Cairo University
Faculty of Sciences
Department of Mathematics

Benjamini-Schramm convergence of graphs

Mostafa SABRI

Mathematics Summer School 2019

Contents

Chapter 1

Background

The aim of this project is to investigate a notion of *convergence* for sequences of graphs. The limit will generally be a *random graph.* One important question we will study is the behavior of the *spectrum* of the graph as it gets large.

This preliminary chapter will thus cover some background on each of the previous keywords :

— can we define the convergence of some objects more abstract than sequences of numbers ?

— what does "random" means ?

— what is graph ?

— what is the spectrum of an operator ?

These questions will lead us to introduce *metric spaces*, which provide a framework which is sufficiently abstract for our purposes and sufficiently concrete to be intuitive. We will also discuss special metric spaces, known as *Banach spaces* and *Hilbert spaces*, which offer a good environment to do *functional analysis.* We will then discuss the notion of randomness by introducing *probability measures.* This naturally leads to consider general measures, and will be the occasion to introduce a basic Hilbert space denoted $L^2(X, \mu)$. We will next move on to discuss basic definitions of graphs (everything being very intuitive). Graphs are equipped with natural operators called *adjacency matrices.* We will define the notion of the *spectrum* for general bounded linear operators and discuss the *spectral theorem.* We conclude this chapter with more probabilistic content which will shed more light on several aspects of the project.

1.1 Metric spaces, Functional Analysis

A metric space is a set endowed with a function that measures the *distance* between its elements. As one imagines, the distance between two elements should be non-negative (in fact positive if they are distinct) and symmetric. Furthermore, it should somehow retain the idea that it measures "the shortest path" between two points. So if x, y, z are three points and we denote $d(x, y)$ the distance between x and y,...etc, then we should have $d(x, z) \leq d(x, y) + d(y, z)$, since we can go from x to z using the path from x to y to z, whose "length" is $d(x, y) + d(y, z)$, so the shortest path should be smaller.

Well, that's all there is about metrics :

Definition 1.1. A *metric space* is a set X endowed with a function $d : X \times X \to [0, \infty)$ satisfying the following properties for any $x, y, z \in X$:

1) $d(x, y) \geq 0,$ (non-negative)

2) $d(x,y) = 0 \iff x = y$,

3) $d(x,y) = d(y,x)$, (symmetric)

4) $d(x,y) \leq d(x,z) + d(z,y)$. (triangle inequality)

Here the second point says two things : first, $d(x,x) = 0$, second, $d(x,y) > 0$ if $y \neq x$.

Example 1.2. The natural distance in the plane $\mathbb{R}^2$ intuitively satisfies these properties (though we needed Euclid's Elements Book 1 for a proof). On $\mathbb{R}^n$, we can define $d(x,y) := (\sum_{k=1}^n |x_k - y_k|^2)^{1/2}$ for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, which is called the *Euclidean distance*, and generalizes the familiar distance in the plane. The proof of the triangle inequality takes some work.

Actually, for any $p \geq 1$, we can define $d(x,y) := (\sum_{k=1}^n |x_k - y_k|^p)^{1/p}$, which still satisfies all the properties of a metric. The triangle inequality in this context is called *Minkowski's inequality.* The proof uses yet another inequality called *Hölder's inequality.* We omit both proofs; see e.g. [26].

Example 1.3. The space $\mathbb{R}^n$ is very special (in the sense that it has a very rich structure). We can actually define a metric on any set X : just put $d(x,y) = 0$ if $x = y$ and $d(x,y) = 1$ otherwise. We call this the *discrete metric.* It is admittedly weird geometrically, but it's a good example to keep in mind.

The reader can search for more examples in books or the internet. An important metric will be defined later in our project.

As mentioned in the second example, $\mathbb{R}^n$ is very special. Besides having a natural distance, it is also a *vector space*, i.e. has a zero element, a notion of addition, and a notion of scalar multiplication. Let us explore such spaces further :

Definition 1.4. Let X be a *vector space* over the field $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

We say that $\|\cdot\| : X \to [0, \infty)$ is a *norm* on X if it satisfies the following properties for each $x, y \in X$ and $\alpha \in \mathbb{K}$:

1) $\|x\| \geq 0$,

2) $\|x\| = 0 \iff x = 0$,

3) $\|\alpha x\| = |\alpha| \cdot \|x\|$,

4) $\|x + y\| \leq \|x\| + \|y\|$.

We call $(X, \|\cdot\|)$ a *normed space.*

Clearly, any normed space is a metric space if we define the distance $d(x,y) := \|x - y\|$. Norms thus measure the "length" of elements, i.e. their distance from 0.

Example 1.5. The space $\mathbb{C}^n$ over the field $\mathbb{C}$, endowed with $\|x\|_p := (\sum_{k=1}^n |x_k|^p)^{1/p}$ is a normed space for any $p \geq 1$.

The same space can be endowed with the norm $\|x\|_\infty := \max_{1 \leq k \leq n} |x_k|$.

Example 1.6. What if we replace $\mathbb{C}^n$ by $\mathbb{C}^\mathbb{N}$; the space of all $x = (x_1, x_2, \ldots)$ with $x_j \in \mathbb{C}$? This is the *space of all sequences*, it's a very big creature.

Let us consider smaller spaces as follows : let $\ell^p := \{x \in \mathbb{C}^\mathbb{N} : \sum_{k=1}^\infty |x_k|^p < \infty\}$. This is the subspace of p-summable sequences. It has a natural norm $\|x\|_p := (\sum_{k=1}^\infty |x_k|^p)^{1/p}$.

Example 1.7. Let $C[a,b]$ be the space of all continuous functions on $[a,b]$. This is a normed space if we endow it with $\|f\| := \max_{t \in [a,b]} |f(t)|$ for $f \in C[a,b]$.

As previously mentioned, any of these normed spaces is in particular a metric space.

Are there interesting spaces which are yet more special than normed spaces ? Well besides having a notion of length, $\mathbb{R}^n$ also has a notion of *angle*. Let's explore such spaces further :

Definition 1.8. Let X be a vector space on $\mathbb{K}$. An *inner product* on X is a map $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{K}$ satisfying the following properties for any $x, y, z \in X$ and $\alpha, \beta \in \mathbb{K}$:

1) $\langle x, x\rangle \geq 0$,
2) $\langle x, x\rangle = 0 \iff x = 0$,
3) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$, (linearity)
4) $\langle y, x\rangle = \overline{\langle x, y\rangle}$. (conjugate symmetry)

We call $(X, \langle\cdot,\cdot\rangle)$ an *inner product space.*

Example 1.9. The basic example is the dot product in $\mathbb{R}^3$: $\langle x, y\rangle = x \cdot y = x_1y_1 + x_2y_2 + x_3y_3$.

In general, on $\mathbb{C}^n$, we can define the inner product $\langle x, y\rangle := \sum_{k=1}^{n} x_k\overline{y_k}$. On $\mathbb{R}^n$ this becomes $\langle x, y\rangle = \sum_{k=1}^{n} x_k y_k$.

The space ℓ^2 can be endowed with the inner product $\langle x, y\rangle := \sum_{k=1}^{\infty} x_k\overline{y_k}$ for $x = (x_k)_{k\geq 1}, y = (y_k)_{k\geq 1} \in \ell^2$.

Recall that for the dot product, we can find the angle θ between two vectors in $\mathbb{R}^3$ via the formula $x \cdot y = \|x\| \cdot \|y\| \cos\theta$. In particular, $|x \cdot y| \leq \|x\| \cdot \|y\|$. In general,

Theorem 1.10. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. Then*

a) *X is in particular a normed space, with norm $\|x\| := \sqrt{\langle x, x\rangle}$,*
b) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. (Cauchy-Schwarz inequality).

Definition 1.11. We say that two vectors x, y in $(X, \langle\cdot,\cdot\rangle)$ are *orthogonal* if $\langle x, y\rangle = 0$. We denote this by $x \perp y$.

More generally, one could define an "angle" θ between vectors in an inner product space via $\langle x, y\rangle = \|x\| \cdot \|y\| \cos\theta$ and recover much of the familiar plane geometry.

So far we introduced inner product spaces, which are special normed spaces, which are special metric spaces. Let us turn back to general metric spaces and remember the problem of the introduction, namely, to define a notion of convergence for something more abstract than sequences of numbers.

Definition 1.12. Let (X, d) be a metric space and $\{x_k\}_{k\geq 1} \subset X$ a sequence of elements of X. We say that

(i) (x_k) is *Cauchy* if for any $\varepsilon > 0$ there is $n_0 = n_0(\varepsilon)$ such that $d(x_n, x_m) < \varepsilon$ for all $n \geq n_0$,

(ii) (x_k) *converges to* $x \in X$ if $d(x_n, x) \to 0$.

The definitions just mimic those of $\mathbb{R}$ and $\mathbb{C}$. As in those spaces, *any convergent sequence is Cauchy.* The proof is easy.

Recall that in $\mathbb{R}$, we also know that any Cauchy sequence converges. In metric spaces, this is not necessarily true. For example, if $X = (0, 1)$ and $x_n = \frac{1}{n}$, we check that (x_n) is Cauchy but does not converge (its "limit" 0 is missing from X).

Definition 1.13. We say that a metric space (X, d) is *complete* if any Cauchy sequence in X converges to some point $x \in X$.

Example 1.14. The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are complete when endowed with $\|\cdot\|_p$. As we saw, $X = (0,1)$ with the natural metric is not complete.

Theorem 1.15. *Any metric space has a* completion*, which is unique up to isometry.*

Roughly speaking, this says that for any metric space X, there is a larger, complete metric space $\widehat{X}$ such $X \subseteq \widehat{X}$ and X is *dense* in $\widehat{X}$.

For example, the completion of $X = (0,1)$ in the natural metric is $\widehat{X} = [0,1]$. To explain the previous notions better, we need some definitions :

Definition 1.16. Let X be a metric space. We say that $M \subseteq X$ is *closed* if

$$(\forall \{x_n\} \subset M : x_n \to x) \implies x \in M\,.$$

In other words, M should contain all its limit points.

Definition 1.17. The *closure* of $M \subseteq X$ is the smallest closed set containing M. We denote it by $\overline{M}$.

We say that M is *dense* in X if $\overline{M} = X$.

So as the language suggests, a dense subset has points in all neighborhoods in X. For instance, $\mathbb{Q}$ is dense in $\mathbb{R}$.

Theorem 1.15 thus says that the bigger space on one hand is complete, and on the other hand is the smallest complete space containing [1] X.

Example 1.18. The space $C[a,b]$ with the norm $\|f\| = \max_{t\in[a,b]} |f(t)|$ is complete.

The space $C[a,b]$ can also be endowed with the norm $\|f\|_p = (\int_a^b |f(t)|^p \,\mathrm{d}t)^{1/p}$. Checking that this is indeed a norm is nontrivial.

It turns out that $(C[a,b], \|\cdot\|_p)$ is not complete. Its completion is the space $L^p[a,b]$, further discussed in the next section.

Before moving to another concept, we record the following :

Lemma 1.19. *If X is a complete metric space and $M \subseteq X$ is closed, then M is complete.*

Conversely, if M is a complete subset of some metric space X, then M is closed.

The basic spaces of functional analysis can now be defined :

Definition 1.20. A *Banach space* is a complete normed space.

A *Hilbert space* is a complete inner product space.

Here completeness is in the metric induced by the norm.

Example 1.21. The ℓ^p spaces are Banach spaces. The spaces $L^p[a,b]$ are also Banach.

The spaces ℓ^2 and $L^2[a,b]$ are Hilbert spaces.

Any finite-dimensional normed space is a Banach space. Similarly, any finite-dimensional inner product space is a Hilbert space.

We still need some metric notions before moving on.

Definition 1.22. We say that a metric space X is *separable* if it has a dense countable subset.

1. Actually, this interpretation is not entirely true : the completion $\widehat{X}$ generally contains an isometric copy of X, not X itself - but it helps to think this way.

A countable subset M is a set... whose elements can be counted. That is, we may write the set completely as an enumeration $M = \{x_1, x_2, x_3, \dots\}$. For example, $\mathbb{N}$ is clearly countable. $\mathbb{Z}$ is also countable. $\mathbb{Q}$ is countable. But $\mathbb{R}$ is uncountable. For any list of real numbers $\{x_1, x_2, \dots\}$, we can find a real $y \notin \{x_1, x_2, \dots\}$. So although $\mathbb{N}$ and $\mathbb{R}$ are both infinite subset, $\mathbb{R}$ has a "bigger infinity" of elements. The German mathematician Cantor delighted in these things

The space $\mathbb{R}$ with the natural metric is separable, as it contains the countable subset $\mathbb{Q}$. The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are separable for analogous reasons.

Definition 1.23. Let (X, d) be a metric space. We say that $M \subseteq X$ is *compact* if any sequence in M has a subsequence which converges to some $x \in M$.

Compact subsets of $\mathbb{R}$ are well-known : they are precisely the subsets which are closed and bounded. The difficult side of this statement is the well-known *Bolzano-Weierstrass* theorem. More generally,

Lemma 1.24. *If X is a finite-dimensional normed space, then $M \subset X$ is compact iff it is closed and bounded.*

In general only one side of the above holds :

Lemma 1.25. *Let X be a metric space. If $M \subseteq X$ is compact, then M is closed and bounded.*

The converse is generally untrue. For example, let X be a normed space and $\widetilde{B} \subset X$ the *closed unit ball*, that is, $\widetilde{B} = \{x \in X : \|x\| \leq 1\}$. It is easy to show that $\widetilde{B}$ is closed and bounded. However :

Lemma 1.26 (Riesz)**.** *If X is infinite-dimensional, then $\widetilde{B}$ is never compact.*

We conclude this section with the notion of a bounded linear operator, which is an object of central importance in functional analysis.

Definition 1.27. Let $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$ be normed spaces. We say that $T : X \to Y$ is a *bounded linear operator* if T is

(i) linear : $T(\alpha x + \beta y) = \alpha Tx + \beta Ty$ for any $x, y \in X$ and $\alpha, \beta \in \mathbb{K}$,

(ii) and bounded, i.e. satisfies

$$\|T\| := \sup_{x \neq 0} \frac{\|Tx\|_2}{\|x\|_1} < \infty \,.$$

We denote $B(X, Y)$ the set of bounded linear operators from X to Y.

We call $\|T\|$ the *operator norm* of T, and sometimes emphasize this by denoting $\|T\|_{op}$.

Example 1.28. The identity map $I : X \to X$, $I : x \mapsto x$, is a bounded linear operator [2], with $\|I\| = 1$. Similarly, the zero operator $\mathbf{0} : x \mapsto 0$ is a b.l.o with $\|\mathbf{0}\| = 0$.

The following theorem shows that the notion of boundedness is only important when moving to infinite dimensions :

2. Note that $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$ is an unbounded function in the context of calculus - so the definition here is different. In functional analysis, we measure how much the operator *increases* the norm of each x, while in calculus we just measure the *size* of the whole range.

Theorem 1.29. *If X is finite-dimensional, then any linear operator on X is bounded.*

Actually, when both X and Y are finite-dimensional, then bounded linear operators are precisely the matrices the reader studied in linear algebra.

Example 1.30. Let $M \subset \ell^p$ be the subspace of sequences $x = (x_1, x_2, \dots)$ with finitely many nonzero terms (i.e. for any x there is n_x such that $x_j = 0$ for all $j > n_x$).

Define $T : M \to M$ by $T : (x_j)_{j\geq 1} \mapsto (jx_j)_{j\geq 1}$. One easily sees that T is unbounded.

A very important operator is unfortunately unbounded :

Example 1.31. Consider $X = C[a,b]$ with $\|f\| = \max_{t\in[a,b]} |f(t)|$. Let $M \subset C[a,b]$ be the subspace of continuously differentiable functions. Define $T : M \to C[a,b]$ by $T : f \mapsto f'$, i.e. T is the differentiation operator. Then T is unbounded.

This is the main reason we still need a theory of unbounded operators - but this will not concern us in this course. We note that integration is however a bounded linear operator.

Let us conclude with operators on Hilbert spaces.

Definition 1.32. Let $(\mathcal{H}_1, \langle\cdot,\cdot\rangle_1)$, $(\mathcal{H}_2, \langle\cdot,\cdot\rangle_2)$ be Hilbert spaces and $T \in B(\mathcal{H}_1, \mathcal{H}_2)$.

We define the *adjoint* T^* of T to be the unique operator $T^* \in B(\mathcal{H}_2, \mathcal{H}_1)$ satisfying

$$\langle Tx, y\rangle_2 = \langle x, T^*y\rangle_1 \qquad \forall x \in \mathcal{H}_1, y \in \mathcal{H}_2 .$$

It can be shown that T^* indeed exists and is unique. Moreover, $\|T^*\| = \|T\|$. The proof of these things uses the *Riesz representation theorem* - a fundamental result in Hilbert space theory.

Example 1.33. Let $A = (a_{i,j})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$ be a matrix, denote the associated operator also by A, so $A : \mathbb{C}^n \to \mathbb{C}^m$ takes $(x_1, \dots, x_n) \mapsto (\sum_{j=1}^n a_{1,j}x_j, \dots, \sum_{j=1}^n a_{m,j}x_j)$. Then $A^* : \mathbb{C}^m \to \mathbb{C}^n$ is the matrix $A^* = (\overline{a_{j,i}})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$. In other words, A^* is the conjugate transpose (just the transpose when entries are real).

