ADDITIVE COMBINATORICS METHODS IN FRACTAL GEOMETRY

PÉTER P. VARJÚ

Abstract. These are notes for a set of three lectures I gave at CIRM in Luminy as part of a minicourse in the Dynamics Beyond Uniform Hyperbolicity conference between 13–24 May 2019. These lectures are closely related to but independent of the first three lectures of the minicourse given by Pablo Shmerkin. Video recordings of the lectures are also available on the CIRM website.

We do not present the full details of proofs, which may be found in the original papers, instead, we aim to indicate the main ideas and hide the technicalities.

If you have any comments, please write to pv270@dpmms.cam.ac.uk.

1. Lecture 4

1.1. Bernoulli convolutions.

This set of lectures are devoted to Bernoulli convolutions. We recall the definition. Fix a number $\lambda \in (0,1)$. The Bernoulli convolution ν_λ with parameter λ is the probability measure on $\mathbf{R}$ that is the law of the random variable

$$\sum_{n=0}^{\infty} \pm\lambda^n,$$

where $\pm$ are independent unbiased random variables.

The aim of these lectures is to discuss recent developments in the dimension theory of Bernoulli convolutions, which led to the following result.

Theorem 1. *Let $\lambda \in (0,1)$ be a number that is not the root of a polynomial with $0, \pm 1$ coefficients. Then*

$$\dim \nu_\lambda = \min(\log 2/\log \lambda^{-1}, 1).$$

For $\lambda < 1/2$, this result is folklore. In that case, ν_λ is the Cantor-Lebesgue measure on a Cantor set and it has dimension $\log 2/\log\lambda^{-1}$. The case $\lambda \in [1/2,1)\cap\overline{\mathbf{Q}}$, that is when λ is algebraic, is due to Hochman [9]. The case $\lambda \in [1/2,1)\backslash\overline{\mathbf{Q}}$, that is when λ is transcendental is due to Varjú [17], building on previous work by several mathematicians including Breuillard, Garsia, Hochman, Mignotte and Solomyak.

1.2. **Exact dimensional measures.** The notion of dimension of measures, which we use in Theorem 1, differs from that used by Shmerkin in his part of the mini-course. We say that a measure μ on $\mathbf{R}$ is exact dimensional, if

$$\dim \mu := \lim_{r\to 0} \frac{\log \mu[x-r,x+r]}{\log r} \tag{1}$$

exists and is constant μ-almost everywhere. We use (1) as the definition of dimension of measures throughout these lectures.

It was proved by Feng and Hu [5] that all self-similar measures (to be defined below), and hence Bernoulli convolution, in particular are exact dimensional.

1.3. **Self-similar measures.** We observe that Bernoulli convolutions satisfy the identity

$$\nu_\lambda = \frac{1}{2}\varphi_1(\nu_\lambda) + \frac{1}{2}\varphi_{-1}(\nu_\lambda),$$

where $\varphi_j : \mathbf{R} \to \mathbf{R}$ is defined by $\varphi_j(x) = \lambda x + j$. In fact, ν_λ is the unique probability measure on $\mathbf{R}$ satisfying this identity, so this yields an alternative way to define Bernoulli convolutions.

Using this property, we can put Bernoulli convolutions in a general framework. Let Λ be a finite set and let $\{\psi_j : j \in \Lambda\}$ be a finite collection of contractive similarities on $\mathbf{R}^d$. Furthermore, let $\{p_j : j \in \Lambda\}$ be a probability vector. Then there is a unique probability measure μ on $\mathbf{R}^d$ that satisfy the identity

$$\mu = \sum_{j\in\Lambda} p_j \psi_j(\mu).$$

Feng and Hu proved that self-similar measures are exact dimensional and we are interested in computing their dimension. There is a folklore conjecture about this, which we state below for $d = 1$, but we first need a definition.

Definition 2. *We say that the iterated function system $\{\psi_j : j \in \Lambda\}$ has exact overlaps, if there is n and*

$$(i_1, \ldots, i_n) \neq (j_1, \ldots, j_n) \in \Lambda^n$$

such that

$$\psi_{i_1} \circ \ldots \circ \psi_{i_n} = \psi_{j_1} \circ \ldots \circ \psi_{j_n}$$

Conjecture 3. *Let μ be a self-similar measure on $\mathbf{R}$ defined by an iterated function system $\{\psi_j : j \in \Lambda\}$ that does not have exact overlaps.*

Then

$$\dim \mu = \min \Big(\frac{\sum p_j \log p_j^{-1}}{\sum p_j \log r_j^{-1}}, 1 \Big), \tag{2}$$

where r_j is the contraction factor of ψ_j.

Note that the right hand side of (2) depends on the iterated function system. The same self-similar measure could be realized by different iterated function systems. Indeed, we may replace Λ by Λ^n and the maps by all n-fold compositions. If there are exact overlaps, we may remove repeated occurrences of the same map and get the same self-similar measure if we adjust the weights p_j appropriately. This has the effect of reducing the value of the right hand side of (2). It is possible to state a similar conjecture for self-similar measures in $\mathbf{R}^d$, but it requires some technical conditions to avoid "too much" of the measure being "trapped" in lower dimensional subspaces. This is subtle and technical and we will not discuss it further.

We end this section by discussing the exact overlaps property for Bernoulli convolutions. We first note

$$\varphi_{i_0} \circ \varphi_{i_1} \circ \ldots \circ \varphi_{i_n}(x) = \lambda^n x + (i_n \lambda^n + \ldots + i_1 \lambda + i_0).$$

Therefore, we can write

$$\varphi_{i_0} \circ \ldots \circ \varphi_{i_n}(x) - \varphi_{j_0} \circ \ldots \circ \varphi_{j_n}(x) = 2P(\lambda),$$

where P is a polynomial of degree at most n with coefficients $\pm 1, 0$. Since this family of polynomials will play a special role in this lectures, we introduce the notation $\mathcal{P}_n$ for them. This means that the Bernoulli convolution ν_λ has exact overlaps if and only if $P(\lambda) = 0$ for some $P \in \bigcup \mathcal{P}_n$. Therefore, Theorem 1 is an instance of the above conjecture.

