \to S_2$ is an ϵ-isomorphism between f_1 and f_2 and assume that $\beta : S_2' \to S_3$ is an ϵ'-isomorphism between f_2 and f_3. Without loss of generality we can assume (by reversing arrows if necessary) that $\epsilon' \geq \epsilon$. We have the following inclusions:

$$\beta^{-1}(\mathrm{supp}_{\epsilon'+\epsilon}(f_3)) \subseteq \beta^{-1}(\mathrm{supp}_{\epsilon'}(f_3)) \subseteq \beta^{-1}(S_3) = S_2',$$

$$\beta^{-1}(\mathrm{supp}_{\epsilon'+\epsilon}(f_3)) \subseteq \mathrm{supp}_{\epsilon}(f_2) \subseteq S_2,$$

$$\alpha(\mathrm{supp}_{\epsilon'+\epsilon}(f_1)) \subseteq \mathrm{supp}_{\epsilon'}(f_2) \subseteq S_2 \cap S_2'.$$

Let $T_2 = \beta^{-1}(\mathrm{supp}_{\epsilon'+\epsilon}(f_3)) \cup \mathrm{supp}_{\epsilon'}(f_2)$ (observe that $T_2 \subseteq S_2 \cap S_2'$) and let $T_1 = \alpha^{-1}(T_2)$, $T_3 = \beta(T_2)$. We have that $\mathrm{supp}_{\epsilon'+\epsilon}(f_1) \subseteq T_1$ and $\mathrm{supp}_{\epsilon'+\epsilon}(f_3) \subseteq T_3$. Let $\gamma : T_1 \to T_3$ be the restriction of $\beta \circ \alpha$ to T_1. Using $T_2 \subseteq S_2 \cap S_2'$ we get that γ is a bijection. To complete the proof of the triangle inequality we show that γ is an $(\epsilon' + \epsilon)$-isomorphism. We have that γ is a bijection and that $|f_1(g_1) - f_3(\gamma(g_1))| \leq \epsilon' + \epsilon$ holds for every $g \in T_1$. It remains to check that γ is a partial isomorphism of weight $\lceil 1/(\epsilon' + \epsilon) \rceil$. This follows form the fact that the composition of a partial isomorphism of weight n and a partial isomorphism of weight m is a partial isomorphism of weight $\min(n, m)$. However the minimum of $\lceil 1/\epsilon \rceil$ and $\lceil 1/\epsilon' \rceil$ is at least $\lceil 1/(\epsilon' + \epsilon) \rceil$. ■

Lemma 5.5 *Assume that a sequence $\{f_i\}_{i=1}^{\infty}$ of l^2 functions on abelian groups converge in $\hat{d}$ to $f \in l^2(G)$. Then $\langle f \rangle$ is also abelian.*

Proof. Let $g_1, g_2 \in \mathrm{supp}(f)$ be two elements. Let $\epsilon = \min(|f(g_1)|/2, |f(g_2)|/2, 1/4)$. Then by convergence of f_i there is an index i such that there is an ϵ-isomorphism ϕ between f and f_i. Since $g_1, g_2 \in \mathrm{supp}_{\epsilon} f$ we have that ϕ is defined on g_1, g_2 and $\phi(g_1)\phi(g_2)\phi(g_1)^{-1}\phi(g_2)^{-1} = 1$ implies that $g_1 g_2 g_1^{-1} g_2^{-1} = 1$ because $\epsilon < 1/4$. ■

For every real number $a > 0$ let $\mathcal{M}_a$ denote the subset of $\mathcal{M}$ consisting of equivalence classes of functions $f \in l^2(G)$ with $\|f\|_2 \leq a$.

Proposition 5.6 *The metric space $(\mathcal{M}_a, \hat{d})$ is compact for every $a > 0$.*

For the proof of proposition 5.6 we will need the next lemma. Let F_r denote the free group in r generators.

Lemma 5.7 *Assume that $\{G_n\}_{n=1}^{\infty}$ is a sequence of groups and for every n we have a sequence of elements $\{g_{n,i}\}_{i=1}^{\infty}$ in G_n. Then there is a sequence of elements $\{g_i\}_{i=1}^{\infty}$ in some group G and a set $S \subseteq \mathbb{N}$ such that for every $r \in \mathbb{N}$ and word $w \in F_r$ there is a natural number N_w such that if $k \in S$ and $k > N_w$ then $w(g_{k,1}, g_{k,2}, \ldots, g_{k,r}) = 1$ if and only if $w(g_1, g_2, \ldots, g_r) = 1$.*

Proof. Let $\{w_i\}_{i=1}^{\infty}$ be an arbitrary ordering of the words in $\cup_{r=1}^{\infty} F_r$ with $w_i \in F_{r_i}$. We construct a sequence of infinite subsets $S_i \subseteq \mathbb{N}$ in a recursive way. Assume that $S_0 = \mathbb{N}$. If S_{i-1} is already constructed then we construct S_i in a way that S_i is an infinite subset in S_{i-1} and either $w_i(g_{s,1}, g_{s,2}, \ldots, g_{s,r_i}) = 1$ holds for every $s \in S_i$ or $w_i(g_{s,1}, g_{s,2}, \ldots, g_{s,r_i}) \neq 1$ holds for every $s \in S_i$. This can be clearly achieved since S_{i-1} is infinite and thus at least one of the two options holds infinitely many times for indices inside S_{i-1}. We then choose a sequence $\{s_i\}_{i=1}^{\infty}$ such that $s_i \in S_i$ and $s_i < s_j$ hold for every pair $i < j$. We obtain for $\{s_i\}_{i=1}^{\infty}$ that for every $r \in \mathbb{N}$ and word $w \in F_r$ either $w(g_{s_i,1}, g_{s_i,2}, \ldots, g_{s_i,r}) = 1$ holds with finitely many exceptions or $w_r(g_{s_i,1}, g_{s_i,2}, \ldots, g_{s_i,r}) \neq 1$ holds with finitely many exceptions. Let W denote the collection of words for which the first case holds. Let G be the group with generators $\{g_i\}_{i=1}^{\infty}$ and relations $\{w(g_1, g_2, \ldots, g_r) = 1 | r \in \mathbb{N}, w \in F_r \cap W\}$. It is clear form the construction of W that every relation that G satisfies in its generators is already listed in W. This follows from the fact that if a word w is not in W then for an arbitrary finite subset W' in W there is a witness among the groups G_{s_i} in which w does not hold but all words in W' hold. Now we have that $S = \{s_i\}_{i=1}^{\infty}$ and G with $\{g_i\}_{i=1}^{\infty}$ satisfies the lemma. ■

Proof of proposition 5.6. Let $\{f_n : G_n \to \mathbb{C}\}_{n=1}^{\infty}$ be a sequence of functions of l^2 norm at most a. For every n let $\{g_{n,i}\}_{i=1}^{\infty}$ be an ordering of the elements in $\mathrm{supp}(f_n)$ in such a way that $|f_n(g_{n,i})| \geq |f_n(g_{n,j})|$ whenever $i < j$. If $\mathrm{supp}(f_n)$ is finite then, to make the list infinite, we add additional elements from outside $\mathrm{supp}(f_n)$ to the list. If the group G_n is finite then we enlarge G_n to an infinite group containing G_n (say $G_n \times \mathbb{Z}$) such that f_n takes the value 0 on the new group elements and then we can make the list infinite with elements from outside G_n.

Let $S \subseteq \mathbb{N}$, G and $\{g_i\}_{i=1}^{\infty}$ be chosen for the sequences $\{g_{n,i}\}_{i=1}^{\infty}$ according to lemma 5.7. Let $S' \subseteq S$ be an infinite subset of S such $a_i := \lim_{n\to\infty, n\in S'} f_n(g_{n,i})$ exists for every $i \in \mathbb{N}$. Now we define the function $f : G \to \mathbb{C}$ such that $f(g_i) = a_i$ inside the set $\{g_i\}_{i=1}^{\infty}$ and $f(g) = 0$ for the rest of the elements. It is clear that f is well defined since $g_{n,i} \neq g_{n,j}$ holds for every n if $i \neq j$ and thus $g_i \neq g_j$. It is clear that $\|f\|_2 \leq \liminf_{n\in S'} \|f_n\|_2$ and thus $\|f\|_2 \leq a$.

To create an ϵ-isomorphism between f and f_n (if $n \in S'$ is big enough) we consider the sets $T_n = \{g_{n,i} : i \leq a^2/\epsilon^2\}$ and the set $T = \{g_i : i \leq a^2/\epsilon^2\}$. Let $\alpha_n : T_n \to T$ be the bijection defined by $\alpha_n(g_{n,i}) = g_i$. It is clear that $\mathrm{supp}_\epsilon(f_n) \subseteq T_n$ holds for every n and that $\mathrm{supp}_\epsilon(f) \subseteq S$. The construction guarantees that $|f_n(g) - f(\alpha_n(g)| \leq \epsilon$ holds if $n \in S'$ is big enough. Furthermore the property given by lemma 5.7 shows that α_n is a partial isomorphism of weight m for an arbitrary $m \in \mathbb{N}$ if $n \in S'$ is big enough. This completes the proof.

5.2 Convergence notions on compact Abelian groups

In this chapter we deal with compact abelian groups, Haar measure, Fourier transform and Pontrjagin duality. The tools that we use are covered in the textbook [77]. Compact abelian groups in this thesis will be assumed to be second countable and thus the Pontrjagin dual group is always countable. For a compact abelian group G we denote by $L^2(G)$ the Hilbert space of Borel measurable complex valued functions f on G with $\int |f|^2 \, d\mu \leq \infty$ where μ is the normalized Haar measure.

If $H \subseteq G$ is a closed subgroup of G then we have that $\tau_H : G \to G/H$ is continuous and Haar measure preserving. Let $L^2_H(G)$ denote the Hilbert sub-space $\tau_H \circ L^2(G/H)$ in $L^2(G)$. We have that $L^2_{H_1}(G) \cap L^2_{H_2}(G) = L^2_{\langle H_1, H_2\rangle}(G)$. It follows that for $f \in L^2(G)$ there is a unique largest closed subgroup $H(G,f)$ such that $f \in L^2_{H(G,f)}(G)$. In other words $H(G,f)$ is the largest closed subgroup of G such that there is a unique function $f' \in L^2(G/H(G,f))$ with $f = \tau_{H(G,f)} \circ f'$. It is clear that $H(G/H(G,f), f')$ is trivial. We say that the function $f' \in L^2(G/H(G,f))$ is the *economic representation* of $f \in L^2(G)$.