Example 1.34. More trivially, $I^* = I$ and $\mathbf{0}^* = \mathbf{0}$.

Definition 1.35. We say that $T : \mathcal{H} \to \mathcal{H}$ is *self-adjoint* if $T^* = T$.

Note that self-adjointness only makes sense if the operator goes from the space to itself.

Self-adjoint operators are the bread and butter of quantum mechanics (though unfortunately, they are typically unbounded).

Example 1.36. Let $X = L^2[a,b]$, φ a continuous function and $M_\varphi : X \to X$ be the *multiplication operator* $M_\varphi : f \mapsto \varphi f$. Then M_φ is self-adjoint iff φ is real-valued. In fact, $M_\varphi^* = M_{\overline{\varphi}}$.

More classes of interest :

Definition 1.37. We say that $U : \mathcal{H} \to \mathcal{H}$ is *unitary* if it is bijective and $U^{-1} = U^*$.

We say that $T : \mathcal{H} \to \mathcal{H}$ is *normal* if $TT^* = T^*T$.

Further reading : See e.g. [26, Chapters 1–3] for much more details.

1.2 (Probability) Measures

Perhaps the best known "measure" is what we call the *Lebesgue measure*, though it's certainly not the simplest one.

The purpose of the Lebesgue measure is to extend the notion of length from intervals to arbitrary subsets of $\mathbb{R}$, similarly, the notion of volume of cubes in $\mathbb{R}^n$ to arbitrary subsets. What is the "length" of $\mathbb{Q}$ for example ?

Unfortunately this ideal aim turns out to be impossible to achieve. A paradox by Banach and Tarski says that one can break the unit ball in $\mathbb{R}^3$ into a finite number of pieces, rotate and translate them, then reassemble them to obtain two copies of the unit ball ! In particular, since the volume of the ball, call it τ, should be the sum of the volume of the pieces, we get $\tau = 2\tau$.

The problem here is that the pieces actually don't have a well-defined "volume". We say they are *not measurable.*

So although we can extend the notion of "size" to a wide family of subsets of $\mathbb{R}^n$, we cannot extend it to *all* subsets. The family of "good" sets is called the Borel σ-algebra.

The construction of Lebesgue's measure contains most of the main difficulties, so we can move on directly to discuss general measures on sets X. This is what we really need for the project later on.

Definition 1.38. Let X be a set. A family of subsets $\mathfrak{F}$ of X is called an *algebra* if

(i) $X \in \mathfrak{F}$,

(ii) $\mathfrak{F}$ is closed under finite unions,

(iii) $\mathfrak{F}$ is closed under complements.

Hence, $A_1, \dots, A_n \in \mathfrak{F} \implies \cup_{j=1}^n A_j \in \mathfrak{F}$ and $A \in \mathfrak{F} \implies A^c \in \mathfrak{F}$. Here $A^c := X \setminus A$. In particular, we see that $\emptyset \in \mathfrak{F}$. Also, $\mathfrak{F}$ is closed under finite intersections, due to De Morgan's law $A \cap B = (A^c \cup B^c)^c$.

Example 1.39. Let $X = \{a, b, c\}$ and $\mathfrak{F} = \{\emptyset, \{a\}, \{b\}, \{a, b, c\}\}$. Then $\mathfrak{F}$ is not an algebra.

Definition 1.40. An algebra is called a *σ-algebra* if it is closed under countable unions (implying it's also closed under countable intersections). In this case, we call $(X, \mathfrak{F})$ a *measurable space.* Henceforth $\mathfrak{F}$ always denotes a σ-algebra.

Sets in $\mathfrak{F}$ are said to be *measurable.*

Example 1.41. The set $\mathcal{I}$ of all intervals $I \subset \mathbb{R}$ is certainly not a σ-algebra (not even an algebra; it's a so-called *π-system* if we also consider $\emptyset$). However, it *generates* the Borel σ-algebra $\mathfrak{B}$ of $\mathbb{R}$. This is the smallest σ-algebra containing $\mathcal{I}$. Its elements contain in particular all open sets, closed sets, countable intersections of open sets, countable unions of closed sets... and much more.

More generally, if X is a topological space, we can always endow X with the *Borel σ-algebra* generated by all open sets.

We are now ready to discuss the main entity of this section :

Definition 1.42. Let X be a set with σ-algebra $\mathfrak{F}$. We say that $\mu : \mathfrak{F} \to [0, \infty]$ is a *measure* if

(i) $\mu(\emptyset) = 0$,

(ii) $\mu(\cup_{j=1}^\infty A_j) = \sum_{j=1}^\infty \mu(A_j)$ if $A_j \cap A_k = \emptyset$ for all $j \neq k$.

Property (ii) is called σ-additivity.

Example 1.43. For any set X, we may take $\mathfrak{F} = P(X)$, the power set of X. Now fix $x \in X$ and consider the measure $\mu = \delta_x$, where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise. This is called the *Dirac measure* centered at x.

Example 1.44. For any set X, take $\mathfrak{F} = P(X)$ and let $\mu(A)$ be the number of elements in A (∞ if A is infinite). We call this the *counting measure*.

Example 1.45. Let $X = \mathbb{R}^n$ and take $\mathfrak{F} = \mathfrak{B}(\mathbb{R}^n)$, the σ-algebra of Borel sets in $\mathbb{R}^n$. We take $\mu = \mathcal{L}$ to be the *Lebesgue measure*. Its construction is complicated and will not be discussed. We just mention that it satisfies the following properties :

— the volume of rectangles is as expected : $\mathcal{L}(I_1 \times \cdots \times I_n) = |I_1| \cdots |I_n|$ for any intervals $I_j \subset \mathbb{R}$, where $|[a,b]|$ is the usual length, $b - a$,

— $\mathcal{L}$ is both translation invariant and rotation invariant,

— $\mathcal{L}(A) = 0$ for any countable set. In particular, the "length" of $\mathbb{Q} \subset \mathbb{R}$ is 0,

— the converse is not true : there exists uncountable sets with $\mathcal{L}(A) = 0$.

Definition 1.46. We say a measure μ is *finite* if $\mu(X) < \infty$.

We say that μ is a *probability measure* if $\mu(X) = 1$.

Example 1.47. Let X be finite, $\mathfrak{F} = P(X)$ and μ the normalized counting measure : $\mu(A) = \frac{\#A}{\#X}$, where $\#A$ is the number of elements in A. This is the familiar probability measure from high school, called the uniform measure.

Example 1.48. Let $X = \mathbb{R}$, $\mathfrak{F} = \mathfrak{B}(\mathbb{R})$, $\mathrm{d}\mu = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-c)^2/(2\sigma^2)}\,\mathrm{d}x$. In particular, for intervals $I = [a,b]$, we have $\mu(I) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \mathrm{e}^{-(x-c)^2/(2\sigma^2)}\,\mathrm{d}x$. Then μ is a probability measure on $\mathbb{R}$ called the *Gaussian measure*.

Similarly, we can take $\mathrm{d}\mu = \frac{1}{\pi\gamma} \cdot \frac{\gamma^2}{(x-x_0)^2+\gamma^2}\,\mathrm{d}x$. This probability measure is said to have the *Cauchy distribution*. Both measures are quite important.

Definition 1.49. A property is said to hold μ*-almost everywhere* if it holds outside a set $N \subset X$ of zero measure : $\mu(N) = 0$. We abbreviate this by μ-a.e. If μ is a probability measure, we often say μ*-almost surely*, μ-a.s.

Example 1.50. Let $\mathbf{1}_A$ be the *characteristic function* of A, that is, $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. Then $\mathbf{1}_{\mathbb{Q}} = 0$ $\mathcal{L}$-almost everywhere.

Finally we mention the notion of the "support" of a measure.

If X is a metric space and $f : X \to \mathbb{R}$ is continuous, the support of f is defined by $\operatorname{supp} f = \overline{\{x \in X : f(x) \neq 0\}}$.

We're thus looking where the function "lives", where it is nonzero.

Similarly, we define the support of a measure μ on X by

$$\begin{aligned}\operatorname{supp} \mu &= \{x \in X : \mu(O) > 0 \text{ for any open neighborhood } O \text{ of } x\} \\ &= X \setminus \{x \in X : \text{ there exist an open neighborhood } O_0 \text{ of } x \text{ with } \mu(O_0) = 0\}\end{aligned}$$

Example 1.51. The support of the Lebesgue measure $\mathcal{L}$ on $\mathbb{R}$ is $\mathbb{R}$.

Consider the restricted measure $\mathrm{d}\mu = \mathbf{1}_{[a,b]}\,\mathrm{d}x$, also on $X = \mathbb{R}$. In other words, $\mu(A) = \mathcal{L}(A \cap [a,b])$. Then $\operatorname{supp} \mu = [a,b]$.

Now that we understand the basic properties of a measure μ, we briefly discuss how to integrate with respect to μ.

Recall that in Riemann's integration from first year calculus, one starts by dividing the domain of the function, which a closed interval, into smaller and smaller sub-intervals. In each small interval, we then calculate the areas of two rectangles, one above, one below the curve, then take the sum of these areas and take the limit as the width of the rectangles goes to zero. We say that f is Riemann-integrable if the approximations from above and below tend to the same value.

One trouble with this definition is that it requires functions to be pretty regular. [3] Lebesgue's idea is to do the following instead. On any domain X, $f : X \to \mathbb{R}$, we choose a "subdivision" *depending on the function.* We no longer divide intervals and so on while forgetting about f. Instead, we look at the *range* of f. We divide the domain X into pieces composed of *inverse images*

$$f^{-1}(A) := \{x \in X : f(x) \in A\}\,.$$

The approximation procedure then goes as follows : first consider simple functions of the form

$$s = \sum_{j=1}^{p} \alpha_j \mathbf{1}_{A_j}\,.$$

These functions take only finitely many values $\{\alpha_1, \ldots, \alpha_p\}$, so their graphs look like "steps". This representation is generally not unique, but if we choose $A_j := s^{-1}(\alpha_j)$, then it is unique. As one imagines, we define the integral of this function to be the sum of rectangles. More generally, for any measurable $A \subset X$,

$$\int_A s\,\mathrm{d}\mu := \sum_{j=1}^{p} \alpha_j \mu(A_j \cap A)\,.$$

We then move to more interesting functions, namely non-negative measurable [4] functions $f \geq 0$. We define the integral as

$$\int_A f\,\mathrm{d}\mu := \sup_{s \leq f} \int_A s\,\mathrm{d}\mu\,,$$

where the supremum runs over all simple functions $s \leq f$. This works very well, the integral can be shown to be linear for instance, and so on.

If f is real-valued, we just consider its positive and negative parts. Similarly, if f is complex-valued, we consider its real and imaginary parts.

We omit the details but hope the reader got some idea.

One good thing about Lebesgue's integral is that it behaves very well with respect to limits. We state just one illustration :

Theorem 1.52. *Let μ be a finite measure on X, suppose (f_n) is a sequence of measurable functions on X, $|f_n(x)| \leq M$ for all $n \geq 1$, $x \in X$. Suppose that $f_n(x) \to f(x)$ μ-a.e. Then $\int f_n\,\mathrm{d}\mu \to \int f\,\mathrm{d}\mu$.*

3. The french mathematician Dieudonné expressed a strong opposition to the fact that we "still teach" Riemann's integral. See his *Foundations of Modern Analysis*, Volume 1, Chapter VIII.

4. Imposing that a function is measurable is in general a very weak assumption, no need for any regularity. The definition : you need $f^{-1}(I) \in \mathfrak{F}$ for any interval I.

If μ is a probability measure on X, we will often use the notation

$$\mathbb{E}_\mu(f) := \int_X f \,\mathrm{d}\mu \,,$$

where $\mathbb{E}$ stands for *expectation*.

Let us conclude this section with a brief discussion of L^p spaces. This will provide some link with the previous section. This part is not essential to the project, but should be in the mathematical baggage of any graduate student.

Let $(X, \mathfrak{F}, \mu)$ be a measure space, for simplicity assume $\mu(X) < \infty$ (that's really not necessary). Given $p \geq 1$, we define $\mathcal{L}^p(X, \mu)$ to be the set of all measurable $f : X \to \mathbb{C}$ satisfying $\|f\|_p := (\int_X |f|^p \,\mathrm{d}\mu)^{1/p} < \infty$.

Note that we discussed a quite similar space in Section 1.1. There the functions were continuous and we had a normed space (which wasn't complete). Here functions are only measurable, is it a problem ?

Well yes. Recall that for $\|\cdot\|_p$ to be a norm, we need $\|f\|_p = 0 \implies f = 0$. If the function is not continuous, we may well have $\int |f|^p \,\mathrm{d}\mu = 0$ but f not the zero function (consider a function identically zero except at one point). In general, one can prove that

$$\int |f|^p \,\mathrm{d}\mu = 0 \iff f(x) = 0 \qquad \mu - a.e.$$

This shows that $\|\cdot\|_p$ is not a norm on $\mathcal{L}^p(X, \mu)$.

To solve this problem, we should work with a different space in which a function equal to zero almost everywhere is the zero vector. There are many such functions, they should all be the same vector : zero. The idea of collapsing a set of points (here functions) into a single point is well established in mathematics : we just have to take quotients.

With this intuition, we consider $N(X, \mu) = \{f : f(x) = 0 \ \mu - a.e.\}$, which is a subspace of $\mathcal{L}^p(X, \mu)$, and define the quotient

$$L^p(X, \mu) := \mathcal{L}^p(X, \mu)/N(X, \mu) \,.$$

Strictly speaking, the elements of $L^p(X, \mu)$ are cosets.. this is not a very enlightening point of view. It is better to think of elements of $L^p(X, \mu)$ as ordinary functions, which are well-defined up to sets of measure zero.

The problem now is solved : $L^p(X, \mu)$ becomes a normed space if endowed with $\|\cdot\|_p$. This space is separable (unless X is really bad). Moreover :

Theorem 1.53. *$L^p(X, \mu)$ is a Banach space. $L^2(X, \mu)$ is a Hilbert space.*

We can finally understand why $L^p(X, \mu)$ is the completion of $C(X)$ endowed with $\|\cdot\|_p$, a fact mentioned in Section 1.1. Indeed, $L^p(X, \mu)$ is complete. It is also true that continuous functions of compact support are dense in it, when X is locally compact (proof omitted). This concludes the verification.

We use the shorthand notation $L^p(X)$ when $X \subset \mathbb{R}^n$ is endowed with Lebesgue's measure. People sometimes denote $L^p(\mu)$ when the emphasis is really on the measure (for example, comparing several measures on the same space).

Further reading : My favorite "crash course" in measure theory is Appendix A of the book [33] by Teschl. Keep in mind the content there is very dense; you will need to work hard to fill the details. More comprehensive treatments are [28, 29].

1.3 Graphs

Let us now discuss our main objects of study : graphs.

Definition 1.54. A graph is a pair $G = (V, E)$ consisting of a countable set of *vertices* V and a set of *edges* $E \subseteq \{\{x, y\} : (x, y) \in V^2, x \neq y\}$.

Hence, we just have a set of points called vertices, linked together by bridges that we call edges. In our treatment, we exclude multiple edges between vertices and exclude self-loops. Our graphs are *undirected*, i.e. the edge $\{x, y\}$ can be seen as a double arrow, from x to y and from y to x.

More general graphs can be treated, but this framework is already very rich.

Definition 1.55. If $\{x, y\}$ is an edge, we denote $x \sim y$. This means that y is a *nearest neighbor* of x.

We denote the set of nearest neighbors of x by $\mathcal{N}_x$.

The *degree* of a vertex is the number of its neighbors. We denote $d(x) = |\mathcal{N}_x|$.

Even if the graph is undirected, it is sometimes useful to work with directed bonds :

Definition 1.56. Let $G = (V, E)$ be a graph. If $\{x, y\}$ is an edge, we denote $b = (x, y)$ the *directed edge* with *origin* x and *terminus* y. These are denoted o_b and t_b, respectively.

We let B denote the set of directed edges of G. Since the graph is undirected, each edge induces two directed edges, so that $|B| = 2\,|E|$.

Example 1.57. Draw the following graphs :

— A *complete graph* on 3 vertices. A complete graph is a graph in which $\mathcal{N}_x = V \setminus \{x\}$ for any x, in other words, there is an edge between x and any $y \neq x$.

— A *star graph* on 6 vertices. The situation here is quite the opposite. There is a central vertex which is connected to a vertices. Any other vertex is connected only to the central vertex, so we get a star shape.

— A *d-regular graph* is a graph in which each vertex has exactly d-neighbors. Draw a 3-regular graph on 4 vertices.

Definition 1.58. A *path* in a graph G is a sequence of vertices $(x_0, x_1, \dots)$ which may be finite or infinite, and which satisfies $x_j \sim x_{j+1}$ and $x_{j+1} \neq x_{j-1}$. Sometimes we emphasize this and say *non-backtracking path*.

A *walk* in G is similar but can backtrack. Namely, it is a sequence $(x_0, x_1, \dots, x_n)$, usually finite, which is only required to satisfy $x_{j+1} \sim x_j$.

A *cycle* is a closed path $(x_0, x_1, \dots, x_n, x_0)$.

Example 1.59. A *tree* is a graph which has no cycles. Trees are important types of graphs which appear in a wide variety of problems. In particular, star graphs are trees.