1.4. **Outline of the proof.** We turn to the proof of Theorem 1. We first set out the main parts of the proof; we will discuss some of these in detail in the next two lectures. These developments started with the following result of Hochman [9, Theorem 1.9].

Theorem 4 (Hochman)**.** *Let $\lambda \in [1/2, 1)$ be such that $\dim \nu_\lambda < 1$. Then for all C and for all sufficiently large $n > N(C)$, there is a number $\eta_n \in \mathbf{C}$ such that $P(\eta_n) = 0$ for some $P \in \mathcal{P}_n$ and $|\lambda - \eta_n| < C^{-n}$.*

Informally, this result shows that all parameters λ for which the conclusion of Theorem 1 fails are extremely well approximated by parameters with exact overlaps at all scales.

Theorem 4 already implies Theorem 1 for algebraic parameters. This follows simply from the fact that algebraic numbers repel each other and hence the conclusion of Theorem 4 may hold for an algebraic number λ only if $\lambda = \eta_n$. For example, the following result can be used to justify this.

Theorem 5 (Mignotte [13])**.** *Let η be an algebraic number of degree n and let $n' > n(\log n)^2$ be another integer. Let $\eta' \neq \eta$ be the root of a polynomial in $\mathcal{P}_{n'}$. Then*

$$|\eta' - \eta| > (CM(\eta))^{-2n'},$$

where C is an absolute constant and $M(\lambda)$ is the Mahler measure of λ (to be defined below).

Mignotte's theorem is more general than this, we just stated it in the form we will use it. This theorem is much stronger than what we need to deduce Theorem 1 for algebraic parameters from Theorem 4. Indeed, it would be enough for us to have $|\eta' - \eta| > C^{-n'}$ for any $C = C(\eta)$ and such a statement is significantly easier to prove. However, we will need the full force of Theorem 5 shortly.

In the paper [17], Theorem 5 was substituted by an alternative result (see [17, Proposition 12]), which was derived from earlier results of Garsia [6] and Solomyak [15].

We still owe the reader the definition of Mahler measure.

Definition 6. *Let η be an algebraic number with minimal polynomial $a_n \prod_{j=1}^{n}(x - \eta_j) \in \mathbf{Z}[x]$, that is, a_n is the leading coefficient of the minimal polynomial and $\eta_1, \ldots, \eta_n$ are its roots; one of these numbers equal η. Then the Mahler measure of η is defined as*

$$M(\eta) = |a_n| \prod_{j=1}^{n} \max(1, |\eta_j|).$$

It was observed by Hochman that Theorem 4 already implies all of Theorem 1 if the answer to the following question posed by him is affirmative.

Question 7 (Hochman)**.** *Is there a universal constant C such that $|\eta - \eta'| > C^{-n}$ for all numbers $\eta \neq \eta'$ that are roots of polynomials in $\mathcal{P}_n$?*

Note that this question asks for a strengthening of Theorem 5 in two ways. First, the constant in the separation bound needs to be universal, it is not allowed to depend on the Mahler measure of η. Second, the separation is also required for numbers that have (approximately) the same degree.

Here is the argument of Hochman to deduce Theorem 1 from his Theorem 4 assuming the answer to the above question is affirmative.

Suppose that $\lambda \in [1/2, 1)$ is such that $\dim \nu_\lambda < 1$. By Theorem 4, for all n sufficiently large, there is η_n that is a root of a polynomial in $\mathcal{P}_n$ and $|\lambda - \eta_n| < C^{-n}$. This means that $|\eta_n - \eta_{n+1}| < 2C^{-n}$. Now an affirmative answer to the above question yields $\eta_n = \eta_{n+1}$ provided we set the constant in Theorem 4 suitably. Now we see that the sequence η_n stabilizes and hence its limit λ is a root of a polynomial in $\bigcup \mathcal{P}_n$.

An affirmative answer to Hochman's question seems plausible. However, we do not know how to prove this. On the other hand, the above argument can still be carried out using Mignotte's Theorem 5 instead together with the following result.

Theorem 8 (Breuillard, Varjú [4, Theorem 1])**.** *Let* $\lambda \in [1/2, 1)$ *be such that* $\dim \nu_\lambda < 1$. *Then for any* $\varepsilon > 0$, *there are infinitely many integers* n, *and numbers* $\eta_n \in \mathbf{R}$ *such that* $P(\eta_n) = 0$ *for some* $P \in \mathcal{P}_n$, $|\lambda - \eta_n| < \exp(-n^{100})$ *and* $\dim \nu_{\eta_n} < \dim \nu_\lambda + \varepsilon$.

The theorem holds with any number in place of 100, in fact, one may even take a slowly diverging sequence of exponents as a function of n.