The economic representation can also be described through Fourier transforms. Let $\hat{G}$ denote the Pontrjagin dual of G and let $\hat{f} \in l^2(\hat{G})$ denote the Fourier transform of f. For a closed subgroup $H \subseteq G$ we have a natural embedding of $\widehat{G/H}$ into $\hat{G}$. We have that $f \in L^2_H(G)$ if and only $\widehat{G/H}$ contains the support of $\hat{f}$ inside $\hat{G}$. It follows that the economic representation f' of f is the Fourier transform of the restriction of f to the group $\langle \mathrm{supp}(\hat{f})\rangle$ generated by the support of $\hat{f}$. In particular f' is defined on the dual group of $\langle \mathrm{supp}(\hat{f})\rangle$ which the factor group of G with the subgroup H that is the intersection of the kernels of the characters in $\langle \mathrm{supp}(\hat{f})\rangle$.

Let $f_1 \in L^2(G_1)$ and $f_2 \in L^2(G_2)$ be functions on the compact abelian groups G_1 and G_2 with economic representations $(f_1', G_1/H(G_1, f_1))$ and $(f_2', G_2/H(G_2, f_2))$. We say that f_1 and f_2 are *isomorphic* if and only if there is a continuous isomorphism $\phi : G_1/H(G_1, f_1) \to G_2/H(G_2, f_2)$ such that $f_1' = \phi \circ f_2'$. It is clear that this notion of isomorphism is an equivalence relation. Using the above dual description of economic representations we have that f_1 and f_2 is isomorphic if and only $\hat{f}_1$ is isomorphic to $\hat{f}_2$ in the sense of chapter 5.1.

Note that f_1, f_2 are isomorphic if and only if there is a third function $f_3 \in L^2(G_3)$ and continuous epimorphisms $\alpha_i : G_i \to G_3$ for $i = 1, 2$ such that $f_3(\alpha_i(g)) = f_i(g)$ holds for almost every g with respect to the Haar measure in G_i. This follows from the fact that the economic representations of f_1 and f_2 must factor through α_1 and α_2.

Let $\mathcal{H}$ denote the set of isomorphism classes of Borel measurable L^2 functions on compact abelian groups. We introduce the distance d on $\mathcal{H}$ by $d(f_1, f_2) := \hat{d}(\hat{f}_1, \hat{f}_2)$. The metric d induces a convergence notion on $\mathcal{H}$. If we say $\{f_i\}_{i=1}^{\infty}$ is convergent then we mean convergence in d if not stated explicitly in which other meaning it is convergent. Let $\mathcal{H}_a$ denote the set of functions in $\mathcal{H}$ with L^2-norm at most a. Using the fact that Fourier transform preserves the L^2-norm we have by lemma 5.5 and proposition 5.6 the following statement.

Proposition 5.8 *$(\mathcal{H}_a, d)$ is a compact metric space for every $a > 0$.*

For a set $K \subseteq \mathbb{C}$ let $\mathcal{H}(K)$ denote the set of functions in $\mathcal{H}$ which take values in K. We will prove the next theorem.

Theorem 5.9 *If $K \subseteq \mathbb{C}$ is a compact convex set then $(\mathcal{H}(K), d)$ is a compact metric space.*

Corollary 5.10 *If $\{f_i\}_{i=1}^{\infty}$ is a sequence of $\{0, 1\}$ valued functions in $\mathcal{H}$ converging to f in the metric d then the values of f are in the interval $[0, 1]$.*

Theorem 5.9 is somewhat surprising. The metric d is given in terms of Fourier transforms however it is not trivial to relate the set of values of a function to the properties of its Fourier transform. The condition that K is convex turns out to be necessary in theorem 5.9. Corollary 5.10 is useful when we study limits of sets in abelian groups by the limits of their characteristic functions. We give the proof of theorem 5.9 in a later chapter.

In general if $\{f_i\}_{i=1}^{\infty}$ converges to f (in the sense of this chapter) it is not necessarily true that $\{\|f_i\|_2\}_{i=1}^{\infty}$ converges to $\|f\|_2$. We only have that $\limsup_{i\to\infty} \|f_i\|_2 \geq \|f\|_2$. This motivates the next definition. We say that a sequence $\{f_i\}_{i=1}^{\infty}$ in $\mathcal{H}$ is *tightly convergent* if it converges in d and the limit f satisfies $\lim_{i\to\infty} \|f_i\|_2 = \|f\|_2$. Tight convergence can be metrized by the distance

$$d'(f_1, f_2) := d(f_1, f_2) + |\|f_1\|_2 - \|f_2\|_2|.$$

Convergence in d' is stronger than convergence in d and it has stronger consequences. To formulate our result we need the following notation. For a measurable function f on a compact abelian group A we denote by μ_f the probability distribution of $f(x)$ where x is chosen randomly from A according to the Haar measure. The measure μ_f is a Borel probability distribution on $\mathbb{C}$.

Theorem 5.11 *Let $\{f_i\}_{i=1}^{\infty}$ be a sequence of uniformly bounded functions in $\mathcal{H}$ converging to f in d'. Then μ_{f_i} converges to μ_f in the weak topology of measures.*

Note that the above theorem is not true for convergence in d. A trivial example for a tightly convergent sequence is an L^2-convergent sequence of functions on a fixed compact abelian group A. However there are more interesting examples. We finish this chapter with an example which shows that a sequence of L^2 functions on the circle group $\mathbb{R}/\mathbb{Z}$ can have a limit (even in d') which can not be defined on the circle group. The limit object exists on the torus. Let $f_n(x) = e^{2i\pi x} + e^{2in\pi x}$ defined on $\mathbb{R}/\mathbb{Z}$ for $n \in \mathbb{N}$. It is easy to see that f_n is convergent and the limit is the function $f(x,y) = e^{2i\pi x} + e^{2i\pi y}$ on the torus $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$. Note that the sequence f_n is tightly convergent since $\|f_n\|_2 = \|f\|_2 = \sqrt{2}$.

5.3 Densities of linear configurations in functions on Abelian groups

In this chapter we state our main theorem regarding the convergence of the densities of linear configurations of complexity 1. We will follow the language introduced by Gowers and Wolf in [47]. Recall from the introduction that a linear form is a homogeneous linear multivariate polynomial with coefficients in $\mathbb{Z}$. If $L = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n$ is a linear form then we can evaluate it in an arbitrary abelian group A by giving values from A to the variables x_i and thus it becomes a function of the form $L : A^n \to A$. A system $L_1, L_2, \ldots, L_k$ of linear forms determines a type of linear configuration. An example for a linear configuration is the 3-term arithmetic progression which is encoded by the linear forms $x_1,\ x_1 + x_2,\ x_1 + 2x_2$. Assume that A is a compact abelian group and $\mathcal{F} = \{f_i\}_{i=1}^{k}$ is a system of bounded measurable functions in $L^\infty(A)$. Assume furthermore that $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ is a sytem of linear forms in $\mathbb{Z}(x_1, x_2, \ldots, x_n)$. Then it is usual to define the density of the configuration $\mathcal{L}$ in $\mathcal{F}$ by the formula

$$t(\mathcal{L}, \mathcal{F}) := \mathbb{E}_{x_1,x_2,\ldots,x_n \in A} \prod_{i=1}^{k} f_i(L_i(x_1, x_2, \ldots, x_n)). \tag{15}$$

If $f_i = f$ for every $1 \le i \le k$ in the function system $\mathcal{F}$ then we use the notation $t(\mathcal{L}, f)$ for $t(\mathcal{L}, \mathcal{F})$.

In this chapter we address the following type of problem.

Assume that $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ is a linear configuration and $\mathcal{A}$ is a class of compact abelian groups. Under what conditions on $\mathcal{L}$ and $\mathcal{A}$ is the function $f \mapsto t(\mathcal{L}, f)$ continuous in the metric d when functions are assumed to be uniformly bounded measurable functions on groups in $\mathcal{A}$?

The role of the class $\mathcal{A}$ is to exclude certain degeneracies that occur for number theoretic reasons. For example the linear form $2x$ becomes degenerated on the elementary abelian group $(\mathbb{Z}/2\mathbb{Z})^m$. We will need the following definition introduced by Gowers and Wolf in a slightly different form in [47].

Definition 5.12 Let $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ be a linear configuration. The **true complexity** of $\mathcal{L}$ in a class $\mathcal{A}$ of abelian groups is the smallest number $m \in \mathbb{N}$ with the following property. For every $\epsilon > 0$ there exists $\delta > 0$ such that if $A \in \mathcal{A}$ is any abelian group and $\mathcal{F} = \{f_i\}_{i=1}^{k}$ is a system of measurable functions with $|f_i| \le 1$ and $\|f_j\|_{U_{m+1}} \le \delta$ for some j then $t(\mathcal{L}, \mathcal{F}) \le \epsilon$.

In the above definition $\|.\|_{U_{m+1}}$ denotes Gowers's $m+1$-th uniformity norm. Our main theorem states is the following.

Theorem 5.13 *Let $a > 0$. Let $\mathcal{L}$ be a linear configuration and $\mathcal{A}$ be a family of compact abelian groups such that $\mathcal{L}$ has true complexity at most 1 in $\mathcal{A}$. Then $f \to t(\mathcal{L}, f)$ is continuous with respect to the metric d for measurable functions $f \in L^\infty(A)$ with $A \in \mathcal{A}$ and $|f| \le a$.*

5.4 Ultraproducts and ultralimits

Let ω be a non principal ultrafilter on the natural numbers. Let $\{X_i\}_{i=1}^\infty$ be a sequence of sets. For two elements $x = (x_1, x_2, \dots)$ and $y = (y_1, y_2, \dots)$ in the product $\prod_{i=1}^\infty X_i$ we say that $x \sim_\omega y$ if $\{i \;:\; x_i = y_i\} \in \omega$. It is well known that $\sim_\omega$ is an equivalence relation. The set $\prod_\omega X_i := \left(\prod_{i=1}^\infty X_i\right)/\sim_\omega$ is called the *ultraproduct* of the sets X_i.