Definition 1.60. A graph is said to be *connected* if for any $x, y \in V$, there is a path $(x_0, \dots, x_k)$ in G such that $x_0 = x$ and $x_k = y$.

Example 1.61. Draw a connected graph. Draw a disconnected graph.

Definition 1.62. A graph is said to be *finite* if it has finitely many vertices : $|V| < \infty$.

We always assume the graph is *locally finite*, which means that the degree $d(x) < \infty$ for any vertex x.

Example 1.63. Convince yourself that a 3-regular tree must be infinite.

Remark 1.64. If G is a finite d-regular graph, then $|B| = d\,|V|$. In fact,

$$|B| = \sum_{(x,y)\in B} 1 = \sum_{x\in V}\sum_{y\sim x} 1 = \sum_{x\in V} d(x) = d\,|V|\,.$$

Recall the bounded linear operators discussed in Section 1.1. Graphs come with the following natural operator :

Definition 1.65. Let $\ell^2(G) = \{f : V \to \mathbb{C} : \sum_{v\in V} |f(v)|^2 < \infty\}$. Note that $\ell^2(G) = \mathbb{C}^V$ if the graph is finite (the set of all functions from $V \to \mathbb{C}$).

Let $G = (V, E)$ be a graph. The *adjacency matrix* of G, denoted $\mathcal{A}_G$, is the operator $\mathcal{A}_G : \ell^2(G) \to \ell^2(G)$ defined by

$$(\mathcal{A}_G f)(x) = \sum_{y\sim x} f(y)\,.$$

In other words, the value of $\mathcal{A}_G f$ at x is the sum of $f(y)$ over all nearest neighbors.

Since our graphs are undirected, we get an important property :

Lemma 1.66. *Suppose that* $\sup_{x\in V} d(x) \le D < \infty$.

Then $\mathcal{A}_G$ is a bounded linear operator with $\|\mathcal{A}_G\| \le D$. Moreover, $\mathcal{A}_G$ is self-adjoint.

Proof. This is easy. First recall the *Cauchy-Schwarz inequality :*

$$\sum_{j=1}^{n} |x_j y_j| \le \left(\sum_{j=1}^{n} |x_j|^2\right)^{1/2} \left(\sum_{j=1}^{n} |y_j|^2\right)^{1/2}.$$

Now let $f \in \ell^2(G)$. Then

$$\|\mathcal{A}_G f\|^2 = \sum_{x\in V} |(\mathcal{A}_G f)(x)|^2 = \sum_{x\in V} \left|\sum_{y\sim x} f(y)\right|^2.$$

By Cauchy-Schwarz,

$$\begin{aligned}\left|\sum_{y\sim x} f(y)\right| &\le \sum |f(y)| = \sum_{y\sim x} (1\cdot |f(y)|)\\ &\le \left(\sum_{y\sim x} 1^2\right)^{1/2} \left(\sum_{y\sim x} |f(y)|^2\right)^{1/2} = \sqrt{d(x)} \left(\sum_{y\sim x} |f(y)|^2\right)^{1/2}.\end{aligned}$$

Hence,

$$\begin{aligned}\|\mathcal{A}_G f\|^2 &\le D \sum_{x\in V}\sum_{y\sim x} |f(y)|^2 = D \sum_{(x,y)\in B} |f(y)|^2 = D \sum_{(y,x)\in B} |f(y)|^2\\ &= D \sum_{y\in V}\sum_{x\sim y} |f(y)|^2 \le D^2 \sum_{x\in V} |f(y)|^2 = D^2 \|f\|^2\,.\end{aligned}$$

Since f is arbitrary, this gives $\|\mathcal{A}_G\| \le D$.

For self-adjointness, note that

$$\begin{aligned}\langle \mathcal{A}_G f, g\rangle &= \sum_{x\in V} (\mathcal{A}_G f)(x)\overline{g(x)} = \sum_{x\in V}\sum_{y\sim x} f(y)\overline{g(x)}\\ &= \sum_{(x,y)\in B} f(y)\overline{g(x)} = \sum_{y\in V}\sum_{x\sim y} f(y)\overline{g(x)} = \sum_{y\in V} f(y)\overline{(\mathcal{A}_G g)(y)} = \langle f, \mathcal{A}_G g\rangle\end{aligned}$$

as required. □

It is also useful to consider more general operators. In mathematical physics, the adjacency matrix is regarded as a discrete analog of the Laplacian $-\Delta$ on continuous functions. So what are Schrödinger operators ?

Here it's simple. Endow each vertex x with some weight $W(x) \in \mathbb{R}$. We then define

$$H_G = \mathcal{A}_G + W$$

to be the operator $(H_G f)(x) = (\mathcal{A}_G f)(x) + W(x)f(x)$. In other words, W is a multiplication operator. In this context, it is called a *potential* and H_G is called a *Schrödinger operator*. It is more common to denote potentials by V, but we chose W to avoid confusion with vertices.

Lemma 1.67. *If* $\sup_{x\in V} d(x) \le D < \infty$ *and* $\|W\|_\infty := \sup_{x\in V} |W(x)| < \infty$*, then* H_G *is a bounded self-adjoint operator with* $\|H_G\| \le D + \|W\|_\infty$*.*

The proof is left for the reader. Note that self-adjointness holds because W is *real-valued.*

1.4 The spectrum of an operator

1.4.1 Matrices

Consider the Hilbert space $\mathcal{H} = \mathbb{C}^n$ and a square matrix $A : \mathcal{H} \to \mathcal{H}$. Recall that $0 \ne u \in \mathcal{H}$ is an *eigenvector* of A with corresponding eigenvalue $\lambda \in \mathbb{C}$ if $Au = \lambda u$. The matrix $A = (a_{i,j})_{i,j=1}^n$ is said to be *Hermitian* if $a_{i,j} = \overline{a_{j,i}}$ for all $i, j = 1, \dots, n$. In the language of Section 1.1, this means that A is self-adjoint. A classical result in linear algebra says that any Hermitian matrix is diagonalizable. More precisely, we may choose an orthonormal basis (φ_j) for $\mathcal{H}$ of eigenvectors of A, and the corresponding eigenvalues are real. The matrix $(\langle A\varphi_j, \varphi_i\rangle)_{i,j=1}^n$ will then be diagonal (with the eigenvalues on the diagonal). This result is sometimes expressed by saying that there exists a unitary matrix U and a diagonal matrix D such that $U^{-1}AU = D$; the unitary matrix is simply the matrix that transforms the original basis into the basis (φ_j) above.

Let us take a closer look. Let $\lambda_1 < \lambda_2 < \ldots < \lambda_m$, $(m \le n)$ be the eigenvalues of A. Let $\varphi_1(\lambda_k), \dots, \varphi_{N(\lambda_k)}(\lambda_k)$ be the eigenvectors corresponding to λ_k (one vector if the eigenvalue is simple). Let $\mathcal{H}_k = \operatorname{span}\{\varphi_j(\lambda_k)\}_j$. This "spectral theorem" then tells us that the eigenvectors $\{\varphi_j(\lambda_k)\}_{j,k}$ form an orthonormal basis of $\mathcal{H}$. So any $f \in \mathcal{H}$ has an expansion

$$f = \sum_{k=1}^{m} \sum_{j=1}^{N(\lambda_k)} \langle f, \varphi_j(\lambda_k)\rangle \varphi_j(\lambda_k) \tag{4-1}$$

Since $A\varphi_j(\lambda_k) = \lambda_k \varphi_j(\lambda_k)$, this expansion also gives

$$Af = \sum_{k=1}^{m} \sum_{j=1}^{N(\lambda_k)} \lambda_k \langle f, \varphi_j(\lambda_k)\rangle \varphi_j(\lambda_k) \tag{4-2}$$

Let $P(\lambda_k)$ be the orthogonal projection onto $\mathcal{H}_k$. Then

$$P(\lambda_k) f = \sum_{j=1}^{N(\lambda_k)} \langle f, \varphi_j(\lambda_k)\rangle \varphi_j(\lambda_k)$$

Thus, (4-1) and (4-2) take the form

$$\mathbb{1} = \sum_{k=1}^{m} P(\lambda_k), \qquad A = \sum_{k=1}^{m} \lambda_k P(\lambda_k)$$

If we introduce the "projection-valued measure" $E(J) := \sum_{\lambda_j \in J} P(\lambda_j)$ for any Borel set $J \subset \mathbb{R}$, this becomes

$$\mathbb{1} = \int_{\mathbb{R}} dE(\lambda), \qquad A = \int_{\mathbb{R}} \lambda dE(\lambda)\,. \tag{4-3}$$

The aim of this section is to see how far we can extend these expressions to general self-adjoint operators on arbitrary Hilbert spaces. We'll see that (4-3) generalizes well - this is the famous *spectral theorem*.

1.4.2 Spectra

Throughout the course, all Hilbert/Banach spaces will be over $\mathbb{C}$.

We know that any matrix has a set of eigenvalues; these are the zeros of the characteristic polynomial. Since the previous expansions were indexed by eigenvalues, if we want to generalize the above formulas, it seems natural to first check if any self-adjoint operator has eigenvalues.

The simplest matrices are the diagonal matrices, which act as $(Dv)_j = D_j v_j$, where D_j are the diagonal entries. The natural analog for the Hilbert space $\mathcal{H} = L^2[a,b]$ is a multiplication operator, acting as $(M_\varphi f)(x) = \varphi(x)f(x)$.

So let us study the very simple operator $(Af)(x) = xf(x)$. It is easily checked that A is a bounded self-adjoint operator. If A had an eigenvalue $\lambda \in \mathbb{C}$ corresponding to some eigenvector $0 \neq \varphi \in L^2[a,b]$, we would have

$$(x - \lambda)\varphi(x) = 0$$

almost everywhere. This implies $\varphi(x) = 0$ a.e., so $\varphi = 0$ in L^2, a contradiction. Hence, A has no eigenvalues.

We should thus work with something weaker than eigenvalues. Let E be a Banach space and T a bounded operator. If $\lambda \in \mathbb{C}$ is an eigenvalue, then there is a nonzero u such that $(T - \lambda)u = 0$ (we always denote $T - \lambda := T - \lambda I$); in other words, $u \in \ker(T - \lambda)$ and $T - \lambda$ is not injective. But we know that in finite dimension, $T - \lambda$ is not injective iff it is not bijective. So instead of looking at eigenvalues, we could look at the λ such that $T - \lambda$ is not bijective. But by open mapping theorem, $T - \lambda$ is bijective iff $(T - \lambda)^{-1}$ is a bounded operator. Thus, if

$$\rho(T) = \{\lambda \in \mathbb{C} : (T - \lambda)^{-1} \in B(E)\}\,,$$

where $B(E)$ is the set of bounded operators on E, then we may consider

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

This is what is called the *spectrum* of T. The preceding discussion says that $\sigma(T)$ coincides with the set of eigenvalues of T in the special case where T is a matrix on $\mathbb{C}^n$.

Let us turn back to our example and see if the operator A has a spectrum. We already showed that $A - \lambda$ is injective for any $\lambda \in \mathbb{C}$. Now let $\lambda \notin [a,b]$. Given $\varphi \in L^2[a,b]$, let $\psi(x) = \frac{\varphi(x)}{x-\lambda}$. Then

$$\int_{[a,b]} |\psi(x)|^2 \mathrm{d}x = \int_{[a,b]} \Big|\frac{\varphi(x)}{x - \lambda}\Big|^2 \mathrm{d}x \leq d^{-2}\|\varphi\|^2$$

where $d > 0$ is the distance from λ to $[a, b]$. Hence, $\psi \in L^2[a, b]$ and $(A - \lambda)\psi = \varphi$. The operator $A - \lambda$ is thus surjective. As we already know that $A - \lambda$ is injective, it follows that $\lambda \notin \sigma(A)$.

Now suppose $\lambda \in [a, b]$ and choose $\varphi(x) \equiv 1$. Since the functions $(x - \lambda)^{-1} \notin L^2[a, b]$, we get that $A - \lambda$ is not surjective. Hence, $[a, b] \subset \sigma(A)$. Conclusion: $\sigma(A) = [a, b]$.

More generally, the spectrum of a multiplication operator by a measurable function $g(x)$ is precisely the "essential range" of g (here the essential range of $g(x) = x$ is just the range, i.e. $[a, b]$).

The previous fact is quite true for all bounded operators :

Theorem 1.68. *The spectrum of a bounded operator on a Banach space is never empty.*

The proof uses Liouville's theorem (a bounded entire function must be constant). We also know that

Lemma 1.69. *If T is bounded, then $\sigma(T)$ is compact in $\mathbb{C}$ and $\sigma(T) \subset \overline{D(0, \|T\|)}$, the closed disc of radius $\|T\|$ around 0.*

If T is self-adjoint, then the spectrum is real and $\sigma(T) \subseteq [-\|T\|, \|T\|]$.

We remark that contrary to the case of matrices whose spectrum is just a set of points in $\mathbb{C}$, the operator above has a spectrum that looks continuous. In general, there is a fine decomposition of the spectrum into "pure point", "absolutely continuous" and "singularly continuous" parts, which we may discuss later in the course.

We may ask how far a spectral element differs from an eigenvalue. Let T be bounded. We say that (φ_n) is a *Weyl sequence* associated to $\lambda \in \mathbb{C}$, if $\|\varphi_n\| = 1$ for all n and $\|(T - \lambda)\varphi_n\| \to 0$.

Theorem 1.70. *Let T be bounded.*

— *If there exists a Weyl sequence associated to λ, then $\lambda \in \sigma(T)$,*

— *if T is self-adjoint, then $\lambda \in \sigma(T)$ iff there exists a Weyl sequence associated to λ.*

1.4.3 The spectral theorem

In [22], the spectral theorem is said to be "one of the most important mathematical achievements of all times". We already saw in Section 1.4.1 what it's about; let us give the precise statement.

Theorem 1.71. *We may associate to any self-adjoint operator A on a separable Hilbert space $\mathcal{H}$ a unique projection-valued measure P_A on $\mathbb{R}$ such that*

$$A = \int_{\mathbb{R}} \lambda \mathrm{d}P_A(\lambda)$$

More generally, for any Borel function $f : \sigma(A) \to \mathbb{C}$, we may define an operator

$$f(A) = \int_{\sigma(A)} f(\lambda) \mathrm{d}P_A(\lambda)\,.$$

This operator is bounded if f is bounded. Otherwise, it is well-defined on a proper subspace.

In contrast to Section 1.4.1, the measure P_A is no longer discrete in general. It is supported on the spectrum of A, so it suffices to integrate over $\sigma(A)$ above.

Still, it is useful to work with ordinary scalar measures. Given $\phi, \psi \in \mathcal{H}$, we call

$$\mu_{\phi,\psi}(J) := \langle \chi_J(A)\phi, \psi\rangle\,, \qquad \text{for Borel } J \subseteq \mathbb{R}$$

the *spectral measure of* A at the points ϕ, ψ. In general this is complex-valued, but $\mu_{\phi,\phi}$ is a good non-negative measure as in Section 1.2. Note that $\mu_{\phi,\phi}(\mathbb{R}) = \|\phi\|^2$.

The following theorem gives the properties of $f(A)$. For simplicity, we assume everything is bounded.

Theorem 1.72 (Functional Calculus)**.** *Let $\mathcal{H}$ be a Hilbert space, $T \in B(\mathcal{H})$ be self-adjoint and let $\mathscr{B}(\sigma(T))$ be the set of bounded Borel functions on $\sigma(T)$. Then*

(i) *The map $f \mapsto f(T)$ is an algebraic $*$-homomorphism from $\mathscr{B}(\sigma(T)) \to B(\mathcal{H})$. That is, $(fg)(T) = f(T)g(T)$, $(\alpha f)(T) = \alpha f(T)$, $\mathbf{1}(T) = Id$ and $\overline{f}(T) = f(T)^*$.*

(ii) $\|f(A)\| \leq \|f\|_\infty$.

(iii) *If $f(t) = t$, then $f(T) = T$.*

(iv) *If $f_n, f \in \mathscr{B}(\sigma(T))$ satisfy* $\sup_n \|f_n\|_\infty < \infty$, *and $f_n \to f$ pointwise, then $f_n(T) \to f(T)$ strongly in $B(\mathcal{H})$.*

(v) *If $Tx = \lambda x$, then $f(T)x = f(\lambda)x$.*

(vi) *If $f \geq 0$, then $f(T) \geq 0$.*

(vii) *If $TS = ST$, then $f(T)S = Sf(T)$.*

(viii) (Spectral mapping theorem). *If f is bounded continuous, then $\sigma(f(A)) = f(\sigma(A))$.*

To conclude, recall that a square matrix A is diagonalizable iff there exist a unitary matrix U such that $U^{-1}AU = D$, for some diagonal matrix D. It would be interesting to obtain a similar version for self-adjoint operators in general. As previously noted, the natural analog of diagonal matrices are multiplication operators. The following theorem holds.

Theorem 1.73. *If A is self-adjoint on a separable Hilbert space $\mathcal{H}$, then there exists a measure space (X, μ), a measurable function $\varphi : X \to \mathbb{R}$ and a unitary operator $U : L^2(X, \mu) \to \mathcal{H}$ such that*

$$(U^{-1}AUf)(x) = \varphi(x)f(x)$$

for any $f \in L^2(\mu)$.