Now we may conclude the proof of Theorem 1. Assume to the contrary that there is $\lambda \in [1/2, 1) \backslash \overline{Q}$ such that $\dim \nu_\lambda < 1$. We first apply Theorem 8 to find η of degree at most n such that $|\lambda - \eta| < \exp(-n^{100})$ and $\dim \eta \leq \dim \nu_\lambda + \varepsilon$. Now we choose an integer n' such that $C^{-n'} < |\lambda - \eta| \leq C^{-(n'-1)}$ with an appropriately chosen constant C. Observe that the condition $n' > n(\log n)^2$ of Mignotte's theorem 5 holds. Now we apply Theorem 4 and find an η' that is a root of a polynomial in $\mathcal{P}_{n'}$ and $|\lambda - \eta'| < C^{-n'}$. We chose n' just large enough to guarantee that $\eta \neq \eta'$ but we still have

$$|\eta - \eta'| \leq |\lambda - \eta| + |\lambda - \eta'| < C^{-n'} + C^{-(n'-1)}.$$

This contradicts Mignotte's Theorem 5 if we choose the constants appropriately and if we can get a suitable bound on $M(\eta)$. We will discuss in Lecture 6 how to control the Mahler measure using the information $\dim \eta \leq \dim \nu_\lambda + \varepsilon$.

Observe that the success of the above argument crucially depends on three features of Theorems 4 and 8. First, the approximation in Theorem 4 holds at all sufficiently smalls scales, so we can choose n' in the argument as we want. Second, the approximation in Theorem 8 holds with very good precision and this ensures that the n' is much larger than n. Third, in Theorem 8, we have a bound on the dimension of the Bernoulli convolution for the approximating parameter, which allows us to bound the Mahler measure (as we will discuss it in Lecture 6).

1.5. **Plan for the remaining two lectures.** We will discuss some details of the proof of Theorem 4 and Theorem 8 in Lecture 5. We will discuss the dimension of Bernoulli convolutions for algebraic parameters in Lecture 6, in particular, we will explain the connection to Mahler measure.

2. Lecture 5

We discuss the proof of the following theorem of Hochman in this lecture.

Theorem 9 (Hochman). *Let $\lambda \in [1/2, 1)$ be such that $\dim \nu_\lambda < 1$. Then for all C and for all sufficiently large $n > N(C)$, there is a number $\eta_n \in \mathbf{C}$ such that $P(\eta_n) = 0$ for some $P \in \mathcal{P}_n$ and $|\lambda - \eta_n| < C^{-n}$.*

The proof we present is essentially the same as Hochman's original, but some technical aspects have been influenced by the paper [4].

2.1. **Entropy dimension.** Let μ be a compactly supported probability measure on $\mathbf{R}$ and let X be a random variable with distribution μ. We keep this notation until the end of this lecture.

The entropy dimension of μ is defined as

$$\dim_{\mathrm{ent}} \mu = \lim_{r\to 0} \frac{H(\lfloor r^{-1} X \rfloor)}{\log r^{-1}}$$

provided the limit exists, where $H(\cdot)$ stands for the Shannon entropy of a discrete random variable. The function $x \mapsto \lfloor r^{-1} x \rfloor$ is constant on intervals of length r, so the above formula involves the entropy of μ with respect to a partition of $\mathbf{R}$ to intervals of length r.

We leave it as an exercise to show that for exact dimensional measures, the limit in the definition of entropy dimension exists and equals to its dimension as defined in the previous lecture. In the proof of Theorem 9 we will work with this notion of dimension.

2.2. **Entropy of measures at scales.** The quantity $H(\lfloor r^{-1} X \rfloor)$ that appears in the definition of entropy dimension has a significant drawback; it is not translation invariant. A simple way to fix this issue is to average it over translations. This turns out to be very useful.

Let μ be a compactly supported probability measure on $\mathbf{R}$ and let X be a random variable with distribution μ. The entropy of μ and X at scale r is defined as

$$H(\mu; r) = H(X; r) = \int_0^1 H(\lfloor r^{-1} X + t \rfloor) dt.$$

We also define the conditional entropy between two scales by

$$H(\mu; r_1|r_2) = H(X; r_1|r_2) = H(X; r_1) - H(X; r_2).$$

If the ratio of the scales r_2/r_1 is an integer, we can realize this as the average of conditional Shannon entropies

$$H(X; r_1|r_2) = \int_0^{r_2} H(\lfloor r_1^{-1}(X+t)\rfloor | \lfloor r_2^{-1}(X+t)\rfloor) dt.$$

Indeed, in this case, $\lfloor r_2^{-1}(X+t)\rfloor$ is a function of $\lfloor r_1^{-1}(X+t)\rfloor$, so the conditional entropy is just the difference of the entropies. We leave the verification of the details to the interested reader.

A similar averaging procedure was employed by Wang in [20] in his study of quantitative density of orbits of some groups of toral automorphisms.

2.3. **Properties of entropy.** In this section we record some facts without proof about entropies at scales, which we will use later. For proofs and further details see [19, Section 2].

(1) *Translation invariance.* We have

$$H(X; r) = H(X+t; r), \qquad H(X; r_1|r_2) = H(X+t; r_1|r_2).$$

(2) *Concavity.* Let $\mu_1, \ldots, \mu_k$ be probability measures and let $p_1, \ldots, p_k$ be a probability vector. Write $\mu = p_1\mu_1 + \ldots + p_k\mu_k$. Then

$$H(\mu; r) \geq p_1 H(\mu_1; r) + \ldots + p_k H(\mu_k; r)$$

and

$$H(\mu; r_1|r_2) \geq p_1 H(\mu_1; r_1|r_2) + \ldots + p_k H(\mu_k; r_1|r_2)$$

provided $r_2/r_1 \in \mathbf{Z}$. This last inequality does not hold in general if the ratio of the scales is not an integer.

(3) *Convolution may only increase entropy.* As a corollary of the above two items, we get

$$H(\mu * \nu; r) \geq H(\mu; r)$$

and

$$H(\mu * \nu; r_1|r_2) \geq H(\mu; r_1|r_2)$$

provided $r_2/r_1 \in \mathbf{Z}$. Again, the integrality of the ratio of the scales is important.