Let T be a compact Polish space and let $\{t_i\}_{i=1}^\infty$ be a sequence in T. The ultralimit $\lim_\omega t_i$ is the unique point t in T with the property that for every open set U containing t the set $\{i \;:\; t_i \in U\}$ is in ω. (This definition implies that $\lim_\omega t_i$ is always an element of the closure of the set $\{t_i : i \in \mathbb{N}\}$.) Let $\{f_i : X_i \to T\}_{i=1}^\infty$ be a sequence of functions. We define $f = \lim_\omega f_i$ as the function on $\prod_\omega X_i$ whose value on the equivalence class of $(x_1, x_2, \dots)$ is $\lim_\omega f_i(x_i)$.

Let $\{X_i, \mu_i\}_{i=1}^\infty$ be pairs where X_i is a compact Polish space and μ_i is a probability measure on the Borel sets of X_i. We denote by $\mathbf{X}$ the ultraproduct space $\prod_\omega X_i$. The space $\mathbf{X}$ has the following structures on it.

Strongly open sets: We call a subset of $\mathbf{X}$ strongly open if it is the ultraproduct of open sets $\{S_i \subset X_i\}_{i=1}^\infty$.

Open sets: We say that $S \subset \mathbf{X}$ is open if it is a countable union of strongly open sets. Open sets on $\mathbf{X}$ form a σ-topology. This is similar to a topology but it has the weaker axiom that only countable unions of open sets are required to be open.

Lemma 5.14 $\mathbf{X}$ *with the above σ-topology is countably compact. This means that if* $\mathbf{X}$ *is covered by countably many open sets then there is a finite sub-system which covers* $\mathbf{X}$.

Proof. Since every open set is a countable union of strongly open sets it is clearly enough to prove the statement for covering systems of $\mathbf{X}$ with strongly open sets. Let $\{O_i\}_{i=1}^\infty$ be such a system. Now each O_i is the ultra product of open sets $\{W_{k,i} \subseteq X_k\}_{k=1}^\infty$. Let $W_k := \cup_i W_{k,i}$. We have that $\prod_\omega W_k \supseteq \cup_i O_i = \mathbf{X}$ and thus $\prod_\omega W_k = \mathbf{X}$. It follows that $K := \{k : W_k = X_k\}$ is in ω. For each $k \in K$ let $f(k)$ denote the largest natural number such that $\cup_{i=1}^{f(k)} W_{k,i} \neq X_k$ (if $W_{k,1} = X_k$ then $f(k)$ is defined to be 0). By compactness of X_k we have that $f(k)$ is finite. Let us construct a sequence $\{x_k\}_{k=1}^\infty$ such that $x_k \in X_k \setminus (\cup_{i=1}^{f(k)} W_{k,i})$ if $k \in K$ and x_k is arbitrary if $k \in \mathbb{N} \setminus K$. The equivalence class of $(x_1, x_2, \dots) \in \mathbf{X}$ is covered by some element O_t from the covering system. It follows that the set $K' := \{k : x_k \in W_{k,t}\}$ is in ω. This means that $f(k) < t$ holds for every $k \in K \cap K'$ and thus $\cup_{i=1}^t W_{k,i} = X_k$ holds for $k \in K' \cap K$. Since $K \cap K' \in \omega$ we have that $\cup_{i=1}^t O_i = \mathbf{X}$. ■

Borel sets and measurable sets: A subset of $\mathbf{X}$ is called Borel if it is in the σ-algebra generated by strongly open sets. A subset of $\mathbf{X}$ is called measurable if it is in the the completion of the Borel σ-algebra.

Ultralimit measure: If $S \subseteq \mathbf{X}$ is a strongly open set of the form $S = \prod_\omega S_i$ then we define $\mu(S)$ as $\lim_\omega \mu_i(S_i)$. It is a classical fact that μ extends as a probability measure to the σ-algebra of all measurable sets on $\mathbf{X}$. If $\mathbf{X}$ is he ultra product of finite sets then the statement can be found in [33], (See proposition 2.2). The proof of the general case is not much different. A good exposition of the subject is Evan Warner's PhD thesis [90] where the statement is discussed in its full generality.

Ultralimit functions: Let T be a compact Hausdorff topological space. Let $\{f_i : X_i \to T\}_{i=1}^\infty$ be a sequence of Borel measurable functions. We call functions of the form $f = \lim_\omega f_i$ ultralimit functions. It is easy to see that ultralimit functions can always be modified on a 0 measure set that they become measurable in the Borel σ-algebra on $\mathbf{X}$. This means that ultralimit functions are automatically measurable in the completion of the Borel σ-algebra.

Measurable functions: It is an important fact (see proposition 5.1 in [33] and proposition 3.8 in [90]) that every bounded measurable function on $\mathbf{X}$ is almost everywhere equal to some ultralimit function $f = \lim_\omega f_i$.

Continuity: A function $f : \mathbf{X} \to T$ from $\mathbf{X}$ to a topological space T is called continuous if $f^{-1}(U)$ is open in $\mathbf{X}$ for every open set in T. It follows from lemma 5.14 that the image $f(\mathbf{X})$ of a continuous function $f : \mathbf{X} \to T$ in T is countably compact with respect to the restriction of the topology of T to $f(\mathbf{X})$. If T is metrizable then also $f(\mathbf{X})$ is metrizable and thus contably compactness implies compactness. We will need the next lemma.

Lemma 5.15 *A continuous function $f : \mathbf{X} \to \mathbb{R}^n$ is the ultralimit of uniformly bounded continuous functions $f_i : X_i \to \mathbb{R}^n$.*

Proof. Observe first that it is enough to prove the statement for functions of the form $f : \mathbf{X} \to \mathbb{R}$ and the general statement follows by coordinate wise application. We have that $f(\mathbf{X})$ is a compact subset in $\mathbb{R}$ and thus $f(\mathbf{X}) \in (-a, a)$ for some $a \in \mathbb{R}^+$. Let $\epsilon > 0$ be fixed and let $U_i = (-a - \epsilon + i\epsilon/2, -a + i\epsilon/2)$ for $i = 1, 2, \dots, t = \lceil 4a/\epsilon \rceil$. It is clear that the intervals U_i cover $(-a, a)$. For each $1 \le i \le t$ we have that $f^{-1}(U_i)$ is the union of countably many strongly open sets $\{Q_{i,j}\}_{j=1}^\infty$. By $\cup_{i,j} Q_{i,j} = \cup_i f^{-1}(U_i) = \mathbf{X}$ and lemma 5.14 we have that there is a finite sub-system $\{S_k\}_{k=1}^r$ of $\{Q_{i,j}\}_{i,j}$ which covers $\mathbf{X}$. Let us choose points $\{x_k \in S_k\}_{k=1}^r$. Let $\{S_{k,j} \in X_j\}_{j=1}^\infty$ be sequences of open sets for every $1 \le k \le r$ such that $\prod_\omega S_{k,j} = S_k$ holds. Using that $\{S_k\}_{k=1}^r$ covers $\mathbf{X}$ we have that $T := \{j : \cup_k S_{k,j} = X_j\}$ is in ω. For every $j \in T$ we can choose a partition of unity $\{\rho_{k,j} : X_j \to [0,1]\}_{k=1}^r$ subordinated to the covering $\{S_{k,j}\}_{k=1}^r$. The functions $\rho_{k,j}$ are continuous and their sum is the constant 1 function. Furthermore $\rho_{k,j}$ is supported on $S_{k,j}$. If $j \in \mathbb{N} \setminus T$ we define $\rho_{k,j}$ to be 0. Now let $f_j := \sum_{k=1}^r \rho_{k,j} f(x_k)$ for $j \in \mathbb{N}$. Let $f' := \lim_\omega f_j$ and $\rho_k := \lim_\omega \rho_{k,j}$. By the additivity of ultra limits we have that $f' = \sum_{k=1}^r \rho_k f(x_k)$ and that $\sum_{k=1}^r \rho_k = 1_{\mathbf{X}}$. Now let $x \in \mathbf{X}$ be arbitrary. We have that whenever a set $f^{-1}(U_i)$ contains x then $|f(x) - f(x_i)| \le \epsilon$ because $f(U_i)$ has diameter at most ϵ. since $f'(x)$ is a convex combination of the values $\{f(x_i) : x \in U_i\}$ we have that $|f'(x) - f(x)| \le \epsilon$ holds everywhere.

Now for an arbitrary $\epsilon > 0$ we produced a sequence of continuous function $\{f_j\}_{i=1}^\infty$ such that $|\lim_\omega f_j - f|_{\sup} \le \epsilon$ and that $\|f_j\|_{\sup} \le \|f\|_{\sup}$ holds for every $j \in \mathbb{N}$. Now we produce sequences for $m \in \mathbb{N}$ recursively. If $m = 1$ then let $\{f_j^1\}_{j=1}^\infty$ be a sequence which satisfies the above conditions with $\epsilon = 1/2$ for f. In general if $\{f_j^{m-1}\}_{j=1}^\infty$ is already produced then we produce a new sequence

$\{f_j^m\}_{j=1}^{\infty}$ with $\epsilon = 1/2^m$ for the function $f - \sum_{l=1}^{m-1} \lim_\omega f_j^m$. Note that if $m > 1$ then we have that $\|f_j^m\|_{\sup} \leq 1/2^{m-1}$ holds for every j. It follows that $g_j := \sum_{m=1}^{\infty} f_j^m$ is continuous. It is also clear that $\lim_\omega g_j = f$.