Thus, any self-adjoint operator is unitarily equivalent to a multiplication operator. However, the above representation is not canonical at all. Indeed, we may in general represent A in many different $L^2(X, \mu)$ spaces. However, we can strengthen this theorem to obtain an *ordered representation*, see e.g. [22], which is more complicated (to write and to prove), but has the advantage of giving a complete classification of self-adjoint operators, up to unitary equivalence.

Further reading : The reader can find an outline of the proof of the spectral theorem in Halmos' paper [21]. Everything we need appears in detail in [30, Chapter 4]. In particular, *compact* self-adjoint operators are shown to behave like matrices, so that the spectrum reduces to (the closure of) the set of eigenvalues in this special case. A similarly simple situation holds if the operator has a compact resolvent. We finally mention the important classical references [27, 24].

The previous results generalize in many directions. In particular, everything works with unbounded operators. The spectral theorem also extends to normal operators.

1.5 Some basics of probability theory

1.5.1 Basics

Definition 1.74. A *probability space* is a triple $(\Omega, \mathfrak{F}, \mathbb{P})$, where Ω is a set, $\mathfrak{F}$ a σ-algebra of subsets of Ω and $\mathbb{P}$ a probability measure on Ω.

The set Ω represents the set of *sample points*; we denote $\omega \in \Omega$. The σ-algebra represents the set of *events*, and we would like to measure their probability.

The events we observe are usually of the form "$\{X \in J\}$" :

Example 1.75. Consider the question : what is the probability that the mean temperature of Cairo in August will be between 20 and 25 degrees Celcius ?

To model this vague question, we may base our answer on the meteorological data from the years 1991 to 2020 for example. For each year, a mean temperature is recorded. So we may take the sample $\Omega = \{1991, 1992, \dots, 2020\}$, consider the mean temperature as a map $T : \Omega \to \mathbb{R}$, and one possible answer is that such a probability is unbiased, i.e.

$$\mathbb{P}(T \in [20,25]) = \frac{\#\{\text{years in } \Omega : T(year) \in [20,25]\}}{30}.$$

The previous question is quite typical of probability : the most important quantity was the temperature, based on this we started looking for appropriate sample spaces. As the temperature fluctuates from year to year, we call it a "random variable".

Definition 1.76. A *random variable* is a measurable map $X : \Omega \to \mathbb{R}$.

Recall that the map is measurable if $X^{-1}(J) := \{\omega : X(\omega) \in J\} = \{X \in J\} \in \mathfrak{F}$ for any Borel $J \subset \mathbb{R}$, i.e. we need these to be legitimate events to measure.

Similarly, we can consider more general random objects :

Definition 1.77. A *random vector* is a measurable map $X : \Omega \to \mathbb{R}^k$.

A *random $n \times n$ matrix* is a measurable map $X : \Omega \to \mathcal{M}_{n,n}(\mathbb{R})$

We can also consider random self-adjoint operators and so on.

Definition 1.78. The *mean* or *expectation* of a random variable X is the quantity $\mathbb{E}(X) = \int_\Omega X(\omega)\,\mathrm{d}\,\mathbb{P}(\omega)$. We say that X is *centered* if $\mathbb{E}(X) = 0$.

Random variables can be used to *pushforward* the measure $\mathbb{P}$ to $\mathbb{R}$:

Definition 1.79. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable. We denote $\mathbb{P}_X := X_\star\,\mathbb{P}$ the probability measure on $\mathbb{R}$ defined by

$$\mathbb{P}_X(J) := \mathbb{P}(X^{-1}(J)) = \mathbb{P}(X \in J)$$

for Borel $J \subseteq \mathbb{R}$. We call $\mathbb{P}_X$ the *distribution* of X.

Similarly, the distribution of a random vector $X : \Omega \to \mathbb{R}^k$ is the measure on $\mathbb{R}^k$ defined by $\mathbb{P}_X(D) = \mathbb{P}(X \in D)$.

We next discuss independence.

Definition 1.80. Two events $A, B \in \mathfrak{F}$ are said to be *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

The reason for considering products is intuitive :

Example 1.81. Suppose you have to find the (unbiased) probability of getting the string 11 out of the sample $\{00, 01, 10, 11\}$. By considering strings as a whole, we find the probability is $\frac{1}{|\Omega|} = \frac{1}{4}$. But we could also argue as follows : the first digit can be either 0 or 1, so the probability that the first entry is 1 is $\frac{1}{2}$. Similarly for the second digit. The probability to get 11 is thus $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. This layman method of thinking implicitly uses that the values of the first and second digits are independent of each other (in the sense of everyday language). [5]

We now give a more general definition which includes events as a special case.

Definition 1.82. Let $(X_j)_{j=1}^n$ be random variables defined on $(\Omega, \mathfrak{F}, \mathbb{P})$. We say that (X_j) are *independent* if the distribution of $X = (X_1, \ldots, X_n)$ on $\mathbb{R}^n$ is a product measure $Q_1 \times \cdots \times Q_n$, with Q_j a probability measure on $\mathbb{R}$.

Events A_j are then independent if the indicator random variables $\mathbf{1}_{A_j}$ are independent.

If $Q_1 = \cdots = Q_n$, the random variables are said to be *identically distributed*.

We use the standard abbreviation "i.i.d." for "independent, identically distributed".

In particular, we require that

$$\mathbb{P}\left((X_1, \ldots, X_n) \in A_1 \times \cdots \times A_n\right) = Q_1(A_1) \times \cdots \times Q_n(A_n)$$

for any rectangle $A_1 \times \cdots \times A_n$. Convince yourself that $\mathbb{P}(X_j \in A_j) = Q_j(A_j)$. Then it follows that $\mathbb{P}(X_j \in A_j \text{ for all } 1 \leq j \leq n) = \prod_{j=1}^n \mathbb{P}(X_j \in A_j)$.

One fundamental property is the following :

Theorem 1.83. *If $X_1, \ldots, X_n$ are independent random variables on Ω with $\mathbb{E}\,|X_j| < \infty$, then $\mathbb{E}\,|X_1 \cdots X_n| < \infty$ and*

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n)\,.$$

Remark 1.84. As previously mentioned, the main interest is generally in random variables; this is usually the quantity given at start. Suppose we know that X_j takes values in $[a_j, b_j]$ and we want the random variables to be independent, with distributions Q_j. Then we simply *choose* the sample space $\Omega = [a_1, b_1] \times \cdots \times [a_n, b_n]$, so that $(X_1, \ldots, X_n)$ describe the coordinates of Ω, and we endow Ω with $\mathbb{P} := Q_1 \times \cdots \times Q_n$. If we want them to be i.i.d., we choose $\mathbb{P} = Q_1 \cdots Q_1$.

We often need to deal with infinitely many random variables.

Definition 1.85. Let $(X_t)_{t \in \Lambda}$ be a possibly infinite family of random variables. We say the family is independent if every finite subfamily is independent.

1.5.2 Three fundamental results

With these basic definitions, we can give three important results in probability. Each result gives a kind of information when a series of experiments is performed a very large number of times (infinitely many times in math language).

5. In general, it is useful to imagine that independent events induce a product structure on the sample space, such that the first event is described by the first variable and the second event by the second variable. The intersection of events thus draws a rectangle in sample space; its measure (its probability) is the product of length and width, hence the definition. This interpretation essentially means that independent events imply the existence of "independent vectors" in sample space. This should not be taken too literally; for example Ω is not necessarily a vector space.

The first result is the Borel-Cantelli lemma. The object here is a sequence of events A_n, and we try to answer a basic question : what is the probability that A_n will occur infinitely often if we perform the experiment infinitely many times. To formalize this, we define the event

$$\{A_n \text{ i.o } \} = \cap_{n\geq 1} \cup_{m\geq n} A_m \,.$$

Here i.o stands for "infinitely often". Indeed, $\omega \in \Omega$ belongs to the right-hand-side iff for any n, we may find $m \geq n$ such that $\omega \in A_m$. This means indeed that ω belongs to infinitely many sets $A_{m_1}, A_{m_2}, \ldots$ for some $m_j \geq j$.

We may now state the lemma :

Lemma 1.86 (Borel-Cantelli)**.** *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and $A_n \in \mathfrak{F}$ for $n \in \mathbb{N}$.*

(i) *If $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n \text{ i.o }) = 0$.*

(ii) *If the $\{A_n\}_{n\geq 1}$ are independent and $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A_n \text{ i.o }) = 1$.*

This says that the occurrence of the events infinitely often depends on how big the probability of the events is for very large n. If this probability is so small that the series converges, then almost surely, after some trial n_0, the even will never occur. Conversely, if the series diverges and the events are independent, then almost surely, the event occurs infinitely often.

There is a humorous illustration of the second point, known as the *infinite monkey theorem*. It states that if a monkey hits the keys of a typewriter at random infinitely many times, then almost surely, he will type the complete works of Shakespeare.

To see this, let the sample space be the set of all infinite strings *asbdsaebw*... the monkey may type. The monkey types each letter uniformly at random among the $n = 26$ letters of the alphabet. Let S be the string of length m consisting of the complete works of Shakespeare. Let A_k be the event that "the m-substring starting at position k is S". Then A_{mj+1} are independent for $j = 0, 1, \ldots$ and $\mathbb{P}(A_k) = (\frac{1}{n})^m$. Hence, $\sum_{j=1}^{\infty} \mathbb{P}(A_{mj+1}) = \sum_{j=1}^{\infty} (\frac{1}{n})^m = \infty$.

By Borel-Cantelli, we deduce that the complete works of Shakespeare will actually appear infinitely often in the monkey's output !

Let us move to the second fundamental result, known as the *law of large numbers*. This result says that if we perform an experiment a large number of times, then the average result will converge to the expectation $\mathbb{E}(X)$, which is an average over the *sample space.* [6]

Theorem 1.87 (Law of large numbers)**.** *Let $\{X_n\}_{n\geq 1}$ be an i.i.d. sequence of integrable random variables defined on some $(\Omega, \mathfrak{F}, \mathbb{P})$. Then almost surely,*

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mathbb{E}(X_1)\,.$$

The idea of approximating the average of a function (here the average over sample space in the RHS) by an average over a large number of empirical observations (as in the LHS) is ubiquitous in probability theory and its interactions with mathematical analysis. Sometimes the empirical observations are over "continuous paths" instead of discrete times X_1,X_2,...etc.

6. Apparently, the poor name "law of large numbers" was coined by the French mathematician Poisson. "Large numbers" simply refers to the fact that we won't observe such convergence unless we perform a large number of trials. In this sense, any limit is a law of large numbers... something like "the fundamental theorem of probability" seems more appropriate to the humble author of these notes.

As an illustration of the theorem in (occidental) life, consider a casino offering some game of chance like the roulette. Although the casino may lose money after one or few plays, its earnings on average will tend towards a predictable amount (the expected value), which is at the casino's advantage.

There are several proofs for the law of large number; each has its advantage when it comes to generalizations. In general, the "almost sure" property comes from applying Borel-Cantelli, first statement.

The last theorem of this section is a fundamental result known as the *central limit theorem*. To discuss it rigorously, and also for our future project, let us briefly digress on the topic of *weak convergence* of probability measures.

Definition 1.88. Let E be a metric space and $\mu, \mu_1, \ldots,$ a sequence of probability measures on E. We say that (μ_n) *converges weakly* to μ and denote $\mu_n \xrightarrow{w} \mu$, if

$$\int f \, \mathrm{d}\mu_n \xrightarrow{n\to\infty} \int f \, \mathrm{d}\mu \qquad \text{for any } f \in C_b(E)\,, \tag{5-1}$$

where $C_b(E)$ is the set of bounded continuous functions on E.

It is good to have equivalent criteria for weak convergence. Many are usually collected as a *portemanteau theorem.*[7] One of them goes as follows :

Proposition 1.1. *Let E be a metric space and $\mu, \mu_1, \mu_2, \ldots$ probability measures on E. Then $\mu_n \xrightarrow{w} \mu \iff \mu_n(A) \to \mu(A)$ for any measurable A satisfying $\mu(\partial A) = 0$.*

Here ∂A is the boundary of A. For example, $\partial[a, b] = \{a, b\}$.

In practice, it suffices to check (5-1) for some sub-algebra of $C_b(E)$. For example, when E is compact, it is often doable to establish (5-1) for polynomial functions f, and this turns out to be sufficient. This is sometimes called the *method of moments*, in reference to the fact that $\int t^k \, \mathrm{d}\mu(t) = \mathbb{E}_\mu(t^k)$ is the k-th moment.

Let us state one important result before moving on :

Theorem 1.89. *If E is a compact mectric space, then the set $\mathcal{P}(E)$ of probability measures on E is compact for weak convergence. That is, any sequence of probability measures on E has a weakly convergent subsequence.*

This theorem follows from the *Banach-Alaoglu theorem* in functional analysis. It is also a special case of the quite tough *Prokhorov theorem* in probability.

We finally turn back to the central limit theorem. Recall that we showed in the law of large numbers that if (X_j) are i.i.d. random variables, then $\frac{1}{n}\sum_{k=1}^n X_j \to a$ almost surely, where $a := \mathbb{E}(X_1)$. It follows that $\frac{1}{n}\sum_{k=1}^n (X_j - a) \to 0$, so that the sum $\sum_{k=1}^n (X_j - a) = o(n)$. This notation means that $\sum_{k=1}^n (X_j - a)$ is *negligible* in front of n. So what is the correct order of magnitude of this sum ? That is, can we find some power n^α with $\alpha < 1$ such that the limit $\frac{1}{n^\alpha}\sum_{k=1}^n (X_j - a) \to M \neq 0$ almost surely ?[8]

It turns out that the correct order of magnitude is $n^{1/2}$. However, unfortunately the convergence is not almost sure (and not even "in probability", a weaker concept). Instead, the convergence is in distribution. Before stating the result, we recall the (probably familiar) :

7. In one of his books, Billingsley (jokingly) attributed the theorem to some mathematician called "Jean-Pierre Portmanteau" from some imaginary university. To this day, confusion persists. Actually portemanteau just means "coat hanger", as you can choose "any coat" and you'll get weak convergence.

8. The same question can be rephrased as "what is the speed of convergence in the law of large numbers ?", in this case we search for n^β such that $n^\beta(\frac{1}{n}\sum_{k=1}^n (X_j - a)) \to M \neq 0$.

Definition 1.90. The *variance* of a random variable X with expectation $a := \mathbb{E}(X)$ is the quantity $\mathrm{Var}(X) := \mathbb{E}[(X-a)^2]$.

Definition 1.91. We say that a random variable Y has a *normal distribution* if $\mathbb{P}_Y$ is a Gaussian probability measure, that is, if $\mathrm{d}\,\mathbb{P}_Y = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-a)^2/(2\sigma^2)}\,\mathrm{d}x$ for some $a \in \mathbb{R}$ and $\sigma^2 > 0$. This is usually denoted $N(a, \sigma^2)$. We say that Y has a *standard normal distribution* if $a = 0$ and $\sigma = 1$.

Theorem 1.92 (Central limit theorem)**.** *Let X_j be i.i.d. random variables with $a := \mathbb{E}(X_1)$ and $\sigma^2 := \mathrm{Var}(X_1)$. Define the random variable $Y_n := \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - a)$. Then there is a random variable Y having a standard normal distribution such that $\mathbb{P}_{Y_n} \xrightarrow{w} \mathbb{P}_Y$.*

Equivalently, the theorem says that for any $t \in \mathbb{R}$, $\mathbb{P}(Y_n \leq t) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \mathrm{e}^{-x^2/2}\mathrm{d}x$.

This theorem is often summarized as “the Y_n converge in distribution to $N(0, 1)$”.

This result answers our question of estimating the order of magnitude of $\sum_{k=1}^n (X_k - a)$, which turns out to be $\sqrt{n}$, since the quotient converges (in some sense) to some non-trivial, finite limit. Equivalently, the speed of convergence is $n^{-1/2}$.

We will not need this theorem in our work. However, besides its importance in probability, we stated it here as an illustration of “universality”. Indeed, the theorem says much more than the fact that we have a nontrivial limit. It says that no matter how the X_j are distributed, the random variables Y_n will always converge to $N(0, 1)$, so that the normal distribution can be regarded as a “universal limit”. The quest to prove some forms of “universal laws” for certain classes of random matrices has resulted in an important body of research from renowned mathematicians during the past few decades; see Section 2.1.

Further reading : I recommend [23] for a very pleasant non-technical reading. I personally learned from [14], and often use the encyclopedic [25] in my work, when looking for specific results.

Chapter 2

Benjamini-Schramm convergence

In this chapter we discuss the basic properties of the Benjamini-Schramm topology, carry out some examples, and prove that convergence in the sense of Benjamini-Schramm implies convergence of the associated empirical measures.

What Benjamini and Schramm introduced in [13] is a notion of convergence for a sequence of finite graphs (G_n). An important aspect is that if this sequence has a uniformly bounded degree, then there always exists a limit object up to passing to a subsequence. Moreover, it gives a "correct" notion of convergence from the point of view of spectral theory. For example it has been known at least since [2] that if one is interested in the spectral properties of the $(q+1)$-regular tree $\mathcal{T}_q$, then studying what happens on a sequence of growing balls around some fixed origin is not a good idea, because the spectral behavior of the limit object is completely different from those of $\mathcal{T}_q$. And indeed, such tree balls do not Benjamini-Schramm converge to $\mathcal{T}_q$. On the other hand, a sequence of $(q+1)$-regular graphs with few short cycles do converge to $\mathcal{T}_q$ in the sense of Benjamini-Schramm, and it is known that the mean spectral measures of the corresponding adjacency matrices converge to the density of states of the adjacency matrix on $\mathcal{T}_q$. This is known as the law of *Kesten-McKay*, and it turns out to be a general phenomenon.