(4) *Scaling.* We have

$$H(X; r) = H(sX; sr), \qquad H(X; r_1|r_2) = H(sX; sr_1|sr_2).$$

(5) *Continuity.* The function $\rho \mapsto H(\mu; \exp(-\rho))$ is a monotone increasing Lipschitz function, moreover, we have

$$0 \le H(\mu; r_1|r_2) \le \min(2 \log r_2/r_1, \log r_2/r_1 + \log 2).$$

If $r_2/r_1 \in \mathbf{Z}$ we even have

$$H(\mu; r_1|r_2) \le \log r_2/r_1.$$

2.4. **Proof of Theorem 9.** We prove Theorem 9 by contradiction. Suppose that $\dim \nu_\lambda < 1$ for some $\lambda \in [1/2, 1)$. Fix C and let n be a sufficiently large integer. Suppose to the contrary that there is no $\eta \in \mathbf{C}$ that is a root of a polynomial in $\mathcal{P}_n$ and $|\eta - \lambda| < C^{-n}$.

Fix $\varepsilon > 0$ small enough. As we remarked above, the dimension of ν_λ equals its entropy dimension, hence we have

$$\Big| \frac{H(\nu_\lambda; \lambda^l|1)}{l \log \lambda^{-1}} - \dim \nu_\lambda \Big| < \varepsilon \tag{3}$$

provided l is large enough. This means, in particular, that

$$H(\nu_\lambda; \lambda^n|1) > (\dim \nu_\lambda - \varepsilon) n \log \lambda$$

and if we are able to show that

$$H(\nu_\lambda; \lambda^m|\lambda^n) > (\dim \nu_\lambda + 3\varepsilon)(m - n) \log \lambda \tag{4}$$

for some $m \ge 2n$, then we get

$$H(\nu_\lambda; \lambda^m|1) > (\dim \nu_\lambda + \varepsilon) m \log \lambda,$$

which contradicts (3).

To prove (4), we decompose ν_λ as a convolution of measures in a suitable way. For $I \subset (0, 1]$, we write ν_λ^I for the law of the random variable

$$\sum_{j:\lambda^j \in I} \pm \lambda^j.$$

With this notation, we have $\nu_\lambda = \nu_\lambda^{(0,1]}$ and we have $\nu_\lambda^{I \cup J} = \nu_\lambda^I * \mu_\lambda^J$ for two disjoint sets $I, J \subset (0, 1]$.

Now we observe that $\nu_\lambda = \nu_\lambda^{(0,\lambda^n]} * \nu_\lambda^{(\lambda^n,1]}$. Using the scaling property of entropy and (3), we can write

$$H(\nu_\lambda^{(0,\lambda^n]}; \lambda^m|\lambda^n) = H(\nu_\lambda; \lambda^{m-n}|1) > (\dim \nu_\lambda - \varepsilon)(m - n) \log \lambda. \tag{5}$$

We only need a very small improvement over this to achieve (4).

We have already discussed that entropy cannot decrease if we take convolution of measures. (Strictly speaking this is true only if the ratio of scales is integral, but it still holds with a small error in general.) Now we want to argue that under suitable conditions, entropy increases if we take a convolution. This would enable us to improve the bound in (5) to (4) when we convolve the measure $\nu_\lambda^{(0,\lambda^n]}$ with $\nu_\lambda^{(\lambda^n,1]}$ to get ν_λ.

We are going to use the following result.

Theorem 10 (Varjú [19, Theorem 3]). *For every $\alpha \in (0, 1/2]$, there is a number $c > 0$ such that the following holds. Let μ and ν be two bounded probability measures on $\mathbf{R}$. Let $0 < s_1 < s_2$ and $\beta \in (0, 1/2]$ be numbers. Suppose that*

$$H(\mu; s|2s) \leq \log 2 - \alpha$$

for all $s \in \mathbf{R}_{>0}$ and

$$H(\nu; s_1|s_2) \geq \beta \log(s_2/s_1).$$

Then

$$H(\mu * \nu; s_1|s_2) \geq H(\mu; s_1|s_2) + c\beta(\log \beta^{-1})^{-1} \log s_2/s_1 - c^{-1}.$$

We apply this theorem with $\mu = \nu_\lambda^{(0,\lambda^n]}$, $\nu = \nu_\lambda^{(\lambda^n,1]}$, $s_1 = \lambda^m$ (with a suitable choice of m) and $s_2 = \lambda^n$. If we can show that the conditions of the theorem hold with some $\alpha > 0$ and $\beta > 0$ independently of n, then we can achieve (4) by choosing some

$$\varepsilon < c\beta(\log \beta^{-1})^{-1}/4.$$

The first condition of the theorem follows from the following.

Lemma 11. *Let $\lambda \in [1/2, 1)$ be such that $\dim \nu_\lambda < 1$. Then there is a number $\alpha > 0$ such that*

$$H(\nu_\lambda^I; r|2r) < \log 2 - \alpha$$

for all $I \subset (0, 1]$.

We first observe that it is enough to prove this lemma for $I = (0, 1]$, because $\nu_\lambda = \nu_\lambda^I * \nu_\lambda^{(0,1]\setminus I}$ and convolution may only increase entropy between scales of integral ratio. For the rest of the proof we refer to [4, Lemma 13]. We just mention here the following observation, which plays an important role in the proof. Using the scaling property of entropies and the fact that convolution may only increase entropy we can write

$$H(\nu_\lambda; r|2r) = H(\nu_\lambda^{(0,\lambda]}; \lambda r|2\lambda r) \leq H(\nu_\lambda; \lambda r|2\lambda r).$$

If $H(\nu_\lambda; r|2r)$ is very close to $\log 2$ for some scale r, then we can use the above observation to produce many more scales with this property. Together with some additional ideas this can be used to show that the entropy dimension of ν_λ is very close to 1.