■

Lemma 5.16 *Let T be a compact Polish space. Then the ultra limit of continuous functions $\{f_i : X_i \to T\}$ is continuous.*

Proof. Let $f = \lim_\omega f_i$. Let U be an open set in T. We can choose a countable family of open sets $\{W_j\}_{j=1}^{\infty}$ such that $U = \cup_j W_j$ and $\bar{W}_j \subseteq U$. We claim that $f^{-1}(U) = \cup_j \prod_\omega f_i^{-1}(W_j)$. Let j be fixed and assume that $x \in \prod_\omega f_i^{-1}(W_j)$. It follows from the basic properties of ultra limits that $f(x) \in \bar{W}_j \subseteq U$. This implies that $f^{-1}(U)$ contains $\prod_\omega f_i^{-1}(W_j)$ for every j. To see the other containment of the claim let $x \in f^{-1}(U)$. We have that there is j such that $f(x) \in U_j$. Assume that x is the equivalence class of $(x_1, x_2, \dots)$. We have by the properties of ultra limits that $\{i : f_i(x_i) \in U_j\}$ has to be in ω and thus $x \in \prod_\omega f_i^{-1}(W_j)$. ■

5.5 The Fourier σ-algebra

If A is a compact Abelian group then linear characters are continuous homomrphisms of the form $\chi : A \to \mathcal{C}$ where $\mathcal{C}$ is the complex unit circle with multiplication as the group operation. Note that on compact abelian groups we typically use $+$ as the group operation. However if we think of $\mathcal{C}$ as a subset of $\mathbb{C}$ then we are forced to use multiplicativ notation. On the other hand, if we think of $\mathbb{C}$ as the group $\mathbb{R}/\mathbb{Z}$ then we are basically forced to use additive notation.

Linear characters are forming the Fourier basis in $L^2(A)$. In particular linear characters generate the whole Borel σ-algebra on A. Assume now that $\mathbf{A} = \prod_\omega A_i$ is the ultraproduct of compact abelian groups. Linear characters of $\mathbf{A}$ can be similarly defined as for compact abelian groups. In this case we require them to be continuous in the σ-topology on $\mathbf{A}$.

Proposition 5.17 *A function $\chi \in L^\infty(\mathbf{A})$ is a linear character if and only if $\chi = \lim_\omega \chi_i$ for some sequence $\{\chi_i \in L^\infty(A_i)\}_{i=1}^{\infty}$ of linear characters.*

The proof of the proposition relies on a rigidity result saying that almost linear characters on compact groups can be corrected to proper characters. For a function f we denote by f^* the point wise complex conjugate of f.

Lemma 5.18 *For every $\epsilon > 0$ there is $\delta > 0$ such that if $f : A \to \mathbb{C}$ is a continuous function on a compact abelian group A with the property that $|f(x+a)f^*(x) - f(y+a)f^*(y)| \leq \delta$, $||f(x)| - 1| \leq \delta$ for every $x, y, a \in A$ and $|f(0) - 1| \leq \delta$ then there is a character χ of A such that $|\chi(x) - f(x)| \leq \epsilon$ holds for every $x \in A$.*

Proof. As a tool we introduce group theoretic expected values of random variables taking values in $\mathcal{C}$. Let l denote the arc length metric on the circle group $\mathcal{C} \simeq \mathbb{R}/\mathbb{Z}$ normalized by the total length 2π. It is clear that the metric l is topologically equivalent with the complex metric $|x - y|$ on $\mathcal{C}$. Assume that a random variable X takes its values in an arc of the circle group of length $1/3$. Then there is a lift Y of X to $\mathbb{R}$ such that $Y+\mathbb{Z}=X$ and Y takes its values in an interval of length $1/3$. The lift Y with this property is unique up to an integer shift. Then we define $\mathbb{E}(X)\in\mathbb{R}/\mathbb{Z}$ as $\mathbb{E}(Y)+\mathbb{Z}$. Switching to multiplicative notation in $\mathcal{C}$ this expected value satisfies $\mathbb{E}(X_1X_2)=\mathbb{E}(X_1)\mathbb{E}(X_2)$ where X_1,X_2 take values in an arc of length $1/6$.

Let us define $f_2(x)=f(x)/|f(x)|$. If $\delta<1$ then $f(x)\neq 0$ on A and thus f_2 is defined on A. If $\delta>0$ is small enough then for every fixed t the function $x\mapsto f(x+t)f^*(x)$ takes values in an arc of length at most $1/6$. For every $t\in A$ let $g(t)=\mathbb{E}_x(f(x+t)f^*(x))$ where $\mathbb{E}$ is the group theoretic expected value. If δ is small enough then $|g(t)-f(t)|\leq\epsilon$ holds for every $t\in A$ because $|f(x+t)f^*(x)-f(t)f^*(0)|\leq\delta$ and $f(0)$ is close to 1. Using our multiplicativity property of $\mathbb{E}$ we have for every pair $a,b\in A$ that

$$g(a+b)g^*(b)=\mathbb{E}_x(f(x+a+b)f^*(x)f^*(x+b)f(x))=\mathbb{E}_x(f(x+a+b)f^*(x+b))=$$

$$=\mathbb{E}_x((x+a)f^*(x))=g(a).$$

This implies that g is a linear character of A. ■

Now we are ready to prove proposition 5.17

Proof. The continuity of χ guarantees that $\chi=\lim_\omega f_i$ for some sequence of continuous functions f_i on A_i (see lemma 5.15). The fact that χ is a character implies that there is a sequence δ_i such that f_i satisfies the conditions of lemma 5.18 with δ_i for every i and $\lim_\omega\delta_i=0$. It follows by lemma 5.18 that there is a sequence of linear characters χ_i on A_i such that $\lim_\omega\max(|\chi_i-f_i|)=0$. Thus we have that $\lim_\omega\chi_i=\lim_\omega f_i=\chi$. ■

Proposition 5.17 implies that the set of linear characters of $\mathbf{A}$ (also as a group) is equal to $\prod_\omega\hat{A}_i$. We denote this set by $\hat{\mathbf{A}}$. If $f\in L^2(\mathbf{A})$ then the Fourier transform of f on $\mathbf{A}$ is the function $\hat{f}\in l^2(\hat{\mathbf{A}})$ defined by $\hat{f}(\chi)=(f,\chi)$. If $f=\lim_\omega f_i$ then we have that $\hat{f}=\lim_\omega\hat{f}_i$.

It was observed in [83] that linear characters of $\mathbf{A}$ no longer span $L^2(\mathbf{A})$. This shows that in general we only have $\|\hat{f}\|_2\leq\|f\|_2$ instead of equality. Furthermore the σ-algebra $\mathcal{F}(\mathbf{A})$ generated by linear characters on $\mathbf{A}$ is smaller than the whole ultraproduct σ-algebra on $\mathbf{A}$. (The only exception is the case when $\mathbf{A}$ is a finite group. This can happen if the groups A_i are finite and there is a uniform bound on their size.)

We call $\mathcal{F}(\mathbf{A})$ the **Fourier σ-algebra** on $\mathbf{A}$. The fact that the Fourier σ-algebra is not the complete σ-algebra on $\mathbf{A}$ gives rise to the interesting operation $f\mapsto\mathbb{E}(f|\mathcal{F}(\mathbf{A}))$ that isolates the "Fourier part" of a function $f\in L^2(\mathbf{A})$. Using that linear characters of $\mathbf{A}$ are closed with respect to multiplication we obtain that linear characters are forming a basis in $L^2(\mathcal{F}(\mathbf{A}))$. This implies that if $f\in L^2(\mathbf{A})$ then $\hat{f}=\hat{g}$ where $g=\mathbb{E}(f|\mathcal{F}(\mathbf{A}))$. Thus we have that $\|\hat{f}\|_2=\|\hat{g}\|_2=\|\mathbb{E}(f|\mathcal{F}(\mathbf{A}))\|_2$. In particular $\|f\|_2=\|\hat{f}\|_2$ holds if and only if f is measurable in $\mathcal{F}(\mathbf{A})$.

The Fourier σ-algebra has an elegant description in terms of the second Gowers norm U_2. Recall that the U_2 norm [46],[45] of a function $f\in L^\infty(A)$ on a compact abelian group A is defined by

$$\|f\|_{U_2}=\Big(\mathbb{E}_{x,a,b\in A}f(x)f^*(x+a)f^*(x+b)f(x+a+b)\Big)^{1/4}. \tag{16}$$

The next lemma gives a description of the U_2-norm in terms of Fourier analysis.

Lemma 5.19 *If $f\in L^\infty(A)$ then $\|f\|_{U_2}=\|\hat{f}\|_4$ and thus $\|\hat{f}\|_\infty\leq\|f\|_{U_2}\leq(\|f\|_2\|\hat{f}\|_\infty)^{1/2}$.*

One can define $\|f\|_{U_2}$ by the formula (16) for functions on ultraproduct groups. With this definition we have that $\|f\|_{U_2} = \lim_\omega \|f_i\|_{U_2}$ whenever $f = \lim_\omega f_i$. The main difference from the compact case is that $\|.\|_{U_2}$ is no longer a norm for functions in $L^\infty(\mathbf{A})$. It is only a semi-norm. However the next lemma shows that $\|.\|_{U_2}$ is a norm when restricted to $L^\infty(\mathcal{F}(\mathbf{A}))$ and that $\mathcal{F}(\mathbf{A})$ is the largest σ-algebra with this property.

Lemma 5.20 *If $g \in L^\infty(\mathbf{A})$ then $\|g\|_{U_2} = 0$ if and only if g is orthogonal to $L^2(\mathcal{F}(\mathbf{A}))$. A function $f \in L^\infty(\mathbf{A})$ is measurable in $\mathcal{F}(\mathbf{A})$ if and only if f is orthogonal to every function $g \in L^\infty(\mathbf{A})$ with $\|g\|_{U_2} = 0$. In particular we have that $\|.\|_{U_2}$ is a norm on $L^\infty(\mathcal{F}(\mathbf{A}))$.*

Proof. We can assume that $g = \lim_\omega g_i$ for some sequence of functions $\{g_i \in L^\infty(A_i)\}_{i=1}^\infty$ such that $\|g_i\|_\infty \leq \|g\|_\infty$ holds for every i. Assume first that $\|g\|_{U_2} = 0$. Let $\chi = \lim_\omega \chi_i$ be an ultralimit of linear characters. Using lemma 5.19 we have that $|(g_i, \chi_i)| \leq \|\hat{g}_i\|_\infty \leq \|g_i\|_{U_2}$ and thus

$$|(g, \chi)| = \lim_\omega |(g_i, \chi_i)| \leq \lim_\omega \|g_i\|_{U_2} = \|g\|_{U_2} = 0.$$

It follows that g is orthogonal to the space $L^2(\mathcal{F}(\mathbf{A}))$ spanned by linear characters of $\mathbf{A}$. For the other direction assume that $g \neq 0$ is orthogonal $L^2(\mathcal{F}(\mathbf{A}))$. For every i we choose a linear character χ_i on A_i such that $|(g_i, \chi_i)| = \|\hat{g}_i\|_\infty$. We have by lemma 5.19 and by $\|g_i\|_2 \leq \|g_i\|_\infty \leq \|g\|_\infty$ that $|(g_i, \chi_i)| \geq \|g_i\|_{U_2}^2 \|g\|_\infty^{-1}$. Then we have for $\chi = \lim_\omega \chi_i$ that $0 = |(g, \chi)| \geq (\lim_\omega \|g_i\|_{U_2}^2)\|g\|_\infty^{-1}$. It follows that $\|g\|_{U_2} = 0$.