As previously mentioned, an important advantage of this notion of convergence is the existence of a limit object. A very broad question is whether specific information about the limit object implies something on the convergent sequence when n gets large. It is one such question that is considered in [7], where we show that if (G_n, W_n) is a sequence of colored graphs of uniformly bounded degree and coloring, and if (G_n, W_n) has a local weak limit ρ supported on colored trees, then roughly speaking, AC spectrum of the limit "Schrödinger" operator implies quantum ergodicity for the sequence. A lot of different spectral questions have also been studied in the literature, and relations have been found with conjectures in group theory [3].

2.1 Prelude on random matrices

Though our main concern in these notes will be sequences of *deterministic* graphs, it is useful to briefly review some results in the spectral theory of random matrices, to put things into context. [1]

The study of random matrices did not start for abstract reasons, for example generalizing results known for $X : \Omega \to \mathbb{R}$ to $X : \Omega \to \mathcal{M}_{n,n}(\mathbb{R})$. Instead, and more interestingly, it started due its natural emergence in applications, first to try to make sense of complicated

1. This section is merely an optional survey, no worries if the student doesn't understand everything.

data analysis (Wishart, 1920s) and later as statistical models for heavy nuclei in physics (Wigner, 1950s). While the early works and the more complete results were usually done under the assumption that the matrix is Hermitian, many powerful results have recently appeared in the literature for non-Hermitian matrices, which we'll mention later.

The first family of results we discuss concerns *Wigner matrices.* Consider a set of independent, centered random variables $\{Z_{i,j}\}_{1\le i<j}$ (real or complex-valued) which are independent from a family $\{Y_i\}_{1\le i}$ of i.i.d. centered real-valued random variables. Then a *Wigner matrix* is a matrix self-adjoint $N \times N$ matrix X_N with entries

$$X_N(j,i) = \overline{X}_N(i,j) = \begin{cases} Z_{i,j} & \text{if } i<j \\ Y_i & \text{if } i=j. \end{cases}$$

Let us assume the random variables have finite second moments (the opposite regime of infinite moments with "heavy tails" yields different results). To get some idea of the norm of this matrix, we may estimate trivially

$$\|X_N x\|^2 = \sum_{i=1}^N \left|\sum_{j=1}^N X_N(i,j)x_j\right|^2 \le \sum_{i,j=1}^N |X_N(i,j)|^2 \cdot \|x\|^2 \lesssim N^2\|x\|^2$$

where we used Cauchy-Schwarz and assumed the stronger condition that $|X_N(i,j)|^2 \lesssim 1$. This gives a bound as $\|X_N\| \lesssim N$, so $\sigma(X_N) \subseteq [-N,N]$. However, it turns out that the eigenvalues of Wigner matrices are actually of order $\sqrt{N}$ with overwhelming probability. Even more, Wigner proved the following fundamental result :

Let $(\lambda_i^N)_{i\le N}$ be the eigenvalues of X_N and define the *empirical spectral measure*

$$\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{\frac{\lambda_i^N}{\sqrt{N}}}\ ,$$

which is a probability measure on X_N. It is the uniform probability measure supported on the eigenvalues of $\frac{X_N}{\sqrt{N}}$. Note that $\mu_N(a,b) = \frac{\#\{i:\lambda_i^N\in(\sqrt{N}a,\sqrt{N}b)\}}{N}$.

Theorem 2.1 (Wigner [34])**.** *Assume* $\mathbb{E}(|Z_{1,2}|^2)=1$ *and* $\mathbb{E}(Y_1^2)<\infty$*. Then almost surely,* $\mu_N \xrightarrow{w} \rho_{sc}$*, where* $\mathrm{d}\rho_{sc} := \frac{1}{2\pi}\mathbf{1}_{[-2,2]}\sqrt{4-x^2}\,\mathrm{d}x$ *is the* semi-circle law. [2]

The theorem says that the (finite) spectrum of $\frac{1}{\sqrt{N}}X_N$ get denser and denser up to filling the interval $[-2,2]$. Moreover, we know precisely the relative density of the eigenvalues at the limit. In particular, there is more clustering around 0, and fewer around the edges ± 2.

Since the theorem does not assume anything on the distribution of the random variables, this gives a kind of *universal law.*

There are many directions one can explore from here. For example, in analogy to the central limit theorem, one may ask about the speed of convergence in the previous theorem. Curiously, the speed turns out to be $\frac{1}{N}$ instead of $\frac{1}{\sqrt{N}}$:

Theorem 2.2. *Assume the random variables have finite sixth-moments. Then for any* C^4*-function* f*,* $N(\mathbb{E}_{\mu_N}(f) - \mathbb{E}_{\rho_{sc}}(f))$ *converges to some nontrivial, universal limit.*

See [10] for a proof.

Empirical spectral measures give a rough idea on how the spectrum behaves globally. Can we somehow follow eigenvalues individually ?

2. Actually, Wigner proved the convergence in probability; the a.s. upgrade is based on Borel-Cantelli.

Theorem 2.3. *We have as $N \to \infty$, if $\tilde{\rho}_{sc}(x) := \frac{1}{2\pi}\sqrt{4-x^2}$,*

1. *w.h.p., for bulk E, $\#\{\lambda_j^N \in [E-\eta, E+\eta]\} \approx 2\eta N \tilde{\rho}_{sc}(E)$ for any $\eta = \eta_N > \frac{(\log N)^4}{N}$,*
2. *$\lambda_{\min}^N \to -2$ and $\lambda_{\max}^N \to 2$ a.s.*

See [18] for bulk E (i.e. energies well inside $(-2,2)$). Here "w.h.p." means "with high probability". Note that the restriction on η_N allows to take $\eta_N \to 0$ pretty fast (one cannot hope for more than the ideal scale, $\frac{1}{N}$). This strong result actually implies the *eigenvectors* are also very *delocalized*. We will discuss this later if time allows. The second statement is older and goes back to Füredi and Komlós in the 80s. See [9, Section 2.2] for the present version.

There are still a great number of results on Wigner matrices, but let us now move to the non-Hermitian context. Here the matrix is no longer self-adjoint, so the spectrum is now a subset of $\mathbb{C}$.

Consider a family $(X_{i,j})_{i,j\geq 1}$ of i.i.d. complex-valued random variables with variance $\mathrm{Var}(X_{i,j}) := \mathbb{E}(|X_{i,j}|^2) - |\mathbb{E}(X_{i,j})|^2 = 1$ and consider the random matrix $X_N = (X_{i,j})_{i,j\leq N}$.

Let λ_i^N be the eigenvalues of X_N and define $\mu_N = \frac{1}{N}\sum_{k=1}^N \delta_{\frac{\lambda_k^N}{\sqrt{N}}}$. Then

Theorem 2.4 (Girko's circular law)**.** *We have almost surely $\mu_N \xrightarrow{w} \mathcal{C}$ as $N \to \infty$, where $\mathcal{C}$ is the uniform* circular law *on the unit disc with density $\mathrm{d}\mathcal{C} = \pi^{-1}\mathbf{1}_{\{z\in\mathbb{C}:|z|\leq 1\}}\mathrm{d}z$.*

This gives again a very satisfactory universal law, which is however a lot more difficult to prove. The history starts with Girko [20], who gave very good ideas to prove the theorem, but the argument had technical problems. These were repaired by Bai [8] (who says he worked on this for 13 years from 1984 to 1997 and was in fact the hardest problem he had worked on) under some assumptions on moments. The assumptions were relaxed by many authors subsequently, culminating in the final statement above due to Terry Tao and Van Vu [32] (who did a lot of research on random matrices, Hermitian or not).

Further reading : See [11] for a more detailed survey on Wigner matrices. For non-Hermitian matrices, see [16].

2.2 Basic definitions

We finally introduce the project of these notes : Benjamini-Schramm convergence of graphs.

A *colored rooted graph* is a triple (G, o, W), where $G = (V, E)$ is a graph, o is a marked vertex in G called the *root*, and W is a map from $V \to \mathbb{R}$ which we see as a "coloring"; it can also be regarded as a potential on $\ell^2(V)$. This is a special case of what is called a *network* in [3], but the discussion of this note applies to this general setting as well. All graphs are assumed to be *locally finite*, i.e. each vertex has a finite degree.

If G is connected, we denote by $B_G(x,r)$ the r*-ball* $\{y \in V : d_G(x,y) \leq r\}$, where d_G is the length of the shortest path between x and y in G. We say there is a *rooted isomorphism* φ between two balls $B_{G_1}(x_1,r)$ and $B_{G_2}(x_2,r)$, and denote

$$\varphi : B_{G_1}(x_1,r) \xrightarrow{\sim} B_{G_2}(x_2,r)$$

if $\varphi : B_{G_1}(x_1,r) \to B_{G_2}(x_2,r)$ is a graph isomorphism with $\varphi(x_1) = x_2$.

We define a distance between colored connected graphs by

$$d_{loc}[(G_1,o_1,W_1),(G_2,o_2,W_2)] = \frac{1}{1+\alpha_{1,2}}\,, \tag{2-1}$$

where

$$\alpha_{1,2} = \sup\{r > 0 : \exists\, \varphi : B_{G_1}(o_1, [r]) \xrightarrow{\sim} B_{G_2}(o_2, [r]) \text{ with } |W_2(\varphi(v)) - W_1(v)| < 1/r \ \forall v \in B_{G_1}(o_1, [r])\} .$$

Lemma 2.5. *The function d_{loc} is a pseudo-metric on the set of colored graphs.*

Proof. We prove the triangle inequality.

Denote $\|W_j \circ \varphi - W_i\| = \sup_{x \in B_{G_i}(o_i,t)} |W_j(\varphi(v)) - W_i(v)|$, where the radius t of the ball will be clear from the context. We need to show that $\frac{1}{1+\alpha_{1,2}} \le \frac{1}{1+\alpha_{1,3}} + \frac{1}{1+\alpha_{2,3}}$.

Assume $\alpha_{1,2} = r$ and let $s > r$. Then there is no $\varphi : B_{G_1}(o_1, [s]) \xrightarrow{\sim} B_{G_2}(o_2, [s])$ having $\|W_2 \circ \varphi - W_1\| < 1/s$.

We have two cases : if there is no $\psi : B_{G_1}(o_1, [s]) \xrightarrow{\sim} B_{G_3}(o_3, [s])$, then $\alpha_{1,3} < s$. In this case, $d(G_1, G_2) = \frac{1}{1+r} = \sup_{s>r} \frac{1}{1+s} \le \frac{1}{1+\alpha_{1,3}} = d(G_1, G_3)$. Similarly, if there is no $\psi : B_{G_2}(o_2, [s]) \xrightarrow{\sim} B_{G_3}(o_3, [s])$, then $\alpha_{2,3} < s$, implying $d(G_1, G_2) \le d(G_2, G_3)$. In either case we get $d(G_1, G_2) \le d(G_1, G_3) + d(G_2, G_3)$.

So assume now that there exists $\psi_1 : B_{G_1}(o_1, [s]) \xrightarrow{\sim} B_{G_3}(o_3, [s])$ and $\psi_2 : B_{G_2}(o_2, [s]) \xrightarrow{\sim} B_{G_3}(o_3, [s])$. Then $\psi_2^{-1} \circ \psi_1 : B_{G_1}(o_1, [s]) \xrightarrow{\sim} B_{G_2}(o_2, [s])$. So by assumption, we must have $\|W_2 \circ \psi_2^{-1} \circ \psi_1 - W_1\| \ge 1/s$. Hence,

$$\begin{aligned}\frac{1}{s} &\le \|W_1 - W_2 \circ \psi_2^{-1} \circ \psi_1\| \le \|W_1 - W_3 \circ \psi_1\| + \|W_3 \circ \psi_1 - W_2 \circ \psi_2^{-1} \circ \psi_1\| \\ &= \|W_1 - W_3 \circ \psi_1\| + \|W_3 - W_2 \circ \psi_2^{-1}\| = \|W_3 \circ \psi_1 - W_1\| + \|W_3 \circ \psi_2 - W_2\| .\end{aligned}$$

So $\frac{1}{\|W_3 \circ \psi_1 - W_1\| + \|W_3 \circ \psi_2 - W_2\|} \le s$ and thus

$$\begin{aligned}\frac{1}{1+s} &\le \frac{1}{1 + \frac{1}{\|W_3 \circ \psi_1 - W_1\| + \|W_3 \circ \psi_2 - W_2\|}} = \frac{\|W_3 \circ \psi_1 - W_1\| + \|W_3 \circ \psi_2 - W_2\|}{1 + \|W_3 \circ \psi_1 - W_1\| + \|W_3 \circ \psi_2 - W_2\|} \\ &\le \frac{\|W_3 \circ \psi_1 - W_1\|}{1 + \|W_3 \circ \psi_1 - W_1\|} + \frac{\|W_3 \circ \psi_2 - W_2\|}{1 + \|W_3 \circ \psi_2 - W_2\|} \\ &= \frac{1}{1 + \frac{1}{\|W_3 \circ \psi_1 - W_1\|}} + \frac{1}{1 + \frac{1}{\|W_3 \circ \psi_2 - W_1\|}} .\end{aligned}$$

Finally, we have $\alpha_{1,3} \le \frac{1}{\|W_3 \circ \psi_1 - W_1\|}$ and $\alpha_{2,3} \le \frac{1}{\|W_3 \circ \psi_2 - W_1\|}$ by definition (because the r must satisfy $r < \frac{1}{\|W_j \circ \varphi - W_i\|}$). We thus showed that $\frac{1}{1+s} \le \frac{1}{1+\alpha_{1,3}} + \frac{1}{1+\alpha_{2,3}}$, and this completes the proof as before. □

We say that two colored graphs (G_1, o_1, W_1) and (G_2, o_2, W_2) are *equivalent* if there is a graph isomorphism $\varphi : G_1 \to G_2$ such that $\varphi(o_1) = o_2$ and $W_2 \circ \varphi = W_1$. We denote the equivalence class of (G, o, W) by $[G, o, W]$.

We denote by $\mathcal{G}_*$ the set of equivalence classes of connected colored rooted graphs.

Lemma 2.6. *The function d_{loc} induces a metric on $\mathcal{G}_*$.*

Proof. The value of $\alpha_{1,2}$ is independent of the choice of the representative in the equivalence class. Suppose $d_{loc}[(G, o, W), (G', o', W')] = 0$. We show that $[G, o, W] = [G', o', W']$.

For any $r \in \mathbb{N}$, there is $\varphi_r : B_G(o, r) \xrightarrow{\sim} B_{G'}(o', r)$ with $\|W' \circ \varphi_r - W\|_{B_{G'}(o',r)} < 1/r$.

Let $\varphi_r^{(n)} = \varphi_r|_{B_G(o,n)}$ for $r \ge n$. Then $\varphi_r^{(n)} : B_G(o, n) \to B_{G'}(o, r)$ for $r \ge n$. Actually, $\varphi_r^{(n)} : B_G(o, n) \to B_{G'}(o, n)$ for all $r \ge n$, because φ_r is a graph isomorphism and thus preserves neighbors. So $\varphi_r^{(n)} : B_G(o,n) \to \operatorname{Ran}\varphi_r^{(n)}$ is a graph isomorphism. So $|\operatorname{Ran}\varphi_r^{(n)}| = |B_G(o,n)| = |B_{G'}(o',n)|$, where the last equality holds because φ_n is a graph isomorphism. It follows that $\operatorname{Ran}\varphi_r^{(n)} = B_{G'}(o,n)$. Hence, $\varphi_r^{(n)} : B_G(o,n) \xrightarrow{\sim} B_{G'}(o,n)$ for all $r \geq n$. Since G and G' are locally finite, $\varphi_r^{(n)}$ has a convergent (in fact stationary) subsequence $\varphi_{r_j}^{(n)}$. Denote its limit by $\varphi^{(n)}$. Then $\varphi^{(n)} : B_G(o,n) \xrightarrow{\sim} B_{G'}(o',n)$.