The second condition in Theorem 10 can be verified using the hypothesis that there is no $\eta \in \mathbf{C}$ that is a root of a polynomial in $\mathcal{P}_n$ and

$|\eta - \lambda| < C^{-n}$. We first observe that there is a number $\widetilde{C}$ depending only on λ and C such that $|P(\lambda)| > \widetilde{C}^{-n}$ for all $0 \neq P \in \mathcal{P}_n$. Indeed, write

$$|P(\lambda)| = \prod_{j=1}^{n} |\lambda - \eta_j|,$$

where η_j are the roots of P. By the hypothesis, $|\lambda - \eta_j| > C^{-n}$ for all j. It follows from Jensen's formula (see e.g. [4, Lemma 26] for details) that there are at most K roots of P of absolute value less than $1 - (1 - \lambda)/2$ for some K depending only on λ. This shows

$$|P(\lambda)| > C^{-Kn}\Big(\frac{1-\lambda}{2}\Big)^n,$$

so the claim holds indeed with $\widetilde{C} = C^K(1 - \lambda)/2$.

Now we choose m so that $\lambda^m < \widetilde{C}^{-n}$ and observe that the 2^n points in the support of $\nu_\lambda^{(\lambda^n,1]}$ are separated by a distance of at least λ^m pairwise. This means that $H(\lfloor \lambda^{-m} Y + t \rfloor) = n \log 2$ for all t and hence

$$H(\nu_\lambda^{(\lambda^n,1]}; \lambda^m) = n \log 2.$$

Now we can write

$$\begin{aligned} H(\nu_\lambda^{(\lambda^n,1]}; \lambda^m|\lambda^n) &\geq H(\nu_\lambda^{(\lambda^n,1]}; \lambda^m) - H(\nu_\lambda^{(\lambda^n,1]}; \lambda^n|1) - H(\nu_\lambda^{(\lambda^n,1]}; 1) \\ &\geq n \log 2 - n \log \lambda^{-1} - \log 2 - H(\nu_\lambda; 1). \end{aligned}$$

From this we see that the second condition in Theorem 10 holds indeed for $\nu_\lambda^{(\lambda^n,1]}$ with some β that depends only on λ and the constant C in the indirect hypothesis.

2.5. **Concluding remarks.** Theorem 10 appeared in [19] after Hochman's work [9]. Hochman formulated and proved a similar result for estimating the entropy of convolutions. His version does not give an explicit estimate for the entropy gain in terms of the parameters of the problem. As it is clear form the above discussion, such explicit estimates are not needed for the application in the proof of Theorem 9. However, these are absolutely critical in the proof of Theorem 8, which we mentioned in Lecture 4. Hochman also did not used the averaging procedure under translations.

The proof of Theorem 8 is beyond the scope of these notes, and we only make a few comments about it. The proof is again by contradiction, so we assume that $\dim \nu_\lambda < 1$ for some $\lambda \in [1/2, 1)$, but the conclusion of Theorem 8 does not hold for λ. Using these assumptions, we find a disjoint collection of sets $I_1, \ldots, I_k \subset (0, 1]$ such that we can obtain some lower bound on $H(\nu_\lambda^{I_j}; s_1|s_2)$ for some appropriately chosen scales $s_1 < s_2$. Now we apply Theorem 10 repeatedly to obtain an even stronger lower bound on

$$H(\nu_\lambda^{I_1\cup\ldots\cup I_k}; s_1|s_2),$$

which will contradict the universal upper bound $\log(s_2/s_1) + \log 2$. In these applications of Theorem 10, we always have a good control on α, it depends only on λ. However, the parameter β may take very small values depending on how well λ is approximated by roots of polynomials in $\mathcal{P}_n$ and for which n's. Therefore, the quantitative aspects of Theorem 10 are very crucial. For more details, we refer to the original paper [4].

3. Lecture 6

The aim of this lecture is to discuss the connection between $\dim\nu_\lambda$ and Mahler measure $M(\lambda)$ for algebraic parameters. More specifically, we want to get an upper bound for $M(\lambda)$ in terms of $1-\dim\nu_\lambda$ for $\lambda \in [1/2, 1) \cap \overline{\mathbf{Q}}$.

3.1. **Dimension of Bernoulli convolutions for algebraic parameters and Garsia entropy.** Beside Theorem 4, which we discussed in the previous lecture, Hochman's results [9] also imply a formula for $\dim\nu_\lambda$ for algebraic parameters. This formula is in terms of a quantity, which was introduced by Garsia [7].

Definition 12. *Let $\lambda \in \mathbf{C}$. The Garsia entropy of λ is the number*

$$h(\lambda) = \dim_{n\to\infty} \frac{1}{n} H\Big(\sum_{j=0}^{n-1} \pm\lambda^j\Big),$$

where $H(\cdot)$ is the Shannon entropy of a discrete random variable.

Theorem 13 (Hochman)**.** *Let $\lambda \in [1/2, 1) \cap \overline{\mathbf{Q}}$. Then*

$$\dim\nu_\lambda = \min(h(\lambda)/\log\lambda^{-1}, 1).$$

This result is not stated explicitly in [9], but it is an immediate consequence of the main result of that paper. See [3, Section 3.4] for the details.

As we discussed in Lecture 4, the iterated function system underlying ν_λ has exact overlaps when λ is the root of a polynomial with $\pm1, 0$ coefficients. In that case, ν_λ can be realized as a self-similar measure defined by other iterated function systems, such that the right hand of (2) is smaller than it is in the case of the standard iterated function system $x \mapsto \lambda x \pm 1$. It is not too difficult to see that the infimum of the right hand side of (2) over all iterated function systems realizing ν_λ is precisely the formula for $\dim\nu_\lambda$ in Theorem 13 above.