Now we prove the second part of the statement. If $f \in L^\infty(\mathcal{F}(A))$ then by the first part of the statement f has to be orthogonal to every $g \in L^\infty(\mathbf{A})$ with $\|g\|_{U_2} = 0$. For the other direction assume that $f \in L^\infty(\mathbf{A})$ is orthogonal to every $g \in L^\infty(\mathbf{A})$ with $\|g\|_{U_2} = 0$. Let $h := f - \mathbb{E}(f|\mathcal{F}(\mathbf{A})) \in L^\infty(\mathbf{A})$. Note that since $\mathbb{E}$ is an orthogonal projection it follows that $(f, h) = \|h\|_2^2$. From

$$\mathbb{E}(h|\mathcal{F}(\mathbf{A})) = \mathbb{E}(f - \mathbb{E}(f|\mathcal{F}(\mathbf{A}))|\mathcal{F}(\mathbf{A})) = \mathbb{E}(f|\mathcal{F}(\mathbf{A})) - \mathbb{E}(f|\mathcal{F}(\mathbf{A})) = 0$$

we have that h is orthogonal to the whole space $L^2(\mathcal{F}(\mathbf{A}))$ and so by the first statement it follows that $\|h\|_{U_2} = 0$. It implies by our assumption on f that that $(f, h) = 0$ and thus $\|h\|_2^2 = 0$. Now we have that $h = 0$ and $f = \mathbb{E}(f|\mathcal{F}(\mathbf{A}))$ is measurable in $\mathcal{F}(\mathbf{A})$. ∎

Let $\hat{\mathcal{Q}} : L^2(\mathbf{A}) \to \mathcal{M}$ be such that $\hat{\mathcal{Q}}(f)$ is the isomorphism class of $\hat{f}$ in $\mathcal{M}$. Let furthermore $\mathcal{Q}(f)$ denote the isomorphism class in $\mathcal{H}$ representing the Fourier transform of $\hat{\mathcal{Q}}(f)$. Note that $\mathcal{Q}(f) = \mathcal{Q}(\mathbb{E}(f|\mathcal{F}(\mathbf{A})))$. We have that $\mathcal{Q}(f)$ can be represented as a measurable function on some second countable compact abelian group with $\|\mathcal{Q}(f)\|_2 \leq \|f\|_2$ which in some sense imitates f. However it is not even clear from this definition that if f is a bounded function then $\mathcal{Q}(f)$ is also bounded. The next theorem provides a structure theorem for functions in $L^\infty(\mathcal{F}(\mathbf{A}))$ and describes $\mathcal{Q}(f)$.

Theorem 5.21 *A function $f \in L^\infty(\mathbf{A})$ is measurable in $\mathcal{F}(\mathbf{A})$ if and only if there is a continuous, surjective, measure preserving homomorphism $\phi : \mathbf{A} \to A$ to some second countable compact abelian group A and a function $h \in L^\infty(A)$ such that $f = h \circ \phi$ (up to 0 measure change). Furthermore $d(h, \mathcal{Q}(f)) = 0$ implying that the isomorphism class of h is $\mathcal{Q}(f)$.*

Proof. Assume first that $f = h \circ \phi$ for some homomorphism ϕ and function h as in the statement. Let $h = \sum_{i=1}^{\infty} \lambda_i \chi_i$ be the Fourier decomposition of h converging in $L^2(A)$ where χ_i is a sequence of linear characters of A. We have that $\chi_i \circ \phi$ is a linear character of $\mathbf{A}$ for every i. The measure preserving property of ϕ implies that $f = \sum_{i=1}^{\infty} \lambda_i (\chi_i \circ \phi)$ and thus f is measurable in $\mathcal{F}(\mathbf{A})$.

For the other direction assume that $f \in L^\infty(\mathcal{F}(\mathbf{A}))$. Then $f = \sum_{i=1}^{\infty} a_i \chi_i$ for some (distinct) linear characters $\{\chi_i\}_{i=1}^{\infty}$ of $\mathbf{A}$ where the convergence is in $L^2(\mathbf{A})$ and $\|f\|_2^2 = \sum_{i=1}^{\infty} |a_i|^2$. Let us consider the homomorphism $\phi : \mathbf{A} \to \mathcal{C}^\infty$ such that the i-th coordinate of $\phi(x) = \chi_i(x)$. (Recall that $\mathcal{C}$ is the group $\mathbb{R}/\mathbb{Z}$ or equivalently the complex unit circle with respect to multiplication. The group $\mathcal{C}^\infty$ is a compact abelian group.) Using the continuity of ϕ we have that the image A of ϕ is a closed subgroup in $\mathcal{C}^\infty$. Let ν denote the Borel measure on A satisfying $\nu(S) = \mu(\phi^{-1}(S))$ where μ is the ultralimit measure on $\mathbf{A}$. The fact that ϕ is a homomorphism implies that ν is a group invariant Borel probability measure on A and thus ν is equal to the normalized Haar measure. In other words ϕ is measure preserving with respect to the Haar measure on A.

Let us denote by α_i the i-th coordinate function on A. It is clear that $\{\alpha_i\}_{i=1}^{\infty}$ is a system of linear characters of A. Since ϕ is surjective it induces an injective homomorphism $\hat{\phi} : \hat{A} \to \hat{\mathbf{A}}$ defined by $\hat{\phi}(\chi) = \chi \circ \phi$ with the property that $\hat{\phi}(\alpha_i) = \chi_i$ holds for every i. We have that $h = \sum_{i=1}^{\infty} a_i \alpha_i$ (which is defined up to a 0 measure set on A) is convergent in L^2 and has the property that $f = h \circ \phi$ (up to a 0 measure set). The fact that $\hat{\phi}$ is an injective homomorphism implies that $\hat{d}(\hat{h}, \hat{f}) = 0$ and thus $d(h, \mathcal{Q}(f)) = 0$. ∎

If $\mathcal{L}$ is a system of linear forms and $f \in L^\infty(\mathbf{A})$ then we can define $t(\mathcal{L}, f)$ by the formula (15) using the ultralimit measure on $\mathbf{A}$.

Proposition 5.22 *Let $f \in L^\infty(\mathcal{F}(\mathbf{A}))$ and let $\mathcal{L}$ be a system of linear forms. Then $t(\mathcal{L}, f) = t(\mathcal{L}, \mathcal{Q}(f))$. Furthermore if $\mathcal{L}$ has complexity 1 in a family $\mathcal{A}$ of compact abelian groups, $\mathbf{A}$ is an ultraproduct of groups in $\mathcal{A}$ and $f \in L^\infty(\mathbf{A})$ then $t(\mathcal{L}, f) = t(\mathcal{L}, \mathcal{Q}(f))$.*

Proof. For the first part we use theorem 5.21. We get that $f = h \circ \phi$ for some measure preserving homomorphsim $\phi : \mathbf{A} \to A$. It follows that $t(\mathcal{L}, f) = t(\mathcal{L}, h) = t(\mathcal{L}, \mathcal{Q}(f))$. For the sencond part let $f = \lim_\omega f_i$ and $g = \mathbb{E}(f|\mathcal{F}(\mathbf{A})) = \lim_\omega g_i$ for some functions with $\|f_i\|_\infty \leq \|f\|_\infty$ and $\|g_i\|_\infty \leq \|g\|_\infty$. We have that $\lim_\omega \|f_i - g_i\|_{U_2} = \|f - g\|_{U_2} = 0$. Then using that $\mathcal{L}$ has complexity 1 we obtain $t(\mathcal{L}, \mathcal{Q}(f)) = t(\mathcal{L}, \mathcal{Q}(g)) = t(\mathcal{L}, g) = \lim_\omega t(\mathcal{L}, g_i) = \lim_\omega t(\mathcal{L}, f_i) = t(\mathcal{L}, f)$. ∎

5.6 The ultraproduct descriptions of $\hat{d}$ and d convergence

We give a simple and useful description of $\hat{d}$-convergence using ultrafilters. The price that we pay for the simplicity is that we don't get an explicit metric on $\mathcal{M}$, we only get the concept of convergence.