Now $\varphi^{(n+1)}|_{B_G(o,n)} = \lim \varphi_{r_j}^{(n+1)}|_{B_G(o,n)} = \lim \varphi_{r_j}^{(n)} = \varphi^{(n)}$. So $\varphi^{(m)}|_{B_G(o,n)} = \varphi^{(n)}$ for all $m \geq n$. Hence, if for $v \in G$, say $v \in B_G(o,n)$ for some n, we put $\varphi(v) := \varphi^{(n)}(v)$, then φ is well-defined. Moreover, φ is bijective. In fact, if $w \in G'$, then $w = \varphi^{(n)}(v)$ for some n and v, so $w = \varphi(v)$; injectivity is similar. Finally φ is a graph isomorphism since given $v \sim v'$, say both in $B_G(o,n)$, we have $\{\varphi(v), \varphi(v')\} = \{\varphi^{(n)}(v), \varphi^{(n)}(v')\}$ is an edge in G'. We thus showed that $\varphi : (G,o) \xrightarrow{\sim} (G',o')$. It remains to check the coloring. Note that for any n and $r \geq n$, we have $\|W' \circ \varphi_r^{(n)} - W\|_{B_{G'}(o',n)} < 1/r$, since φ_r satisfies this on the bigger ball $B_{G'}(o,r)$. Since $\varphi^{(n)}$ is the limit of $\varphi_{r_j}^{(n)}$, we get $\|W' \circ \varphi^{(n)} - W\|_{B_{G'}(o',n)} = 0$. This shows that $W' \circ \varphi = W$. Hence, $[G,o,W] = [G',o',W']$ as required. □

Lemma 2.7. *The metric space $(\mathcal{G}_*, d_{loc})$ is a* Polish space*, i.e. separable and complete.*

Proof. Consider the family $\mathcal{C}_n = \{([n],1,W)\}$ of graphs on $[n] = \{1,\dots,n\}$ with W taking values in $\mathbb{Q}$. This is a countable family, since the set of maps W is just $\mathbb{Q}^{[n]}$. We denote $\mathcal{C} = \cup_n \mathcal{C}_n$. Given $[G,o,W] \in \mathcal{G}_*$ and $\varepsilon > 0$, choose $r \in \mathbb{N}$ such that $\frac{1}{1+r} < \varepsilon$. Since G is locally finite, the number of vertices in $B_G(o,r)$ is some $n = n(r) \in \mathbb{N}$. If $\varphi : B_G(o,r) \xrightarrow{\sim} ([n],1)$ is a rooted graph isomorphism, we may choose $W_n : [n] \to \mathbb{Q}$ such that $\|W_n \circ \varphi - W\| < 1/r$. Then $d_{loc}[(G,o,W),([n],1,W_n)] \leq \frac{1}{1+r}$, so $\mathcal{C}$ is dense in $\mathcal{G}_*$.

Next, suppose $([G_n,o_n,W_n])$ is Cauchy in $\mathcal{G}_*$. Assuming $|V(G_n)| = V_n$, we consider equivalently the sequence $([[V_n],1,W_n']])$, where $W_n' = W_n \circ \varphi_n$ for some $\varphi_n : G_n \xrightarrow{\sim} [V_n]$. We may check that for any $r \in \mathbb{N}$ there exists n_r and $\psi_n^m : B_{[V_n]}(1,r) \xrightarrow{\sim} B_{[V_m]}(1,r)$ such that $\|W_m' \circ \psi_n^m - W_n'\| < 1/r$ for all $n,m \geq n_r$. Hence,

(i) The sequence $[B_{[V_n]}(1,r)]$ is stationary in n for any r, (it stabilizes for $n \geq n_r$). So we may define a limit graph $[[V],1]$ iteratively such that for any r, $[B_{[V]}(1,r)] = [B_{[V_{n_r}]}(1,r)]$ (possibly $[V] = \mathbb{N}$).

(ii) More explicitly, $[B_{[V_n]}(1,r),1,W_n] = [B_{[V_{n_r}]}(1,r),1,W_n \circ \psi_{n_r}^n]$ for $n \geq n_r$. The sequence $\{W_n \circ \psi_{n_r}^n(v)\}_{n \geq n_r}$ is Cauchy in $\mathbb{R}$ for any $v \in B_{[V_{n_r}]}(1,r)$. In fact, for any r, $|W_n \circ \psi_{n_r}^n(v) - W_m \circ \psi_{n_r}^m(v)| \leq |W_n \circ \psi_{n_r}^n(v) - W_{n_r}(v)| + |W_m \circ \psi_{n_r}^m(v) - W_{n_r}(v)| \leq \frac{2}{r}$, for all $n,m \geq n_r$. Hence, $W_n \circ \psi_{n_r}^n(v) \to a_v \in [-A,A]$ as $n \to \infty$.

From (i), we may choose compatible isomorphisms $\phi^{(r)} : B_{[V]}(1,r) \xrightarrow{\sim} B_{[V_{n_r}]}(1,r)$ in the sense of the previous lemma. If we define $a : [V] \to [-A,A]$ by $a(v) = a_{\phi^{(r)}(v)}$ if $v \in B_{[V]}(1,r)$, we get that for any r, $d_{loc}[([V_n],1,W_n'),([V],1,a)] \leq \frac{1}{1+r}$ for $n \geq n_r$. Indeed, $[B_{[V_n]}(1,r),1,W_n] = [B_{[V_{n_r}]}(1,r),1,W_n \circ \psi_{n_r}^n] = [B_{[V]}(1,r),1,W_n \circ \psi_{n_r}^n \circ (\phi^{(r)})^{-1}]$.

Hence, the sequence converges to $[[V],1,a]$. □

We next define $\mathcal{G}_*^{D,A}$ to be the subset of equivalence classes $[G,o,W]$ such that G has degree uniformly bounded by D and W takes values in $[-A,A]$.

Lemma 2.8. *The metric space $(\mathcal{G}_*^{D,A}, d_{loc})$ is compact.*

Proof. Clearly, $\mathcal{G}_*^{D,A}$ is closed in $\mathcal{G}_*$ and thus complete. So it suffices to show that it is totally bounded. Suppose this is not true. Then for some $\varepsilon > 0$, there is no finite ε-net. So we may construct a sequence $([G_n,o_n,W_n])$ such that $d_{loc}[(G_n,o_n,W_n),(G_m,o_m,W_m)] \geq \varepsilon$ for all $n \neq m$. Hence, there exists $r \in \mathbb{N}$ such that $\alpha_{n,m} < r$ for all $n \neq m$, where $\alpha_{n,m}$ is as in (2-1). Now observe that there are only finitely many equivalence classes of rooted balls $B_{G_n}(o_n, r)$, since the degree is uniformly bounded by D. So if $\alpha_{n,m} < r$ for all $n \neq m$, there must be a sequence of isomorphic balls $B_{G_{n_j}}(o_{n_j}, r)$ on which the coloring is distant, say $\|W_{n_p} \circ \psi_q^p - W_{n_q}\| \geq 1/r$ for all $p \neq q$, where $\psi_q^p : B_{G_{n_q}}(o_{n_q}, r) \xrightarrow{\sim} B_{G_{n_p}}(o_p, r)$ is a rooted isomorphism. So the sequence $\{W_{n_1}, W_{n_2} \circ \psi_1^2, W_{n_3} \circ \psi_1^3, \ldots\}$ on $B_{G_{n_1}}(o_{n_1}, r)$ has no convergent subsequence. But this is a sequence in $[-A, A]^{B_{G_{n_1}}(o_{n_1}, r)}$, which is compact for the product topology, which coincides with the topology of pointwise convergence, i.e. the topology endowed by $\|W\| = \sup_{v \in B_{G_{n_1}}(o_{n_1}, r)} |W(v)|$. This contradiction completes the proof. □

So far we have defined a metric on connected colored rooted graphs. We now introduce a notion of convergence for unrooted graphs (G_n, W_n), which are not necessarily connected. The idea is to consider the convergence of the law of G_n under uniform rooting.

Since $\mathcal{G}_*$ is a Polish space, we may consider the set of probability measures on $\mathcal{G}_*$, denoted by $\mathcal{P}(\mathcal{G}_*)$. The latter is also a Polish space.

If (G, W) is a finite colored graph, $G = (V, E)$, we denote $(G(v), W)$ the subgraph spanned by the vertices in the connected component on v. We then define $U_{(G,W)} \in \mathcal{P}(\mathcal{G}_*)$ by

$$U_{(G,W)} = \frac{1}{|V|} \sum_{v \in V} \delta_{[G(v), v, W]} \,.$$

This captures the idea of choosing the root v uniformly at random in V.

If (G_n, W_n) is a sequence of finite colored graphs, we say that $\rho \in \mathcal{P}(\mathcal{G}_*)$ is the *local weak limit* of (G_n, W_n) if $U_{(G_n, W_n)}$ converges weakly to ρ in $\mathcal{P}(\mathcal{G}_*)$. This notion of convergence was first introduced in [13] and later generalized in [3]. It is often called the *Benjamini-Schramm convergence.*

Let $C(\mathcal{G}_*^{D,A})$ be the set of continuous functions $f : \mathcal{G}_*^{D,A} \to \mathbb{R}$. Recall that a linear subspace $\mathscr{A} \subset C(\mathcal{G}_*^{D,A})$ is called an *algebra* if is is closed under multiplication and contains the constant function 1. We say that $\mathscr{A}$ separates points if for any $[G, o, W] \neq [G', o', W'] \in \mathcal{G}_*^{D,A}$, there is some $f \in \mathscr{A}$ such that $f([G, o, W]) \neq f([G', o', W'])$.

Lemma 2.9. *Let (G_n, W_n) be a sequence of finite colored graphs, $G_n = (V_n, E_n)$, with degree uniformly bounded by D and coloring $W_n : V_n \to [-A, A]$ for all n. Then*

(1) *(G_n, W_n) has a subsequence which converges in the sense of Benjamini-Schramm, i.e. $U_{(G_{n_j}, W_{n_j})}$ converges weakly to some $\mu \in \mathcal{P}(\mathcal{G}_*^{D,A})$.*

(2) *(G_n, W_n) has a local weak limit ρ iff there is an algebra $\mathscr{A} \subset C(\mathcal{G}_*^{D,A})$ which separates points, such that for all $f \in \mathscr{A}$,*

$$\lim_{n\to\infty} \frac{1}{|V_n|} \sum_{v \in V_n} f\left([G_n(v), v, W_n]\right) = \int_{\mathcal{G}_*^{D,A}} f\left([G, o, W]\right) \mathrm{d}\rho\left([G, o, W]\right) \,.$$

Proof. Both items follow from the compactness of $\mathcal{G}_*^{D,A}$, see [25, Chapter 13]. □

The previous lemma gives a convenient criterion to prove that (G_n, W_n) has a local weak limit. However, it may not be very clear how a continuous function on $\mathcal{G}_*^{D,A}$ looks like. We start with the special case where $A = 0$, i.e. without coloring.

Lemma 2.10. *Let (G_n) be a sequence of finite graphs, $G_n = (V_n, E_n)$, with degree uniformly bounded by D. Then (G_n) has a local weak limit ρ iff for any r-ball $B_F(o, r)$,*

$$\lim_{n\to\infty} \frac{\#\{x : B_{G_n(x)}(x, r) \cong B_F(o, r)\}}{|V_n|} = \rho\left(\{[H, x] : B_H(x, r) \cong B_F(o, r)\}\right) \,.$$

Here, $B_{G_n(x)}(x,r) \cong B_F(o,r)$ *means there exists* $\varphi : B_{G_n(x)}(x,r) \xrightarrow{\sim} B_F(o,r)$.

Proof. Let $\mathcal{C}_{[F_r,o]} = \{[H,x] : B_H(x,r) \cong B_F(o,r)\}$. We first note that $\mathcal{C}_{[F_r,o]}$ is a clopen subset of $\mathcal{G}_*^{D,0}$. Indeed, given $[H_1,x_1] \in \mathcal{C}_{[F_r,o]}$ and $[H_2,x_2] \in \mathcal{G}_*^{D,0}$, denote $\alpha_{1,2} = \sup\{s \in \mathbb{N} : B_{H_1}(x_1,s) \cong B_{H_2}(x_2,s)\}$. Then for any $[H_2,x_2]$ with $\alpha_{1,2} \geq r$, we have $[H_2,x_2] \in \mathcal{C}_{[F_r,o]}$. Hence, $d_{loc}([H_1,x_1],[H_2,x_2]) \leq \frac{1}{1+r}$ implies $[H_2,x_2] \in \mathcal{C}_{[F_r,o]}$, so $\mathcal{C}_{[F_r,o]}$ is open. Next, suppose $([H_n,x_n]) \subset \mathcal{C}_{[F_r,o]}$ converges to some $[H,x]$. Then if $\alpha_n = \sup\{s \in \mathbb{N} : B_{H_n}(x_n,s) \cong B_H(x,s)\}$, we may find n_r such that $\alpha_n \geq r$ for all $n \geq n_r$. In particular, $B_H(x,r) \cong B_{H_{n_r}}(x_{n_r},r) \cong B_F(o,r)$, so $[H,x] \in \mathcal{C}_{[F_r,o]}$. Hence $\mathcal{C}_{[F_r,o]}$ is closed.

Note that $U_{G_n}(\mathcal{C}_{[F_r,o]}) = \frac{\#\{x : B_{G_n(x)}(x,r) \cong B_F(o,r)\}}{|V_n|}$. Since $\mathcal{C}_{[F_r,o]}$ is clopen, $\rho(\partial \mathcal{C}_{[F_r,o]}) = 0$, so if U_{G_n} converges weakly to ρ, then $U_{G_n}(\mathcal{C}_{[F_r,o]}) \to \rho(\mathcal{C}_{[F_r,o]})$ for any $B_F(o,r)$.

For the converse, note that since $\mathcal{C}_{[F_r,o]}$ is clopen, its indicator function $\chi_{\mathcal{C}_{[F_r,o]}}$ is continuous. Next, $\mathcal{C}_{[F_{r_1},o]} \cap \mathcal{C}_{[F'_{r_2},o']} = \mathcal{C}_{[F_{r_1},o]}$ if $B_F(o,r_1) \cong B_{F'}(o',r_2)$ and is empty otherwise, so $\chi_{\mathcal{C}_{[F_{r_1},o]}} \chi_{\mathcal{C}_{[F'_{r_2},o']}} = \chi_{\mathcal{C}_{[F_{r_1},o]}}$ or 0. Now if the limit in the lemma holds for any $B_F(o,r)$, this means that in Lemma 2.9(2), the limit is true for any $\chi_{\mathcal{C}_{[F_r,o]}}$. This implies it also holds for linear combinations thereof. Finally, it trivially holds for the constant functions 1 and 0. So the limit holds for the algebra of functions $\mathscr{A} = \{\alpha \chi_{\mathcal{C}_{[F_{r_1},o]}} + \beta \chi_{\mathcal{C}_{[F'_{r_2},o']}}\} \cup \{0,1\}$. Note that $\mathscr{A}$ separates points: if $[G,o] \neq [G',o']$, then $d_{loc}([G,o],[G',o']) > 0$, so we may find r such that $B_G(o,r)$ is not isomorphic to $B_{G'}(o,r)$ as rooted graphs. Taking $B_F(o,r) = B_G(o,r)$, we get $\chi_{\mathcal{C}_{[F_r,o]}}([G,o]) = 1$ but $\chi_{\mathcal{C}_{[F_r,o]}}([G',o']) = 0$. It now follows from by Lemma 2.9(2) that U_{G_n} converges weakly to ρ. □

We now discuss the general case. We first have a partial analogy with Lemma 2.10.

Lemma 2.11. *Let* (G_n, W_n) *be a sequence of finite colored graphs,* $G_n = (V_n, E_n)$, *with degree uniformly bounded by* D *and coloring* $W_n : V_n \to [-A, A]$ *for all* n. *If* (G_n, W_n) *has a local weak limit* ρ, *then for any* $r \in \mathbb{N}$ *and any* r*-ball* $(B_F(o,r), o, W_F)$,

$$\lim_{n\to\infty} \frac{\#\{x : \exists \varphi_n^x : B_{G_n(x)}(x,r) \xrightarrow{\sim} B_F(o,r) \text{ with } \|W_F \circ \varphi_n^x - W_n\|_{B_{G_n(x)}(x,r)} < 1/r\}}{|V_n|}$$
$$= \rho\left(\{[H,x,W] : \exists \varphi : B_H(x,r) \xrightarrow{\sim} B_F(o,r) \text{ with } \|W_F \circ \varphi - W\|_{B_H(x,r)} < 1/r\}\right).$$

Proof. Given an r-ball $(B_F(o,r), o, W_F)$ with $r \in \mathbb{N}$, let

$$\mathcal{C}_F = \{[H,x,W] : \exists \varphi : B_H(x,r) \xrightarrow{\sim} B_F(o,r) \text{ with } \|W_F \circ \varphi - W\|_{B_H(x,r)} < 1/r\}.$$

Then

$$\mathcal{C}_F = \left\{[H,x,W] : d_{loc}([F,o,W_F],[H,x,W]) \leq \frac{1}{1+r}\right\}.$$

Hence, $\mathcal{C}_F$ is closed. It is also open: if $[H,x,W] \in \mathcal{C}_F$, then there exists $\varphi : B_F(o,r) \xrightarrow{\sim} B_H(x,r)$ with $\|W \circ \varphi - W_F\|_{B_F(o,r)} < 1/r$. Choose $s \in \mathbb{N}$, $s > r$, such that $0 < \frac{1}{s} < \frac{1}{r} - \|W \circ \varphi - W_F\|_{B_F(o,r)}$. If $d_{loc}([H,x,W],[H',x',W']) < \frac{1}{1+s}$, there exists $\psi : B_H(x,s) \xrightarrow{\sim} B_{H'}(x',s)$ with $\|W' \circ \psi - W\|_{B_H(x,r)} < 1/s$. As $\|W' \circ \psi - W\|_{B_H(x,r)} = \|W' \circ \psi \circ \varphi - W \circ \varphi\|_{B_F(o,r)}$, it follows that $\|W' \circ \psi \circ \varphi - W_F\|_{B_F(o,r)} < 1/s + \|W \circ \varphi - W_F\|_{B_F(o,r)} < 1/r$. But $\psi \circ \varphi : B_F(o,r) \xrightarrow{\sim} B_{H'}(x',r)$ since $s > r$. Hence, $[H',x',W'] \in \mathcal{C}_F$ and $\mathcal{C}_F$ is open.