3.2. **Garsia entropy and Mahler measure.** Theorem 13 reduces the problem of computing the dimension of Bernoulli convolutions for algebraic parameters to computing the quantity $h(\lambda)$. There has been much work on this issue recently. Akiyama, Feng, Kempton and Persson [1] gave a general algorithm to compute $h(\lambda)$ to arbitrary precision for given λ. Feng and Feng developed further algorithms and carried out extensive numerical calculations. Here are a few sample results they obtained. They showed that $\dim\nu_\lambda \geq 0.980368$ for all $\lambda\in[1/2,1)$ and the minimum of $\dim\nu_\lambda$ for $\lambda\in[1/2,1)$ is obtained in the interval $\lambda_0\pm 10^{-4}$, where $\lambda_0\approx 0.5283$ is a root of the polynomial x^3+x^2+x-1. They also computed $\dim\nu_{\lambda_0} = 0.98040931953\pm 10^{-11}$.

Breuillard and Varjú gave a general estimate for $h(\lambda)$ in terms of the Mahler measure.

Theorem 14 (Breuillard, Varjú[3])**.** *Let λ be an algebraic number. Then*

$$0.4\min(\log M(\lambda),\log 2)\leq h(\lambda)\leq \min(\log M(\lambda),\log 2). \tag{6}$$

Moreover, for every $\varepsilon>0$, there is C such that $h(\lambda)<\log 2-\varepsilon$ implies $M(\lambda)<C$.

Combining this result with Theorem 13 we get the following. For every $\varepsilon>0$, there is C such that $\dim\nu_\lambda<1-\varepsilon$ implies $M(\lambda)<C$ for all $\lambda\in[1/2,1)\cap\overline{Q}$. This is precisely the result that is required to complete the argument in Lecture 4.

The upper bound in (6) is easy to prove. By simple number theoretical considerations it follows that the random variable $\sum_{j=0}^{n-1}\pm\lambda^j$ takes at most $C_\varepsilon M(\lambda)^{n+\varepsilon}$ many different values, and it certainly cannot take more than 2^n. Now the upper bound follows from the fact that the Shannon entropy of a random variable is at most the logarithm of the number of values taken by the random variable.

The purpose of the rest of this lecture is to explain the proof of the lower bound in (6) and the proof of the last claim, which was not stated in [3], but it follows immediately from the same proof.

3.3. **Entropy at scales with Gaussian smoothing.** We introduce a quantity that will play a prominent role in our proof.

Let μ be a compactly supported probability measure on $\mathbf{R}^d$ and let X be a random vector with law μ. Let $A\in\mathrm{GL}_d(\mathbf{R})$. In this lecture, we define the entropy of μ and X at scale A to be

$$H(\mu;A)=H(X;A)=H(X+AG)-H(AG),$$

where G is a standard Gaussian that is independent of X and $H(\cdot)$ is the differential entropy of an absolute continuous random variable. That is, if f denotes the density of $X + AG$, then we have

$$H(\mu; A) = \int_{\mathbf{R}^d} f(x) \log f(x)^{-1} dx.$$

Informally one may think of $H(\mu; A)$ as the entropy of μ with respect to a partition whose atoms are translates of the image of the unit ball under A.

It is instructive to compare this quantity with what we used in the previous lecture. It is easy to verify that with the notation of the previous lecture

$$H(X; s) = H(X + sI) - H(sI),$$

where I is a uniform random variable on $[0, 1]$ that is independent of X and $H(\cdot)$ is differential entropy. Therefore, the difference between the quantity we use now compared with that of the previous lecture is that now we use Gaussian smoothing as opposed to a uniform distribution on an interval. The properties of Gaussian smoothing will be critical to the success of our argument, because we require estimates that are independent of dimension and the Gaussian distribution behaves very well in this regard.

The conditional entropy of μ and X between the scales $A_1, A_2 \in \mathrm{GL}_d(\mathbf{R})$ is defined as

$$H(\mu; A_1|A_2) = H(X; A_1|A_2) = H(X; A_1) - H(X; A_2).$$

We make a final comment on our notation. The expression $H(\cdot)$ sometimes stands for Shannon and sometimes for differential entropy depending on whether the random variable is discrete or absolutely continuous. This should not cause confusion as the type of the random variable will be always clear from the context.

3.4. **Properties of entropy.** We record a few key properties of entropy at a scale defined above. The proofs may be found in [3, Section 2].

(1) If X is a discrete random vector, we have

$$0 \le H(X|A) \le H(X),$$

where $H(\cdot)$ is Shannon entropy on the right hand side.

(2) Let X_1 and X_2 be independent random vectors and let $A_1, A_2 \in \mathrm{GL}_d(\mathbf{R})$ such that $\|A_1^T x\| \le \|A_2^T x\|$ for all $x \in \mathbf{R}^d$, then

$$0 \le H(X_1; A_1|A_2) \le H(X_1+X_2; A_1|A_2) \le H(X_1; A_1|A_2)+H(X_2; A_1|A_2).$$

The condition on A_1 and A_2 may be equivalently reformulated as the image of the unite ball under A_1 is contained in its image under A_2. This property implies that A_2G can be realized as the sum of two independent random vectors one of which has the same distribution as A_1G, and this plays a key role in the proof.