Theorem 5.23 *Let $a > 0$. Assume that $\{f_i\}_{i=1}^{\infty}$ is a sequence in $\mathcal{M}_a$ that converges to f in $\hat{d}$ then f is isomorphic to $\lim_\omega f_i$ for every (non-principal) ultrafilter ω. Consequently a sequence $\{f_i\}_{i=1}^{\infty}$ in $\mathcal{M}_a$ is convergent in $\hat{d}$ if and only if the isomorphism class of $\lim_\omega f_i$-limit doesn't depend on the choice of the ultrafilter ω.*

Proof. For every i let $\alpha_i : T_i \to S_i$ be an ϵ_i-isomorphism between f_i and f with $T_i \subseteq G_i, S_i \subseteq G$ such that $\lim_{i\to\infty} \epsilon_i = 0$. Assume that $\{h_i\}_{i=1}^{\infty}$ represents an element h in $\prod_\omega G_i$ that is in $\mathrm{supp}(g)$

where $g = \lim_\omega f_i$. We have for some set $S \in \omega$ that $|f_i(h_i)| > |g(h)|/2$ and $\epsilon_i \le |g(h)|/4$ for $i \in S$. It follows that $\alpha_i(h_i) \in \mathrm{supp}_{|g(h)|/4}(f)$ holds for every $i \in S$. Since $\mathrm{supp}_{|g(h)|/4}$ is finite we have that $\lim_\omega \alpha_i(h_i)$ exists and it is an element in G that we denote by $\beta(h)$. The map $\beta : \mathrm{supp}(g) \to \mathrm{supp}(f)$ is a partial isomorphism of arbitrary high weight and so it extends to an isomorphism from $\langle g \rangle$ to $\langle f \rangle$. It is clear that β is also an isomorphism between f and g. ■

Corollary 5.24 *Let $a > 0$. Assume that $\{f_i\}_{i=1}^\infty$ is a sequence of functions with $f_i \in L^\infty(A_i)$ and $\|f_i\|_\infty \le a$ for some sequence $\{A_i\}_{i=1}^\infty$ of compact abelian groups. If $\{f_i\}_{i=1}^\infty$ converges to $f \in \mathcal{H}_a$ in the metric d then $f = \mathcal{Q}(\lim_\omega f_i)$ for an arbitrary (non-principal) ultrafilter ω.*

Proof. Since the Fourier transform of $f' = \lim_\omega f_i$ is the ultralimit of the Fourier transforms of f_i we have by theorem 5.23 that $\hat{d}(\hat{f}', \hat{f}) = 0$. It follows that $\mathcal{Q}(f') = f$. ■

Corollary 5.25 *Let $a > 0$. Assume that $\{f_i\}_{i=1}^\infty$ is a convergent sequence of functions with $f_i \in L^\infty(A_i)$ and $\|f_i\|_\infty \le a$ for some sequence $\{A_i\}_{i=1}^\infty$ of compact abelian groups. Then the limit f of $\{f_i\}_{i=1}^\infty$ can be represented as a function on some compact abelian group A such that the dual group of A is a subgroup in $\prod_\omega \hat{A}_i$.*

Proof. We have by corollary 5.24 that $f = \mathcal{Q}(\lim_\omega f_i)$. This means that $\hat{f}$ has an injective embedding into $\hat{\mathbf{A}}$ where $\mathbf{A} = \prod_\omega A_i$. By $\hat{\mathbf{A}} = \prod_\omega \hat{A}_i$ the proof is complete. ■

Corollary 5.25 gives a useful restriction on the structure of the group on which the limit function of a convergent seqence is defined. For example if A_i are growing groups of prime order then the limit function is defined on a compact group whose dual group is torsion-free. On the other hand if p is a fix prime and f_i is defined on $\mathbb{Z}_p^i$ then the limit function is defined on the compact group $\mathbb{Z}_p^\infty$.

5.7 Proofs of theorems 5.9, 5.11, 5.13

For the proofs of theorem 5.9 and theorem 5.11 assume that $\{f_i\}_{i=1}^\infty$ is a convergent sequence in $\mathcal{H}(K)$ for some convex compact set $K \subseteq \mathbb{C}$. Corollary 5.24 implies that the limit is $\mathcal{Q}(f)$ where $f = \lim_\omega f_i$. Note that f takes its values in K. We have that $\mathcal{Q}(f) = \mathcal{Q}(g)$ where $g = \mathbb{E}(f|\mathcal{F}(\mathbf{A}))$. It follows by theorem 5.21 that $g = h \circ \phi$ for some measure preserving homomorphism $\phi : \mathbf{A} \to A$ and the isomorphism class of h is $\mathcal{Q}(g)$. Since g is a projection of f to a σ-algebra we have that g (and thus h) takes its values in K. This completes the proof of theorem 5.9.

For the proof of theorem 5.11 assume that f_i is tightly convergent and $K = \{x : x \in \mathbb{C}, \|x\| \le a\}$. Then, using the above notation we have that $\|g\|_2 = \|h\|_2 = \lim_{i\to\infty} \|f_i\|_2 = \lim_\omega \|f_i\|_2 = \|f\|_2$ where we use tightness in the second equality. This is only possible if $f = g$ and thus $\mu_h = \mu_f = \lim_\omega \mu_{f_i}$ holds. Since this is true for every ultrafilter ω we obtain that $\lim_{i\to\infty} \mu_{f_i} = \mu_h$ holds with respect to weak convergence of measures.

To prove theorem 5.13 assume that $\mathcal{L}$ has complexity 1 and f_i is a d convergent sequence as above. Using the above notation and proposition 5.22 we have that $\lim_\omega t(\mathcal{L}, f_i) = t(\mathcal{L}, f) = t(\mathcal{L}, \mathcal{Q}(f))$ where (using corollary 5.24) $\mathcal{Q}(f)$ is equal to the d-limit of the sequence $\{f_i\}_{i=1}^\infty$. Since this is true for every ultrafilter ω the proof is complete.

5.8 Proof of theorem 5.1

For the proof of theorem 5.1 we will need the next proposition which is interesting on its own right.

Proposition 5.26 *Let B be a compact abelian group with torsion-free dual group and let $f : B \to [0,1]$ be an arbitrary measurable function. Then there are subsets $S_p \subseteq \mathbb{Z}_p$ for every prime number p such that the functions 1_{S_p} converge to f.*

Lemma 5.27 *For every $\epsilon > 0$ there is $N(\epsilon)$ such that if A is a finite abelian group with $|A| \geq N(\epsilon)$ and $f : A \to [0,1]$ is a function then there is a function $h : A \to \{0,1\}$ such that $\|f - h\|_{U_2} \leq \epsilon$.*

Proof. Let us fix $\epsilon > 0$. Let $f : A \to [0,1]$ be a function on a finite abelian group. Let h be the random function on A whose distribution is uniquely determined by the following properties: 1.) h is $\{0,1\}$-valued, 2.) $\{h(a) \mid a \in A\}$ is an independent system of random variables and 3.) $\mathbb{E}(h(a)) = f(a)$ holds for every $a \in A$. We claim that with a large probability the function $h - f$ has U_2 norm at most ϵ if $|A|$ is big enough. Obsereve that $X_a := h(a) - f(a)$ is a random variable for each $a \in A$ with 0 expectation and $\|X_a\|_\infty \leq 1$. The random variables X_a are all independent. Let $\chi : A \to \mathbb{C}$ be a linear character. Then we have that $(h - f, \chi) = |A|^{-1}\sum_{a \in A} X_a \chi(a)$. By Chernoff's bound we have that $\mathbb{P}(|(h - f, \chi)| \geq \epsilon^2)$ is exponentially small in $|A|$. This implies that if $|A|$ is large enough then with probability close to 1 we have that $\|\hat{h} - \hat{g}\|_\infty \leq \epsilon^2$ and thus by lemma 5.19 we get $\|h - g\|_{U_2} \leq \epsilon$ holds in these cases. ■

Proof of proposition 5.26. For a number n let $a(n)$ denote the minimum of $d(1_S, f)$ where S is a subset in $\mathbb{Z}_n$. The statement of the proposition is equivalent with $\lim_{p\to\infty} a(n) = 0$ where p runs through the prime numbers. Assume by contradiction that there is $\epsilon > 0$ and a growing infinite sequence $\{p_i\}_{i=1}^\infty$ of prime numbers with $a(p_i) > \epsilon$. Let $A_i = \mathbb{Z}_{p_i}$ and $\mathbf{A} = \prod_\omega A_i$. We have that $\hat{\mathbf{A}} = \prod_\omega \hat{A}_i \simeq \prod_\omega A_i = \mathbf{A}$. Since $\mathbf{A}$ is not only an abelian group but a field of 0 characteristic with uncountably many elements we have that $\mathbf{A}$ (and thus $\hat{\mathbf{A}}$) as an abelian group is isomorphic to an infinite direct sum of $\mathbb{Q}^+$. It follows that the torsion-free group $\hat{B}$ has an embedding $\hat{\phi} : \hat{B} \to \hat{\mathbf{A}}$ into $\hat{\mathbf{A}}$. This embedding induces a continuous homomorphsim $\phi : \mathbf{A} \to B$ in the way that $\phi(x)$ denotes the unique element in B such that $\chi(\phi(x)) = \hat{\phi}(\chi)(x)$ holds for every $\chi \in \hat{B}$.

Let $g = f \circ \phi$. We have that $g : \mathbf{A} \to [0,1]$ is a measurable function and thus $g = \lim_\omega g_i$ for a system of functions $\{g_i : A_i \to [0,1]\}_{i=1}^\infty$. By lemma 5.27 for every i we can find a $0-1$ valued function g_i' such that $\lim_{i\to\infty} \|g_i' - g_i\|_{U_2} = 0$. By choosing a subsequence we can assume that both $\{g_i'\}_{i=1}^\infty$ and $\{g_i\}_{i=1}^\infty$ are d-convergent. Let $g' = \lim_\omega g_i'$. We have that $\|g - g'\|_{U_2} = 0$ and thus since g is measurable in $\mathcal{F}(\mathbf{A})$ we have that $g = \mathbb{E}(g'|\mathcal{F}(\mathbf{A}))$. By corollary 5.24 we obtain that the d limit of $\{g_i'\}_{i=1}^\infty$ is f. This implies that $0 = \lim d(g_i', f) \geq \liminf a(p_i) \geq \epsilon$ which is a contradiction.

Now we are ready to prove theorem 5.1. First observe that in Proposition 5.26 we can assume with no additional cost that the sets S_p have density at least $\mathbb{E}(f)$. This follows from the fact that their densities converge to $\mathbb{E}(f)$ and so it is enough to set a few values to 1 (with density tending to 0). This observation together with Proposition 5.26 and theorem 15 imply that if $f : A \to [0,1]$ is a measurable function with $\mathbb{E}(f) = \delta$ on an abelian group with torsion-free dual then $\rho(\delta, \mathcal{L}) \leq t(\mathcal{L}, f)$. It remains to find a function where equality holds. For every p prime let $S_p \subseteq \mathbb{Z}_p$ be such that $|S_p|/p \geq \delta$ and that $t(\mathcal{L}, 1_{S_p})$ is minimal possible. We can choose a d-convergent subsequence

$\{f_i\}_{i=1}^{\infty}$ from 1_{S_p} such that $\lim_{i\to\infty} t(\mathcal{L}, f_i) = \rho(\delta, \mathcal{L})$. Let f be the limit of $\{f_i\}_{i=1}^{\infty}$. By theorem 15 we have that $t(\mathcal{L}, f) = \lim_{i\to\infty} t(\mathcal{L}, f_i) = \rho(\delta, \mathcal{L})$. Corollary 5.25 guarantess that f is defined on a group whose dual is torsion-free.