Note that $U_{(G_n,W_n)}(\mathcal{C}_F) = \frac{\#\{x : [B_{G_n(x)}(x,r),x,W] \in \mathcal{C}_F\}}{|V_n|}$. Since $\mathcal{C}_F$ is clopen, $U_{(G_n,W_n)}(\partial \mathcal{C}_F) = 0$, so if $U_{(G_n,W_n)}$ converges weakly to ρ, then $U_{G_n}(\mathcal{C}_F) \to \rho(\mathcal{C}_F)$. □

To obtain a converse, one needs to assume moreover that the limit holds for all elements of the form $\mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}$, in order to argue as before.

Under the hypotheses of the lemma, it is also true that

$$\lim_{n\to\infty} \frac{\#\{x : B_{G_n(x)}(x,r) \cong B_F(o,r)\}}{|V_n|} = \rho\left(\{[H,x,W] : B_H(x,r) \cong B_F(o,r)\}\right) ,$$

since the sets on the RHS are still clopen. So as expected, if (G_n, W_n) converge as colored graphs, they converge in particular as graphs without coloring.

2.3 Examples

We start with simple examples without coloring.

2.3.1 Cycle graphs

The cycle graph with n vertices C_n converges to $\mathbb{Z}$ in the sense of Benjamini-Schramm. More precisely, C_n has the local weak limit $\delta_{[\mathbb{Z},o]}$, where $o \in \mathbb{Z}$ is an arbitrary root. Indeed, given $r \in \mathbb{N}$, if $B_F(o,r)$ is isomorphic to an r-ball in $\mathbb{Z}$, then $\frac{\#\{x:B_{C_n}(x,r)\cong B_F(o,r)\}}{n} = 1$ for all $n > r$. If $B_F(o,r)$ is not isomorphic to an r-ball in $\mathbb{Z}$, then $\frac{\#\{x:B_{C_n}(x,r)\cong B_F(o,r)\}}{n} = 0$ for all $n > r$. So the limit of $\frac{\#\{x:B_{C_n}(x,r)\cong B_F(o,r)\}}{n}$ is 1 (or 0) if $B_F(o,r)$ is isomorphic to an r-ball in $\mathbb{Z}$ (or not). Since $\delta_{[\mathbb{Z},o]}(\{[H,x] : B_H(x,r) \cong B_F(o,r)\})$ has the same values then the claim follows from Lemma 2.10.

2.3.2 Lattice cubes

The cubes $\Lambda_n = \{1,\ldots,n\}^d$ converge to $\mathbb{Z}^d$ in the sense of Benjamini-Schramm. Indeed, if $B_F(o,r)$ is isomorphic to an r-ball in $\mathbb{Z}^d$, then $\frac{\#\{x:B_{\Lambda_n}(x,r)\cong B_F(o,r)\}}{n^d} = \frac{(n-2r)^d}{n^d} \to 1$. Otherwise, $\frac{\#\{x:B_{\Lambda_n}(x,r)\cong B_F(o,r)\}}{n^d} \le \frac{(2r)^d}{n^d} \to 0$. It follows as before that U_{Λ_n} has the local weak limit $\delta_{[\mathbb{Z}^d,o]}$, where $o \in \mathbb{Z}^d$ is arbitrary.

2.3.3 Regular graphs with few cycles

Let $G_N = (V_N, E_N)$ be a sequence of $(q+1)$-regular connected graphs with $|V_N| = N$. As in [5, 4], we define the property

(BST) For all $R > 0$,

$$\lim_{N\to\infty} \frac{|\{x \in V_N : \rho_{G_N}(x) < R\}|}{N} = 0\,,$$

where $\rho_{G_N}(x)$ is the *injectivity radius* at x, i.e. the largest ρ such that $B_{G_N}(x,\rho)$ is a tree. This property holds in particular if the girth of G_N grows to infinity.

We claim that (G_N) satisfies **(BST)** iff (G_N) converges to the $(q+1)$-regular tree $\mathcal{T}_q$ in the sense of Bejamini-Schramm, i.e. iff (G_N) has the local weak limit $\delta_{[\mathcal{T}_q,o]}$, where $o \in \mathcal{T}_q$ is an arbitrary root.

Indeed, let $B_F(o,r)$ be an r-ball and assume (G_N) satisfies **(BST)**. If $B_F(o,r)$ is isomorphic to a ball in $\mathcal{T}_q$, then $\frac{\#\{x:B_{G_N}(x,r)\cong B_F(o,r)\}}{|V_N|} = \frac{\#\{x:\rho_{G_N}(x)\ge r\}}{N} \to 1$. If $B_F(o,r)$ is not isomorphic to a ball in $\mathcal{T}_q$, then $\frac{\#\{x:B_{G_N}(x,r)\cong B_F(o,r)\}}{|V_N|} \le \frac{\#\{x:\rho_{G_N}(x)<r\}}{N} \to 0$. It follows as before that (G_N) has the local weak limit $\delta_{[\mathcal{T}_q,o]}$.

Conversely, if (G_N) has the local weak limit $\delta_{[\mathcal{T}_q,o]}$, given $R > 0$, pick a ball $B_F(o,R)$ in $\mathcal{T}_q$, Then $\frac{\#\{x:\rho_{G_N}(x)\ge R\}}{N} = \frac{\#\{x:B_{G_N}(x,R)\cong B_F(o,R)\}}{|V_N|} \to 1$, so **(BST)** follows.

2.3.4 Graphs with bounded degree

If we assume that $G_N = (V_N, E_N)$ is a sequence of graphs, $|V_N| = N$, with degree uniformly bounded by D, then **(BST)** no longer guarantees convergence. For instance if G_{2N} are 3-regular and G_{2N+1} are 4-regular, and if (G_N) satisfies **(BST)**, then G_{2N} will converge to $\mathcal{T}_2$ while G_{2N+1} will converge to $\mathcal{T}_3$.

2.3.5 Tree balls

Fix a root o in the $(q+1)$-regular tree $\mathcal{T}_q$ and let $G_N = B_{\mathcal{T}_q}(o, N)$. Then (G_N) does not converge to $\mathcal{T}_q$ in the sense of Benjamini-Schramm. Indeed, if $B_F(o, r)$ is isomorphic to a ball in $\mathcal{T}_q$, then $\frac{\#\{x : B_{G_N}(x,r) \cong B_F(o,r)\}}{|V_N|} = \frac{|V_{N-r}|}{|V_N|} \to \frac{1}{q^r}$, since $|V_n| = 1 + (q+1)\sum_{j=1}^{n} q^{j-1} = 1 + (q+1)\frac{1-q^n}{1-q}$, so that $\frac{|V_{N-r}|}{|V_N|} = \frac{(q-1)q^{-N} + (q+1)(q^{-r} - q^{-N})}{(q-1)q^{-N} + (q+1)(1-q^{-N})} \to q^{-r}$. This already shows the local weak limit cannot be $\mathcal{T}_q$. For $B_F(o, r)$ which are not isomorphic to an r-ball in $\mathcal{T}_q$, the value of $\frac{\#\{x : B_{G_N}(x,r) \cong B_F(o,r)\}}{|V_N|}$ is 0 if $B_F(o, r)$ is also not isomorphic to any $B_{G_N}(x, r)$ with $x \in S_n$, $N - r + 1 \le n \le N$, where S_n is the n-th sphere. If $B_F(o, r)$ is isomorphic to $B_{G_N}(x, r)$ with $x \in S_{N-j+1}$, then $\frac{\#\{x : B_{G_N}(x,r) \cong B_F(o,r)\}}{|V_N|} = \frac{|S_{N-j+1}|}{|V_N|} = \frac{(q+1)q^{N-j}}{1 + \frac{q+1}{q-1}(q^N - 1)} \to \frac{q-1}{q^j}$. Based on this information, we construct a random rooted tree (T_q^*, o) called the *canopy tree,* cf. [2] and [19, Chapter 14] :

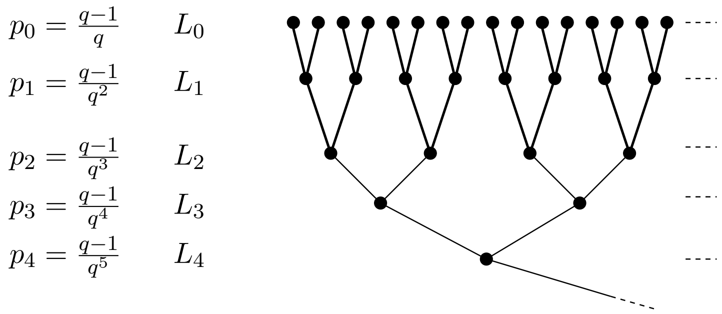

Figure 2.1 – The canopy tree, as introduced in [2].

This is a fixed infinite tree T_q^*; the randomness comes from the root. More precisely, this tree is not transitive, so the position of the root matters. We divide the tree into infinite levels $(L_j)_{j=0}^{\infty}$ as in Figure 2.1. For fixed j, all trees (T_q^*, o) with $o \in L_j$ are equivalent. By some abuse of notation we let $[T_q^*, o] = L_j$ in this case. We then define $\rho \in \mathcal{P}(\mathcal{G}_*^{D,A})$ by $\rho = \sum_{j=0}^{\infty} \frac{q-1}{q^{j+1}} \delta_{L_j}$. This is indeed a probability measure, since $\sum_{j=0}^{\infty} \frac{q-1}{q^{j+1}} = 1$. By construction, $B_{T_q^*}(o, r)$ is isomorphic to an r-ball $B_F(o, r)$ in $\mathcal{T}_q$ precisely when o is in any level L_j with $j \ge r$. In other words, $\rho(\{[H, x, W] : B_H(x, r) \cong B_F(o, r)\}) = \sum_{j=r}^{\infty} \frac{q-1}{q^{j+1}} = \frac{q-1}{q^{r+1}} \sum_{j=0}^{\infty} q^{-j} = \frac{q-1}{q^{r+1}} \frac{q}{q-1} = \frac{1}{q^r}$, which is the limiting value we obtained along G_N. Similarly, if $B_F(o, r)$ is not isomorphic to any $B_{G_N}(x)$ with $x \in S_n$, $N - r \le n \le N$, we find $\rho(\{[H, x, W] : B_H(x, r) \cong B_F(o, r)\}) = 0$, and if $B_F(o, r)$ is isomorphic to $B_{G_N}(x, r)$ with $x \in S_{N-j+1}$, then $\rho(\{[H, x, W] : B_H(x, r) \cong B_F(o, r)\}) = \frac{q-1}{q^j} \delta_{L_{j-1}}(L_{j-1}) = \frac{q-1}{q^j}$. This completes the proof that ρ is the local weak limit of (G_N).

We now turn to colored graphs.

2.3.6 Colored graphs with few cycles

Let (G_N, W_N) be a sequence of colored connected graphs $G_N = (V_N, E_N)$ with degree uniformly bounded by D, coloring $W_N : V_N \to [-A, A]$ for all N and $|V_N| = N$. Let $\mathcal{T}_*^{D,A}$ be the subset of colored rooted trees in $\mathcal{G}_*^{D,A}$.

We show that if (G_N, W_N) has a local weak limit ρ which is concentrated on $\mathcal{T}_*^{D,A}$, then (G_N) satisfies **(BST)**. Conversely, if (G_N) satisfies **(BST)**, and if (G_{N_j}, W_{N_j}) is a subsequence with a local weak limit ρ, then ρ is concentrated on $\mathcal{T}_*^{D,A}$.

Indeed, we may assume the R in **(BST)** is in $\mathbb{N}^*$. We observe that $\frac{\#\{x:\rho_{G_N}(x)<R\}}{N} = \frac{\#\{x:B_{G_N}(x,R) \text{ is not a tree }\}}{N} = U_{(G_N,W_N)}(\{[H,x,W] : B_H(x,R) \text{ is not a tree }\})$. Let $\mathscr{C}_R = \{[H,x,W] : B_H(x,R) \text{ is not a tree }\}$. Then $\mathscr{C}_R$ is clopen. Indeed, if $[H,x,W] \in \mathscr{C}_R$ and $d_{loc}[(H,x,W),(H',x',W')] < \frac{1}{1+R}$, then $B_{H'}(x',R) \cong B_H(x,R)$, so $[H',x',W'] \in \mathscr{C}_R$. Hence, $\mathscr{C}_R$ is open. If $[H_n,x_n,W_n] \subset \mathscr{C}_R$ converges to some $[H,x,W]$, then $B_H(x,R) \cong B_{H_n}(x_n,R)$ for all $n \ge n_R$ and thus $[H,x,W] \in \mathscr{C}_R$. So $\mathscr{C}_R$ is closed.

If (G_N, W_N) has a local weak limit ρ concentrated on $\mathcal{T}_*^{D,A}$, then $U_{(G_N,W_N)}(\mathscr{C}_R) \to \rho(\mathscr{C}_R) = 0$. Hence (G_N) satisfies **(BST)**. Conversely, if (G_N) satisfies **(BST)** and ρ is the local weak limit of a subsequence, then we need to show that for any $\mathscr{M} \subset \mathcal{G}_*^{D,A}$, we have $\rho(\mathscr{M}) = \rho(\mathscr{M} \cap \mathcal{T}_*^{D,A})$. For this, note that $(\mathcal{T}_*^{D,A})^c = \cup_{R\in\mathbb{N}} \mathscr{C}_R$, with $\mathscr{C}_1 \subseteq \mathscr{C}_2 \subseteq \dots$. Hence, $\rho[(\mathcal{T}_*^{D,A})^c] = \lim_{R\to\infty} \rho(\mathscr{C}_R)$. But by hypothesis, $\rho(\mathscr{C}_R) = \lim_{j\to\infty} U_{(G_{N_j},W_{N_j})}(\mathscr{C}_R) = \lim_{j\to\infty} \frac{\#\{x:\rho_{N_j}(x)<R\}}{N_j} = 0$. So ρ is concentrated on $\mathcal{T}_*^{D,A}$.

2.3.7 The Anderson model

In this example we closely follow our presentation in [6].

Let $\Omega = [-A,A]^{\mathbb{Z}^d}$ and define $\mathbb{P}$ on Ω by $\mathbb{P} = \otimes_{v\in\mathbb{Z}^d} \nu$ for some probability measure ν on $[-A,A]$. Given $\omega = (\omega_v) \in \Omega$, define $W^\omega(v) = \omega_v$ for $v \in \mathbb{Z}^d$. Then the $\{\omega_v\}_{v\in\mathbb{Z}^d}$ are i.i.d. random variables with common distribution ν.

Let $\Lambda_n = \{1,\dots,n\}^d$. Given $\omega \in \Omega$, we define $W_n^\omega(v) = \omega_v$ for $v \in \Lambda_n$.

We will show that for $\mathbb{P}$-a.e. ω, the sequence of graphs (Λ_n, W_n^ω) has a local weak limit ρ which is concentrated on $\{[\mathbb{Z}^d, 0, W^\omega] : \omega \in \Omega\}$, and acts by taking the expectation w.r.t. $\mathbb{P}$. More precisely, denoting $D = 2d$,

$$\int_{\mathcal{G}_*^{D,A}} f([G,o,W])\,\mathrm{d}\rho([G,o,W]) = \mathbb{E}[f([\mathbb{Z}^d,0,W^\omega])]\,.$$

Let $\mathscr{A} = \cup_{r\in\mathbb{N}} \mathscr{A}_r$, where

$$\mathscr{A}_r = \{f \in C(\mathcal{G}_*^{D,A}) : f([G,o,W]) = f([G',o',W']) \\ \text{if } [B_G(o,r),o,W] = [B_{G'}(o',r),o',W']\}\,.$$

Then $\mathscr{A}$ is an algebra of continuous functions containing 1. To see that it separates points, let $[G,o,W] \neq [G',o',W']$. Then we may find $r \in \mathbb{N}$ with $d_{loc}([G,o,W],[G',o',W']) > \frac{1}{1+r}$. Define $\mathcal{C}_G = \{[H,x,V] : d_{loc}([G,o,W],[H,x,V]) \le \frac{1}{1+r}\}$. We showed in Lemma 2.11 that $\chi_{\mathcal{C}_G}$ is continuous. Moreover, $\chi_{\mathcal{C}_G}([H,x,V]) = \chi_{\mathcal{C}_G}([H',x',V'])$ if $[B_H(o,r),o,V] = [B_{H'}(o',r),o',V']$. Indeed, if $\chi_{\mathcal{C}_G}([H,x,V]) = 1$, there is $\varphi : B_H(x,r) \xrightarrow{\sim} B_G(o,r)$ with $\|W \circ \varphi - V\|_{B_H(x,r)} < 1/r$. If $\psi : B_{H'}(x',r) \xrightarrow{\sim} B_H(o,r)$ with $V \circ \psi = V'$, then $\varphi \circ \psi : B_{H'}(x',r) \xrightarrow{\sim} B_G(o,r)$ has $\|W \circ \varphi \circ \psi - V'\| < 1/r$ and thus $\chi_{\mathcal{C}_G}([H',x',V']) = 1$. Similarly, if $\chi_{\mathcal{C}_G}([H,x,V]) = 0$, no such φ exists and $\chi_{\mathcal{C}_G}([H',x',V']) = 0$. Hence, $\chi_{\mathcal{C}_G} \in \mathscr{A}_r$. Finally, $\chi_{\mathcal{C}_G}([G,o,W]) = 1$ while $\chi_{\mathcal{C}_G}([G',o',W']) = 0$. We thus showed that $\mathscr{A}$ separates points.