(3) *Scaling.* For any $B \in \mathrm{GL}_d(\mathbf{R})$, we have

$$H(BX; BA) = H(X; A), \qquad H(BX; BA_1|BA_2) = H(X; A_1|A_2).$$

3.5. **Proof of Theorem 14.** We follow [3]. Let $A \in \mathrm{GL}_d(\mathbf{R})$ be a diagonalizable matrix whose eigenvalues are the Galois conjugates of λ with modulus less than 1 and let $x \in \mathbf{R}^d \backslash \{0\}$. We write

$$X_{A,x}^{(m,n)} = \sum_{i=m}^{n} \pm A^i x.$$

Observe that

$$H(X_{A,x}^{0,n-1}) = H\Big(\sum_{i=0}^{n-1} \pm\lambda^i\Big).$$

To see this, we may expand $X_{A,x}^{0,n-1}$ in an eigenbasis of A. Any coordinate in this basis will be a constant multiple of a Galois conjugate of $\sum_{i=0}^{n-1} \pm\lambda^i$ and hence the law of $X_{A,x}^{0,n-1}$ and $\sum_{i=0}^{n-1} \pm\lambda^i$ yield the same probability vector.

Lemma 15. *If* $\|A\| \le 1$, *then*

$$h_\lambda \ge \lim_{l\to\infty} \frac{1}{l} H(X_{A,x}^{(0,\infty)}; A^l | \,\mathrm{Id}).$$

The statement of this lemma is not surprising. Intuitively, we expect that $H(X_{A,x}^{(0,\infty)}; A^l|\,\mathrm{Id})$ is close to $H(X_{A,x}^{(0,l-1)}; A^l|\,\mathrm{Id})$, because the terms $\pm A^j x$ for $j \ge l$ should not change significantly entropy at scale A^l. On the other hand $H(X_{A,x}^{(0,l-1)}; A^l|\,\mathrm{Id})$ is a lower bound for the Shannon entropy $H(X_{A,x}^{(0,l-1)})$, which equals to $H(\sum_{i=0}^{l-1} \pm\lambda^i)$, which appears in the definition of $h(\lambda)$. A formal proof follows.

Proof. We can write

$$\begin{aligned} H\Big(\sum_{i=0}^{l-1}\pm\lambda^i\Big) =& H(X_{A,x}^{(0,l-1)}) \\ \geq& H(X_{A,x}^{(0,l-1)};A^l) \\ =& H(X_{A,x}^{(0,l-1)};A^l)+H(A^lX_{A,x}^{(0,\infty)};A^l)-H(X_{A,x}^{(0,\infty)};\mathrm{Id}) \\ \geq& H(X_{A,x}^{(0,\infty)};A^l)-H(X_{A,x}^{(0,\infty)};\mathrm{Id}) \\ =& H(X_{A,x}^{(0,\infty)};A^l|\,\mathrm{Id}). \end{aligned}$$

Now we can complete the proof if we divide both sides by l and take the limit $l\to\infty$. □

In the next step, we reduce the estimate to a single conditional entropy by a standard argument in entropy theory.

Lemma 16. *Under the same assumptions as above, we have*

$$h(\lambda)\geq H(X_{A,x}^{(0,\infty)};A|\,\mathrm{Id}).$$

Proof. We can write

$$H(X_{A,x}^{(0,\infty)};A^l|\,\mathrm{Id})=\sum_{i=0}^{l-1}H(X_{A,x}^{(0,\infty)};A^{i+1}|A^i).$$

We observe that

$$H(X_{A,x}^{(0,\infty)};A^{i+1}|A^i)\geq H(X_{A,x}^{(i,\infty)};A^{i+1}|A^i)=H(X_{A,x}^{(0,\infty)};A|\,\mathrm{Id}),$$

so

$$\frac{1}{l}H(X_{A,x}^{(0,\infty)};A^l|\,\mathrm{Id})$$

is the average of terms all of which are at least $H(X_{A,x}^{(0,\infty)};A|\,\mathrm{Id})$, and this proves the claim in light of the previous lemma. □

The above estimates are not very wasteful, it seems very plausible that the lover bound in the last lemma can be arbitrarily close to the truth with a suitable choice of A and x. However, this estimate is not very useful, because it involves the entropy of a very complicated measure in high dimension. The next step is very drastic and potentially wasteful, but makes the problem much easier to study. We remove all but one term from the infinite sum defining $X_{A,x}^{(0,\infty)}$.

Lemma 17. *Let x and A be as above and let $U,V\in O(d)$ be arbitrary orthogonal matrices.*

$$h(\lambda)\geq H(\pm x;UAV|\,\mathrm{Id}).$$

Proof. By the previous lemma, we immediately have

$$h(\lambda) \geq H(X_{A,x}^{(0,\infty)}; A|\,\mathrm{Id}) \geq H(\pm x; A|\,\mathrm{Id}).$$

Using the scaling property of entropy, we can write

$$h(\lambda) \geq H(\pm Ux; UA|U\,\mathrm{Id}).$$

Finally we note that by the rotational symmetry of the standard Gaussian that is involved in the definition of entropy at scales, we can multiply any matrix scale from the right by an arbitrary rotation matrix without changing entropy. This and the fact that $x \in \mathbf{R}^d \backslash \{0\}$ is arbitrary completes the proof. □

Now we choose the parameters in the above lemma in a convenient way. By a theorem of Horn (see [3, Proof of Proposition 13] for the details), we can choose A in such a way that its eigenvalues are all the conjugates of λ with modulus less than 1 and its singular values are $1, 1, \ldots, 1, \prod \min(|\lambda_j|, 1)$, where λ_j runs through the Galois conjugates of λ, so the last singular value equals the absolute value of the product of the eigenvalues.

We note that we may assume that λ is an algebraic unit, that is the leading and constant coefficients of its minimal polynomial are ± 1. Indeed, if this is not the case, we have $h(\lambda) = \log 2$ always, because only algebraic units can be roots of polynomials with coefficients $\pm 1, 0$. Now it follows that

$$\prod \min(|\lambda_j|, 1) = 1/\prod \max(|\lambda_j|, 1) = M(\lambda)^{-1},$$

where the product runs over all Galois conjugates of λ.