5.9 Connection to dense graph limit theory and concluding remarks

Let H and G be finite graphs. The density of H in G is the probability that a random map from $V(H)$ to $V(G)$ takes edges to edges. We denote this quantity by $t(H, G)$. One can generalize this notion of density for the case when G is replaced by a symmetric bounded measurable function $W : \Omega^2 \to \mathbb{C}$ where (Ω, μ) is a probability space. Then $t(H, W)$ is defined by

$$t(H, W) := \int_{x_1, x_2, \ldots, x_n \in \Omega} \prod_{(i,j)\in E(H)} W(x_i, x_j) \;\; d\mu^n$$

where the verices of H are indexed by $\{1, 2, \ldots, n\}$. It is easy to check that if $\Omega = V(G)$, μ is the uniform distribution on $V(G)$ and $W : V(G)^2 \to \{0, 1\}$ is the adjacency matrix of G then $t(H, G) = t(H, W)$.

In the framework of dense graph limit theory, a sequence of graphs $\{G_i\}_{i=1}^{\infty}$ is called convergent if for every fixed graph H the sequence $\{t(H, G_i)\}_{i=1}^{\infty}$ is convergent. It was proved in [63] that for a convergent graph sequence $\{G_i\}_{i=1}^{\infty}$ there is a limit object of the form of a symmetric measurable function $W : \Omega^2 \to [0, 1]$ (called a *graphon*) such that for every graph H we have $\lim_{i\to\infty} t(H, G_i) = t(H, W)$.

In the above theorem Ω can be chosen to be $[0, 1]$ with the uniform measure however in many cases it is more natural to use other probability spaces. We investigate the case when (Ω, μ) is a compact abelian group A with the normalized Haar measure. Let $f : A \to \mathbb{C}$ be a bounded measurable function and let $W_f : A^2 \to \mathbb{C}$ be defined by $W_f(x, y) := f(x + y)$. As it was pointed out in the introduction, for a finite graph H the density $t(H, W_f)$ is equal to $t(\mathcal{L}, f)$ where $\mathcal{L}_H := \{x_i + x_j \;:\; (i, j) \in E(H)\}$. Using this correspondence and our results in this chapter we get the following theorem on graph limits.

Theorem 5.28 *Let $\{f_i : A_i \to K\}_{i=1}^{\infty}$ be a sequence of measurable functions on compact abelian groups with values in a compact convex set $K \subseteq \mathbb{C}$. Assume that $\lim_{i\to\infty} t(H, W_{f_i})$ exists for every graph H. Then there is a measurable function $f : A \to K$ on a compact abelian group A such that $\lim_{i\to\infty} t(H, W_{f_i}) = t(H, W_f)$ holds for every graph H.*

Proof. By chosing a subsequence we can assume by theorem 5.9 that $\{f_i\}_{i=1}^{\infty}$ is convergent in d with limit $f : A \to K$. Then by theorem 15 we obtain that $\lim_{i\to\infty} t(\mathcal{L}_H, f_i) = t(\mathcal{L}_H, f)$ holds for every graph H. This completes the proof. ■

Theorem 5.28 is closely related to the results in [69]. Let $f : G \to [0, 1]$ be a measurable function on a compact but not necessarily commutative group. Assume that the technical condition $f(g) = f(g^{-1})$ holds for every $g \in G$. Then the function $W : G^2 \to [0, 1]$ defined by $W(x, y) = f(xy^{-1})$ is symmetric. We call graphons of this type Cayley graphons. It was proved in [69] that limits of Cayley graphons are also Cayley graphons. This theorem implies that one can talk about limits of functions on compact topological groups and the limit objects are also functions on compact topological groups.

Another direction of generalization in the commutative setting is when we consider densities of linear configurations of higher complexity. As it was showed in [85], this refinement of the limit concept leads to more complicated limit objects that are measurable functions on nilmanifolds and nilspaces.

As we promised in the introduction of the chapter we finish by showing that theorem 5.1 implies the qualitative version of Roth's theorem. Assume by contradiction that $\rho(\delta,\mathcal{L})=0$ holds for some $\delta>0$. Then there is a function $f:A\to[0,1]$ such that $t(\mathcal{L},f)=0$ with $\mathbb{E}(f)=\delta$. It is easy to see that if S is the support of f then $t(\mathcal{L},1_S)=0$ also holds and $\mathbb{E}(1_S)\geq\delta$. Since A is the inverse limit of finite dimensional torus groups we have that there is a factor map $\tau:A\to\mathbb{T}_n$ to a finite dimensional torus such that $\mathbb{E}(1_S|\tau)>3/4$ holds on a positive measure set $\tau^{-1}(Q)$ where $Q\subseteq\mathbb{T}_n$ is Borel measurable. We have that

$$0=t(\mathcal{L},1_S)\geq t(\mathcal{L},1_S1_{\tau^{-1}(Q)})\geq t(\mathcal{L},1_{\tau^{-1}(Q)})/4=t(\mathcal{L},1_Q)/4$$

where the only nontrivial inequality is the second one. To see this observe that for almost every 3-term arithmetic progression inside $\tau^{-1}(Q)$ a random translate with some element from $\ker(\tau)$ is with probability at least $1/4$ inside $\tau^{-1}(Q)\cap S$. This is true because $\mathbb{E}(1_S1_{\tau^{-1}(Q)})>3/4$ holds inside $\tau^{-1}(Q)$. It remains to show that on $\mathbb{T}_n$ there is no positive density set Q with 0 density copies of $\mathcal{L}$. By Lebesgue density theorem we can find intervals $I_1,I_2,\dots,I_n\subseteq\mathbb{T}_1$ for every $\epsilon>0$ such that Q intersects $C:=\times_{i=1}^n I_i$ in a way that it has density at least $1-\epsilon$ in C. It is easy to see that if ϵ is small enough then $C\cap Q$ most contain a positive density copies of $\mathcal{L}$.

References

[1] D. J. Aldous, *Exchangeability and related topics*, In Ecole d'été de probabilité s de Saint-Flour, XIII—1983, volume 1117 of Lecture Notes in Math., Springer, Berlin, 1985, pp. 1–198

[2] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski, *Random sampling and approximation of MAX-CSPs*, J. Comput. System Sci. **67** (2003), no. 2, pp. 212–243.

[3] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy: Efficient testing of large graphs, *Proc. 40th Ann. Symp. on Found. of Comp. Sci.,* IEEE (1999), 656–666.

[4] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy: Efficient testing of large graphs, *Combinatorica* **20** (2000), 451–476.

[5] N. Alon, A. Naor, *Approximating the Cut-Norm via Grothendieck's Inequality*, SIAM J. Comput. **35** (2006), no. 4, pp. 787–803.

[6] N. Alon, A. Shapira, *Every monotone graph property is testable*, SIAM J. Comput. **38** (2008), no. 2, pp. 505–522.

[7] N. Alon, J. Spencer: *The Probabilistic Method*, Wiley–Interscience, 2000.

[8] M. Atiyah: *The geometry and physics of knots*, Cambridge University Press, Cambridge, 1990.

[9] A. Backhausz, B. Szegedy, *Action convergence of operators and graphs*, arXiv:1811.00626

[10] A. Backhausz, B. Szegedy, *On the almost eigenvectors of random regular graphs*, Annals of Probability **47** (2019), no. 3, pp. 1677–1725.

[11] I. Benjamini, O. Schramm: *Recurrence of distributional limits of finite planar graphs*, Electronic J. Probab. **6** (2001), no. 23, pp. 1–13.

[12] I. Benjamini, O. Schramm, A. Shapira, *Every Minor-Closed Property of Sparse Graphs is Testable*, Advances in Math. **223** (2010), no. 6, pp. 2200–2218.

[13] J. Bochnak, M. Coste, M. Roy: *Real algebraic geometry*, Springer-Verlag, Berlin, 1998.

[14] B. Bollobás, O. Riordan, *Sparse graphs: Metrics and random models*, Random Structures & Algorithms **39**, no. 1, pp. 1–38.

[15] C. Borgs, J.T. Chayes, H. Cohn, Y. Zhao, *An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence*, The Annals of Probability **46** (2018), no. 1, pp. 337-396.

[16] C. Borgs, J.T. Chayes, H. Cohn, Y. Zhao, *An L^p theory of sparse graph convergence I: limits, sparse random graph models, and power law distributions*, Transactions of the American Mathematical Society **372** (2019), no. 5, pp. 3019–3062.

[17] C. Borgs, J.T. Chayes, J. Kahn, L. Lovász, *Left and right convergence of graphs with bounded degree*, Random Structures & Algorithms **42**, no. 1, pp. 1–28.

[18] C. Borgs, J. Chayes, L. Lovász, *Moments of Two-Variable Functions and the Uniqueness of Graph Limits* Geometric and functional analysis **19** (2010), no. 6, pp. 1597–1619.

[19] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, B. Szegedy, K. Vesztergombi *Graph limits and parameter testing*, Proc. of the 38th ACM Symp. Theory of Comp., 2006, pp. 261–270.

[20] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Adv in Math. **219** (2008), no. 6, pp. 1801–1851.

[21] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi, *Convergent graph sequences I: Subgraph frequencies, metric properties and testing*, Advances in Math. **219** (2008), no. 6, pp. 1801–1851.

[22] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent graph sequences II. Multiway Cuts and Statistical Physics* Ann. of Math. **176** (2012), pp. 151–219.

[23] P. Candela, *Notes on nilspaces: algebraic aspects*, Discrete Analysis, (2017), no. 15, pp. 59.

[24] P. Candela, *Notes on compact nilspaces*, Discrete Analysis, (2017), no. 16, pp. 57.