Using Lemma 2.9, it now suffices to show that there exists $\Omega_0 \subseteq \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that for any $\omega \in \Omega_0$ and any $f \in \mathscr{A}$, we have

$$\lim_{n\to\infty} \frac{1}{n^d} \sum_{x\in\Lambda_n} f([\Lambda_n, x, W_n^\omega]) = \mathbb{E}[f([\mathbb{Z}^d, 0, W^\omega])] \,. \tag{3-1}$$

For this, we first adapt the strong law of large numbers in [17, Theorem 2.3.5]. Given $f \in \mathscr{A}_r$, let

$$Y_x = Y_x^{(n)} = f([\Lambda_n, x, W_n^\omega]) - \mathbb{E}[f([\Lambda_n, x, W_n^\omega])] \quad \text{and} \quad S_n = \frac{1}{n^d}\sum_{x\in\Lambda_n} Y_v \,.$$

Then $\mathbb{E}[Y_x] = 0$. Moreover, Y_x only depends on $(\omega_z)_{z\in B_{\Lambda_n}(x,r)}$, since $f([\Lambda_n, x, W_n^\omega]) = f([\Lambda_n, x, \tilde{W}_n^\omega])$ if $W_n^\omega = \tilde{W}_n^\omega$ on $B_{\Lambda_n}(x,r)$. It follows that Y_x and Y_y are independent if $d_{\Lambda_n}(x,y) > 2r$. Now

$$\mathbb{E}\Big[\sum_{x\in\Lambda_n} Y_x\Big]^4 = \sum_{x\in\Lambda_n} \mathbb{E}(Y_x^4) + 6\sum_{x,y\in\Lambda_n} \mathbb{E}(Y_x^2Y_y^2) + 4\sum_{x,y\in\Lambda_n} \mathbb{E}(Y_xY_y^3 + Y_yY_x^3)$$
$$+ 12\sum_{x,y,z\in\Lambda_n} \mathbb{E}(Y_xY_yY_z^2 + Y_xY_y^2Y_z + Y_x^2Y_yY_z) + 24\sum_{x,y,z,t\in\Lambda_n} \mathbb{E}(Y_xY_yY_zY_t)\,.$$

The first three sums are $O(n^d)$ and $O(n^{2d})$. For the fourth, note that $\mathbb{E}(Y_xY_yY_z^2) = 0$ if $d(x,y) > 4r$, since either $d(x,z) > 2r$ and Y_x is independent of the pair (Y_y, Y_z), or $d(y,z) > 2r$ and Y_y is independent of (Y_x, Y_z). Thus, we have either $\mathbb{E}(Y_xY_yY_z^2) = \mathbb{E}(Y_x)\,\mathbb{E}(Y_yY_z^2) = 0$ or $\mathbb{E}(Y_yY_xY_z^2) = \mathbb{E}(Y_y)\,\mathbb{E}(Y_xY_z^2) = 0$. Hence, $|\sum_{x,y,z\in\Lambda_n} \mathbb{E}(Y_xY_yY_z^2)| \le n^{2d}(4r)^d(2\|f\|_\infty)^4$. The other terms of this sum are treated similarly.

Finally, for $\mathbb{E}(Y_xY_yY_zY_t)$ to be non zero, each point must be at distance $\le 2r$ from one of the three others. Hence, we must have $[d(x,y) \le 2r$ and $d(z,t) \le 2r]$ (or a permutation thereof) or $[d(x,\bullet) \le 8r$ for $\bullet = y,z,t.]$ It follows that $\sum_{x,y,z,t\in\Lambda_n} |\mathbb{E}(Y_xY_yY_zY_t)| \le 3n^{2d}(2r)^{2d}(2\|f\|_\infty)^4 + n^d(8r)^{3d}(2\|f\|_\infty)^4$. In any case $\mathbb{E}(|S_n|^4) \le C_{r,f,d}n^{-2d}$.

By the Borel-Cantelli Lemma, if $A_{n,f}^\varepsilon = \{|S_n| > \varepsilon\}$, then $\mathbb{P}(A_{n,f}^\varepsilon \text{ i.o.}) = 0$. Thus, if

$$\Omega_{0,f}^\varepsilon = \{A_{n,f}^\varepsilon \text{ occurs finitely often}\}\,,$$

we have $\mathbb{P}(\Omega_{0,f}^\varepsilon) = 1$. Since $C(\mathcal{G}_*^{D,A})$ is separable, $\mathscr{A}$ is separable, and we may choose a countable dense subset $\{f_j\} \subset \mathscr{A}$. We then let $\Omega_0 = \cap_{\varepsilon\in\mathbb{Q}^+} \cap_{j\in\mathbb{N}} \Omega_{0,f_j}^\varepsilon$. Then $\mathbb{P}(\Omega_0) = 1$.

Let $\omega \in \Omega_0$. Given $j \in \mathbb{N}$ and $\varepsilon > 0$ let $0 < \varepsilon' < \varepsilon$, $\varepsilon' \in \mathbb{Q}^+$. Then $\omega \in \Omega_{0,f_j}^{\varepsilon'}$, so there is n_ω such that $|S_n| \le \varepsilon' < \varepsilon$ for any $n > n_\omega$. Hence, $S_n \to 0$ for any $\omega \in \Omega_0$.

Now if $f \in \mathscr{A}$, say $f \in \mathscr{A}_r$, we have

$$\Big|\frac{1}{n^d}\sum_{x\in\Lambda_n} f([\Lambda_n, x, W_n^\omega]) - \mathbb{E}[f([\mathbb{Z}^d, 0, W^\omega])]\Big|$$
$$\le |S_n| + \Big|\frac{1}{n^d}\sum_{x\in\Lambda_n} \mathbb{E}[f([\Lambda_n, x, W_n^\omega])] - \mathbb{E}[f([\mathbb{Z}^d, 0, W^\omega])]\Big|\,.$$

Assume $n > r$. If $x \in \{r+1, \dots, n-r\}^d =: C_n^r$, there is $\varphi : B_{\Lambda_n}(x,r) \xrightarrow{\sim} B_{\mathbb{Z}^d}(0,r)$. In fact, we take $\varphi(v) = v - x$. Denoting $W_x^\omega(v) = W^\omega(v+x)$, we get $[B_{\Lambda_n}(x,r), x, W_n^\omega] =$

$[B_{\mathbb{Z}^d}(0,r),0,W_x^\omega]$, so $f([\Lambda_n,x,W_n^\omega]) = f([\mathbb{Z}^d,0,W_x^\omega])$. By usual measure-preserving transformations, we check that $\mathbb{E}[f([\mathbb{Z}^d,0,W_x^\omega])] = \mathbb{E}[f([\mathbb{Z}^d,0,W^\omega])]$. Hence,

$$\Big|\frac{1}{n^d}\sum_{x\in\Lambda_n} f([\Lambda_n,x,W_n^\omega]) - \mathbb{E}[f([\mathbb{Z}^d,0,W^\omega])]\Big|$$
$$\le |S_n| + \frac{1}{n^d}\sum_{x\notin C_n^r}\Big|\mathbb{E}[f([\Lambda_n,x,W_n^\omega])] - \mathbb{E}[f([\mathbb{Z}^d,0,W^\omega])]\Big| \le |S_n| + \frac{(2r)^d}{n^d}(2\|f\|_\infty)\,.$$

Taking $n\to\infty$, it follows that if $\omega\in\Omega_0$, then (3-1) is true for any $f\in\{f_j\}$, the dense subset of $\mathscr{A}$. Arguing as in [25, Corollary 15.3], the proof is complete.

2.4 Convergence of spectral measures

Our aim in this section is to show that the Benjamini-Schramm convergence implies the convergence of the *mean* spectral measures. This can be interpreted as the assertion that the integrated density of states of the limit operator has a finite-volume approximation. Though this can be proved directly, we will first prove a convergence result for *rooted* spectral measures as in [31, Chapter 2], which is of independent interest.

Let $[G,o,W]\in\mathcal{G}_*^{D,A}$, let $\mathcal{A}$ be the adjacency matrix on G and define the (Schrödinger) operator $H=\mathcal{A}+W$. This is a bounded self-adjoint operator. We sometimes denote $H=H_{(G,W)}$ to avoid confusion. We define the rooted spectral measure $\mu_o^{(G,W)}$ by

$$\mu_o^{(G,W)}(J) = \langle\delta_o,\chi_J(H)\delta_o\rangle \qquad \text{for Borel } J\subseteq\mathbb{R}\,.$$

Lemma 2.12. *Suppose $[G_n,o_n,W_n]\subset\mathcal{G}_*^{D,A}$ converges to $[G,o,W]$ in the metric topology of $(\mathcal{G}_*^{D,A},d_{loc})$. Then $\mu_{o_n}^{(G_n,W_n)}$ converges weakly to $\mu_o^{(G,W)}$. So for any continuous $\varphi:\mathbb{R}\to\mathbb{R}$, we have*

$$\lim_{n\to\infty}\langle\delta_{o_n},\varphi(H_{(G_n,W_n)})\delta_{o_n}\rangle = \langle\delta_o,\varphi(H_{(G,W)})\delta_o\rangle\,.$$

Proof. Since all operators $H_n=\mathcal{A}_n+W_n$ and $H=\mathcal{A}+W$ are uniformly bounded by some $A+D$, the supports of the spectral measures is compact, so it suffices to show that for any $k\in\mathbb{N}$, $\int t^k\mathrm{d}\mu_{o_n}(t)\to\int t^k\mathrm{d}\mu_o(t)$; see [25, Chapter 13].

Given $k\in\mathbb{N}$, choose an arbitrary integer $r\ge k$. Then we may find n_r such that for $n\ge n_r$, there exists $\varphi_r : B_{G_n}(o_n,r)\xrightarrow{\sim} B_G(o,r)$ with $\|W\circ\varphi_r - W_n\|_{B_{G_n}(o,r)} < 1/r$. Now

$$\int t^k\,\mathrm{d}\mu_{o_n}(t) = \langle\delta_{o_n},H_n^k\delta_{o_n}\rangle = \sum_{u_0,\dots,u_{k-1}} H_n(o_n,u_0)H_n(u_0,u_1)\dots H_n(u_{k-1},o_n)\,,$$

and $H_n(v,w)=(\mathcal{A}_n\delta_w)(v)+W_n(v)\delta_w(v)$. So the RHS only depends on $B_{G_n}(o_n,k)$ and its coloring. As $r\ge k$ and $\varphi_r : B_{G_n}(o_n,r)\xrightarrow{\sim} B_G(o,r)$, if we let $\mathcal{H}_n=\mathcal{A}+W_n\circ\varphi_r^{-1}$ on G, we get $\langle\delta_{o_n},H_n^k\delta_{o_n}\rangle = \langle\delta_o,\mathcal{H}_n^k\delta_o\rangle$. So for $n\ge n_r$,

$$\Big|\int t^k\,\mathrm{d}\mu_{o_n}(t) - \int t^k\,\mathrm{d}\mu_o(t)\Big| = |\langle\delta_o,(\mathcal{H}_n^k-H^k)\delta_o\rangle| = \Big|\Big\langle\delta_o,\sum_{i=1}^k\mathcal{H}_n^{k-i}(\mathcal{H}_n-H)H^{i-1}\delta_o\Big\rangle\Big|$$
$$\le C_{k,D,A}\|W_n\circ\varphi_r^{-1}-W\|_{B_G(o,r)} \le \frac{C_{k,D,A}}{r}\,.$$

Since $r\ge k$ is arbitrary, this completes the proof. ☐

If (G, W) if a finite colored graph, $G = (V, E)$, with degree uniformly bounded by D and coloring in $[-A, A]$, we define the mean spectral measure

$$\mu^{(G,W)} = \frac{1}{|V|} \sum_{x \in V} \mu_x^{(G,W)} .$$

Corollary 2.13. *Suppose (G_n, W_n) is a sequence of finite colored graphs with degrees uniformly bounded by D and coloring $W_n : V_n \to [-A, A]$ for all n. If (G_n, W_n) has a local weak limit ρ, then $\mu^{(G_n,W_n)}$ converges weakly to $\int_{\mathcal{G}_*^{D,A}} \mu_o^{(G,W)} \,\mathrm{d}\rho([G, o, W])$. So for any continuous $\varphi : \mathbb{R} \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \frac{1}{|V_n|} \operatorname{tr}[\varphi(H_{(G_n,W_n)})] = \int_{\mathcal{G}_*^{D,A}} \langle \delta_o, \varphi(H_{(G,W)})\delta_o\rangle \,\mathrm{d}\rho([G, o, W]) .$$

Proof. Given continuous $\varphi : \mathbb{R} \to \mathbb{R}$, define the transform $\widehat{\varphi} : \mathcal{G}_*^{D,A} \to \mathbb{R}$ by $\widehat{\varphi}([G, o, W]) = \int \varphi(t) \,\mathrm{d}\mu_o^{(G,W)}(t)$. Then $\widehat{\varphi}$ is continuous on $\mathcal{G}_*^{D,A}$ by Lemma 2.12. It is also bounded since $\mathcal{G}_*^{D,A}$ is compact. By hypothesis, $U_{(G_n,W_n)}$ converges weakly to ρ. It follows that $\int \widehat{\varphi} \,\mathrm{d}U_{(G_n,W_n)} \to \int \widehat{\varphi} \,\mathrm{d}\rho$, i.e. $\frac{1}{|V_n|} \sum_{x \in V_n} \widehat{\varphi}([G_n, x, W_n]) \to \int \widehat{\varphi}([G, o, W]) \,\mathrm{d}\rho([G, o, W])$. Since $\widehat{\varphi}([G_n, x, W_n]) = \langle \delta_x, \varphi(H_{(G_n,W_n)})\delta_x\rangle$, the assertion follows. □

Bibliography

[1] Miklós Abért, Andreas Thom, and Bálint Virág. Benjamini-schramm convergence and pointwise convergence of the spectral measure. *available on author's homepage.*

[2] Michael Aizenman and Simone Warzel. The canopy graph and level statistics for random operators on trees. *Math. Phys. Anal. Geom.*, 9(4):291–333 (2007), 2006.

[3] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12:no. 54, 1454–1508, 2007.

[4] Nalini Anantharaman. Quantum ergodicity on regular graphs. *Comm. Math. Phys.*, 353(2):633–690, 2017.

[5] Nalini Anantharaman and Etienne Le Masson. Quantum ergodicity on large regular graphs. *Duke Math. J.*, 164(4):723–765, 2015.

[6] Nalini Anantharaman and Mostafa Sabri. Quantum ergodicity for the Anderson model on regular graphs. *J. Math. Phys.*, 58(9):091901, 10, 2017.

[7] Nalini Anantharaman and Mostafa Sabri. Quantum ergodicity on graphs: From spectral to spatial delocalization. *Ann. of Math. (2)*, 189(3):753–835, 2019.

[8] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[9] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.

[10] Zhidong Bai, Xiaoying Wang, and Wang Zhou. CLT for linear spectral statistics of Wigner matrices. *Electron. J. Probab.*, 14:no. 83, 2391–2417, 2009.

[11] G. Ben Arous and A. Guionnet. Wigner matrices. In *The Oxford handbook of random matrix theory*, pages 433–451. Oxford Univ. Press, Oxford, 2011.

[12] Itai Benjamini. *Coarse geometry and randomness*, volume 2100 of *Lecture Notes in Mathematics*. Springer, Cham, 2013. Lecture notes from the 41st Probability Summer School held in Saint-Flour, 2011, Chapter 5 is due to Nicolas Curien, Chapter 12 was written by Ariel Yadin, and Chapter 13 is joint work with Gady Kozma, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[13] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13, 2001.

[14] Rabi Bhattacharya and Edward C. Waymire. *A basic course in probability theory.* Universitext. Springer, second edition, 2016.

[15] Charles Bordenave. Lecture notes on random graphs and probabilistic combinatorial optimization. *available on author's homepage.*

[16] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probab. Surv.*, 9:1–89, 2012.

[17] Rick Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010.

[18] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.

[19] Pete Gabor. Probability and geometry on groups. *available on author's homepage.*

[20] V. L. Girko. The circular law. *Teor. Veroyatnost. i Primenen.*, 29(4):669–679, 1984.

[21] P. R. Halmos. What does the spectral theorem say? *Amer. Math. Monthly*, 70:241–247, 1963.

[22] A. Ya. Helemskii. *Lectures and exercises on functional analysis*, volume 233 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2006. Translated from the 2004 Russian original by S. Akbarov.

[23] Richard Isaac. *The pleasures of probability*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1995. Readings in Mathematics.

[24] Tosio Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[25] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.

[26] Erwin Kreyszig. *Introductory functional analysis with applications*. John Wiley & Sons, New York-London-Sydney, 1978.

[27] Michael Reed and Barry Simon. *Methods of modern mathematical physics. I*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.

[28] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[29] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[30] Mostafa Sabri. Functional analysis II. *available on my webpage (CU scholars or french page).*

[31] Justin Salez. Some implications of local weak convergence for sparse random graphs. *PhD Thesis, Hal Id: tel-00637130.*

[32] Terence Tao and Van Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[33] Gerald Teschl. *Mathematical methods in quantum mechanics*, volume 157 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2014. With applications to Schrödinger operators.

[34] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.