We choose A in the above way and we choose U and V such that $UAV = \mathrm{diag}(1, \ldots, 1, M(\lambda)^{-1})$. Finally, we set x to be a constant multiple of the last coordinate vector. With this choice, the last lemma gives

$$h(\lambda) \geq H(\pm t e_d; \mathrm{diag}(1, \ldots, 1, M(\lambda^{-1}))|\,\mathrm{Id}).$$

Now we can integrate out the first $d-1$ coordinates and get the following.

Lemma 18. *We have for all $t \in \mathbf{R}_{>0}$*

$$h(\lambda) \geq H(\pm t; M(\lambda^{-1})|1).$$

The notation in the last lemma should not be confused with the notation of the previous lecture, we still use Gaussian smoothing. Now we reduced the problem to the estimation of a 1 dimensional integral, which only involves the Mahler measure of λ. The proof of Theorem 14 is now just a calculus exercise, and we refer to [3, Section 3.2] and [17, Theorem 9] for the details.

4. Further reading

Peres, Schlag and Solomyak wrote a beautiful survey of Bernoulli convolutions [14], which exposes the state of the art up until 1998. Solomyak wrote another nice survey [16]. Surveys covering the more recent developments are [18] and [8]. These also discuss new developments about absolute continuity, which we omitted completely.

The techniques of Hochman has been extended to more general settings including self-similar measures in $\mathbf{R}^d$ [10], self-affine measures [2, 11] and Furstenberg measures [12].

References

[1] S. Akiyama, D.-J. Feng, T. Kempton, and T. Persson, *On the Hausdorff dimension of Bernoulli convolutions*, 2018. arXiv:1801.07118v1. ↑12

[2] B. Bárány, M. Hochman, and A. Rapaport, *Hausdorff dimension of planar self-affine sets and measures with overlaps*, 2017. arXiv:1712.07353. ↑17

[3] E. Breuillard and P. P. Varjú, *Entropy of Bernoulli convolutions and uniform exponential growth for linear groups*, 2018. arXiv:1510.04043v3, To appear in J. Anal. Math. ↑11, 12, 13, 14, 16

[4] E. Breuillard and P. P. Varjú, *On the dimension of Bernoulli convolutions*, 2018. arXiv:1610.09154v3, To appear in Ann. Probab. ↑5, 6, 9, 10, 11

[5] D.-J. Feng and H. Hu, *Dimension theory of iterated function systems*, Comm. Pure Appl. Math. **62** (2009), no. 11, 1435–1500. MR2560042 (2010i:37049) ↑2

[6] A. M. Garsia, *Arithmetic properties of Bernoulli convolutions*, Trans. Amer. Math. Soc. **102** (1962), 409–432. MR0137961 (25 #1409) ↑4

[7] A. M. Garsia, *Entropy and singularity of infinite convolutions*, Pacific J. Math. **13** (1963), 1159–1169. MR0156945 (28 #187) ↑11

[8] S. Gouëzel, *Méthodes entropiques pour les convolutions de Bernoulli, d'après Hochman, Shmerkin, Breuillard, Varjú*, 2018. (exposé au séminaire Bourbaki 2018), `http://www.math.sciences.univ-nantes.fr/~gouezel/articles/bourbaki_bernoulli.pdf`. ↑17

[9] M. Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. of Math. (2) **180** (2014), no. 2, 773–822. MR3224722 ↑1, 3, 10, 11

[10] M. Hochman, *On self-similar sets with overlaps and inverse theorems for entropy in* $\mathbb{R}^d$, 2015. arXiv:1503.09043. ↑17

[11] M. Hochman and A. Rapaport, *Hausdorff dimension of planar self-affine sets and measures with overlaps*, 2019. arXiv:1904.09812. ↑17

[12] M. Hochman and B. Solomyak, *On the dimension of Furstenberg measure for* $SL_2(\mathbb{R})$ *random matrix products*, Invent. Math. **210** (2017), no. 3, 815–875. MR3735630 ↑17

[13] M. Mignotte, *Approximation des nombres algébriques par des nombres algébriques de grand degré*, Ann. Fac. Sci. Toulouse Math. (5) **1** (1979), no. 2, 165–170. MR554376 ↑4

[14] Y. Peres, W. Schlag, and B. Solomyak, *Sixty years of Bernoulli convolutions*, Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), 2000, pp. 39–65. MR1785620 (2001m:42020) ↑17

[15] B. Solomyak, *On the random series* $\sum \pm\lambda^n$ *(an Erdős problem)*, Ann. of Math. (2) **142** (1995), no. 3, 611–625. MR1356783 (97d:11125) ↑4

[16] B. Solomyak, *Notes on Bernoulli convolutions*, Fractal geometry and applications: a jubilee of Benoît Mandelbrot. Part 1, 2004, pp. 207–230. MR2112107 ↑17

[17] P. P. Varjú, *On the dimension of Bernoulli convolutions for all transcendental parameters*, Ann. of Math. (2) **189** (2019), no. 3, 1001–1011. ↑1, 4, 16

[18] P. P. Varjú, *Recent progress on Bernoulli convolutions*, European Congress of Mathematics. To appear. ↑17

[19] P. P. Varjú, *Absolute continuity of Bernoulli convolutions for algebraic parameters*, 2018. arXiv:1602.00261v4, To appear in J. Amer. Math. Soc. ↑7, 9, 10

[20] Z. Wang, *Quantitative density under higher rank abelian algebraic toral actions*, Int. Math. Res. Not. IMRN **16** (2011), 3744–3821. MR2824843 (2012i:11079) ↑7

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
e-mail address: pv270@dpmms.cam.ac.uk