[25] P. Candela, O. Sisask, *Convergence results for systems of linear forms on cyclic groups, and periodic nilsequences*, SIAM Journal on Discrete Mathematics **28** (2014), no. 2, pp. 786–810.

[26] S. Chatterjee, S. R. S. Varadhan, *The large deviation principle for the Erdős-Rényi random graph*. European J. of Comb., **32** (2011), no. 7, pp. 1000–1017.

[27] F. Chung, R.L. Graham, R.M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, pp. 345–362.

[28] W. Cooper, T. Kaiser, D. Král, J. A. Noel, *Weak regularity and finitely forcible graph limits*, Electronic Notes in Discrete Mathematics, **49** (2015), pp. 139–143.

[29] M. Einsiedler, T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011.

[30] G. Elek, *On limits of finite graphs*, Combinatorica **27** (2007), no. 4, pp. 503–507.

[31] G. Elek: *The combinatorial cost*, Enseign. Math. bf53 (2007), pp. 225–235.

[32] G. Elek, G. Lippner, *Borel oracles. An analytical approach to constant-time algorithms*, Proc. Amer. Math. Soc. **138** (2010), no. 8, pp. 2939–2947.

[33] G. Elek, B. Szegedy, *A measure-theoretic approach to the theory of dense hypergraphs*, Adv. in Math. **231** (2012), no. 3-4, pp. 1731–1772.

[34] P. Frankl,V. Rödl, *Extremal problems on set systems*, Random Structures and Algorithms **20** (2002), no. 2, pp. 131–164.

[35] M. H. Freedman, A. Kitaev, C. Nayak, J. Slingerland, K. Walker, Zhenghan Wang, *Universal manifold pairings and positivity*, Geometry & Topology **9** (2005), no. 4, pp. 2303–2317.

[36] M. H. Freedman, L. Lovász, A. Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, Journal of the American Mathematical Society **20** (2007), no. 1, pp. 37–51.

[37] Personal communication with M.H. Freedman, L. Lovász, A. Schrijver

[38] P. Frenkel, *Convergence of graphs with intermediate density*, Transactions of the American Mathematical Society **370** (2018), no. 5, pp. 3363–3404.

[39] A. Frieze, R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, pp. 175–220.

[40] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), no. 1, pp. 204–256.

[41] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press., 2014.

[42] S. Gerke, A. Steger, *The sparse regularity lemma and its applications*, Surveys in combinatorics **327** (2500), pp. 227–258.

[43] R. Glebov, T. Klimosova, D. KralI *Infinite dimensional finitely forcible graphon*, Proceedings of the London Mathematical Society **118** (2019), no. 4, pp. 826–856.

[44] W. T. Gowers, *Lower bounds of tower type for Szemerédi's Uniformity Lemma*, Geom. Funct. Anal. **7** (1997), 322–337

[45] W. T. Gowers, *Fourier analysis and Szemerédi's theorem*, Proceedings of the International Congress of Mathematics **1** (1998), pp. 617–629.

[46] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, pp. 465-588

[47] W. T. Gowers, J. Wolf, *The true complexity of a system of linear equations*, Proc. London Math. Soc. **100** (2010), no. 1, pp. 155–176.

[48] B.J. Green and T. Tao, *Linear equations in primes*, Ann. of Math. **171** (2010), pp. 1753–1850.

[49] O. Goldreich, S. Goldwasser, D. Ron, *Property testing and its connection to learning and approximation*, J. ACM **45** (1998), no. 4, pp. 653–750.

[50] A. Hassidim, J.A. Kelner, H.N. Nguyen, K. Onak: *Local Graph Partitions for Approximation and Testing*, in FOCS '09, 50th Ann. IEEE Symp. on Found. Comp. Science, 2009, pp. 22–31.

[51] H. Hatami, L. Lovász B. Szegedy, *Limits of locally-globally convergent graph sequences*, Geom. Funct. Anal. **24** (2014), no. 1, pp. 269–296.

[52] B. Host, B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, pp. 397–488.

[53] B. Host, B. Kra, *Parallelepipeds, nilpotent groups, and Gowers norms*, Bull. Soc. Math. France **136** (2008), no. 3, pp. 405–437.

[54] B. Host, B. Kra, *Analysis of two step nilsequences*, Annales de l'institut Fourier **58** (2008), no. 5, pp. 1407–1453.

[55] A. Kechris and B.D. Miller: *Topics in orbit equivalence theory*, Lecture Notes in Mathematics **1852**. Springer-Verlag, Berlin, 2004.

[56] A. Kechris, S. Solecki and S. Todorcevic, *Borel chromatic numbers*, Advances in Mathematics **141** (1999), no. 1, pp. 1–44.

[57] Y. Kohayakawa, V. Rödl: *Szemerédi's regularity lemma and quasi-randomness*, in Recent Advances in Algorithms and Combinatorics, CMS Books Math./Ouvrages Math. SMC **11**, Springer, New York, 2003, pp. 289–351.

[58] J. Komlós, M. Simonovits: *Szemerédi's Regularity Lemma and its applications in graph theory*, in D. Miklos et. al, eds., Combinatorics, Paul Erdos is Eighty , Bolyai Society Mathematical Studies **2**, 1996, pp. 295–352.

[59] D. Kunszenti-Kovacs, L. Lovász, B.Szegedy, *Measures on the square as sparse graph limits*, Journal of Combinatorial Theory, Series B **138** (2019), pp. 1–40.

[60] L. Lovász, *The rank of connection matrices and the dimension of graph algebras*, European Journal of Combinatorics **27** (2006), no. 6, pp. 962–970.

[61] L. Lovász, *Large networks and graph limits*, AMS, 2012.

[62] L. Lovász: *Large graphs, graph homomorphisms and graph limits* (forthcoming book).

[63] L. Lovász, B. Szegedy, *Limits of dense graph sequences* Journal of Combinatorial Theory, Series B **96** (2006), no. 6, pp. 933–957

[64] L. Lovász, B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, pp. 933–957.

[65] L. Lovász, B. Szegedy, *Szemerédi's Lemma for the analyst*, Geom. Func. Anal. **17** (2007), no. 1, pp. 252–270.

[66] L. Lovász, B. Szegedy, *Regularity partitions and the topology of graphons*, in An Irregular Mind, Szemerédi is 70, J. Bolyai Math. Soc. and Springer-Verlag 2010, pp. 415–446.

[67] L. Lovász, B. Szegedy, *Testing properties of graphs and functions*, Israel J. of Math. **178** (2010), no. 1, pp. 113–156.

[68] L. Lovász, B. Szegedy, *Finitely forcible graphons*, J. Combin. Theory Ser. B **101** (2011), no. 5, pp. 269–301.

[69] L. Lovász, B. Szegedy, *The automorphism group of a graphon*, Journal of Algebra **421** (2015), pp. 136–166.

[70] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular k-uniform hypergraphs.* Random Structures Algorithms **28** (2006), no. 2, pp. 113–179.

[71] J. Nesetril, P. Ossona de Mendez, *A unified approach to structural limits, and limits of graphs with bounded tree-depth*, American Mathematical Society **263** 2020, no. 1272.

[72] A. Razborov *On the Minimal Density of Triangles in Graphs*, Comb. Prob. and Comp, **17** (2008), no. 4, pp. 603–618.

[73] K. Roth, *On certain sets of integers*, J. London Math Soc. **1** (1953), no. 1, pp. 104–109.

[74] K. Roth, *Irregularities of sequences relative to arithmetic progressions*, IV. Period. Math. Hungar. **2** (1972),no. 1-4, pp. 301–326.

[75] V. Rödl, M. Schacht, *Regular partitions of hypergraphs: regularity lemmas.*, Combin. Probab. Comput. **16** (2007), no. 6, pp. 833–885.

[76] V. Rödl, J. Skokan, *Regularity lemma for k-uniform hypergraphs.* Random Structures Algorithms **25** (2004), no. 1, pp. 1–42.

[77] W. Rudin, *Fourier Analysis on Groups*, Wiley Classics Library, 1962.

[78] O. Schramm, *Hyperfinite graph limits*, Elect. Res. Announce. Math. Sci. **15** (2008), pp. 17–23.

[79] J. Solymosi, *A note on a question of Erdös and Graham.* Combin. Probab. Comput. **13** (2004), no. 2, pp. 263–267.

[80] B. Szegedy, *Edge coloring models and reflection positivity*, J. Amer. Math. Soc. **20** (2007), 969–988

[81] B. Szegedy: *Limits of kernel operators and the spectral regularity lemma*, Europ. J. Combin. **32** (2011), no. 7, pp. 1156–1167.

[82] B. Szegedy, *Limits of functions on groups*, Trans. Amer. Math. Soc. **370** (2018), 8135-8153

[83] B. Szegedy, *Sparse graph limits, entropy maximization and transitive graphs*, arXiv:1504.00858.

[84] B. Szegedy, *Gowers norms, regularization and limits of functions on abelian groups*, arXiv:1010.6211.

[85] B. Szegedy, *On higher order Fourier analysis*, arXiv:1203.2260.

[86] E. Szemerédi, *On sets of integers containing no k-elements in arithmetic progression*, Acta Arithmetica, **27** (1975), pp. 299–345.

[87] E. Szemerédi, *Regular partitions of graphs*, Problémes combinatoires et théorie des graphes, Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976, CNRS, Paris, 1978, pp. 399–401.

[88] T. Tao, *Szemerédi's regularity lemma revisited*, Contributions to Discrete Mathematics, **1** (2006)

[89] T. Tao, *Higher Order Fourier Analysis*, Graduate Studies in Mathematics **142** 2012.

[90] E. Warner, *Ultraproducts and the Foundations of Higher Order Fourier Analysis*, PhD Thesis, 2012

[91] H. Weyl: *The Classical Groups, their Invariants and Representations*, Princeton Mathematical Series **1**, Princeton Univ. Press, Princeton, 1946.

[92] D. Williams: *Probability with Martingales*, Cambridge Univ. Press, 1991.

[93] T. Ziegler, *Universal characteristic factors and Fürstenberg averages* J. Amer. Math. Soc. **20** (2007), no. 1, pp. 53—97.