Metric Diophantine Approximation—From Continued Fractions to Fractals

Simon Kristensen

2000 Mathematics Subject Classification 11J83 · 28A80

Introduction

How does one prove the existence of a real number with certain desired Diophantine properties without knowing a procedure to construct it? And if one also requires the digits in the decimal expansion, say, of this number to be special in some way, is the task then completely impossible? The present notes aim at introducing a number of methods for accomplishing this. Our main tools will be methods from classical Diophantine approximation, from dynamical systems and not least from measure theory. We will assume an acquaintance with basic measure and probability theory and some elementary number theory, but otherwise the notes aim at being self-contained.

The notes are structured as followed. We begin in Sect. 1 with some first and elementary observations on Diophantine approximation and recall some results on continued fractions. Here, we set the scene for the following sections and deduce some first metrical results. In Sect. 2, we relate the machinery of continued fractions to that of ergodic theory. We will use this machinery to deduce Khintchine's theorem in metric Diophantine approximation, which can be seen as a starting point for the metric theory of Diophantine approximation. In Sect. 3, we introduce several notions from fractal geometry. We will discuss Hausdorff measures and Hausdorff dimension, box counting dimension and Fourier dimension. We relate these to sets of arithmetical interest arising both from Diophantine approximation and from representations of real numbers in some integer base. In Sect. 4, we turn our attention

S. Kristensen (✉)
Department of Mathematics, Aarhus University,
Ny Munkegade 118, 8000 Aarhus C, Denmark
e-mail: sik@math.au.dk

J. Steuding (ed.), *Diophantine Analysis*, Trends in Mathematics,
DOI 10.1007/978-3-319-48817-2_2

to higher dimensional problems. The underlying reason for this is two-fold. In a first instance, approximation of real numbers by real algebraic numbers is a higher dimensional problem. A full description of this is unfortunately beyond the scope of these notes, but we briefly outline some results and dwell a little on a conjecture on digit distribution for algebraic irrational numbers. We then turn to our second objective. Here, we study simultaneous and dual approximation of vectors of real numbers and their relation. We will also outline a proof of a higher dimensional variant of Khintchine's theorem. This will be used as a stepping stone for discussing some famous open problems in Diophantine approximation: the Duffin–Schaeffer conjecture and the Littlewood conjecture.

Several null sets of interest arise from the Khintchine type results described. One is the set of elements for which the simple approximation properties which may be derived from variants of the pigeon hole principle cannot be improved beyond a constant. In Sect. 5, we will give a general framework for studying the fractal structure of sets of such elements. In Sect. 6, we discuss the other interesting null sets arising from the Khintchine type theorems. We will outline methods for getting the Hausdorff dimension of these null sets, we will discuss approximation of elements in the ternary Cantor set by algebraic numbers, and finally we will give some results on Littlewood's conjecture. In this final part, ideas from continued fractions, uniform distribution theory, Hausdorff dimension and Fourier analysis come together in a nice blend.

1 Beginnings

Any course on Diophantine approximation should begin with the celebrated result of Dirichlet [18]:

Theorem 1.1 *Let $x \in \mathbb{R}$ and let N be a positive integer. There exist numbers $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ with $q \leq N$ such that*

$$\left|x - \frac{p}{q}\right| < \frac{1}{qN}.$$

Proof Let $[x]$ denote the integer part of x and $\{x\}$ its fractional part, so that $x = [x] + \{x\}$. Divide the interval $[0, 1)$ into N subintervals $[k/N, (k+1)/N)$, where $k = 0, 1, \ldots, N-1$, of length $1/N$. The $N+1$ numbers $\{rx\}$, $r = 0, 1, \ldots, N$, fall into the interval $[0,1)$ and so two, $\{rx\}$ and $\{r'x\}$ say, must fall into the same subinterval, $[k/N, (k+1)/N)$ say. Suppose without loss of generality that $r > r'$. Then

$$\left|\{rx\} - \{r'x\}\right| = \left|rx - [rx] - r'x + [r'x]\right| = |qx - p| < \frac{1}{N},$$

where $q = r - r'$, $p = [rx] - [r'x] \in \mathbb{Z}$ and $1 \leq q \leq N$. Dividing by q finishes the proof.

As an immediate corollary of Dirichlet's theorem, we obtain a non-uniform estimate.

Corollary 1.2 *Let $x \in \mathbb{R}$. For infinitely many pairs (p, q) with $p \in \mathbb{Z}$ and $q \in \mathbb{N}$,*

$$\left|x - \frac{p}{q}\right| < \frac{1}{q^2}. \tag{1.1}$$

Proof If $x \in \mathbb{Q}$, the result is trivial, as we do not require the rationals p/q to be on lowest terms. Suppose now that $x \in \mathbb{R} \setminus \mathbb{Q}$. Fix some $N_1 \in \mathbb{N}$ and choose (p_1, q_1) as in Dirichlet's theorem. By this theorem,

$$\left|x - \frac{p_1}{q_1}\right| < \frac{1}{q_1 N_1} \leq \frac{1}{q_1^2}.$$

As x is irrational, the left hand side must be non-zero. Consequently, it is possible to choose an integer N_2 such that

$$\frac{1}{N_2} < q_1 \left|x - \frac{p_1}{q_1}\right|. \tag{1.2}$$

Taking this value for N in Dirichlet's Theorem gives a pair of points (p_2, q_2) with the desired approximation property. Furthermore, $(p_1, q_1) \neq (p_2, q_2)$ since otherwise (1.2) would contradict the choice of p_2, q_2. Continuing in this way, we obtain a sequence of pairs p_n, q_n satisfying (1.1).

A first natural question in view of the corollary of Dirichlet's theorem is the following: Can the rate of approximation on the right hand side be improved? In general, the answer is negative due to a measure theoretical result. The following is the easy half of Khintchine's theorem, which is our first example of a metric result: it gives a condition for a certain set to be a null-set, so that almost all numbers will lie in its complement. Throughout these notes, for a Borel set $E \subseteq \mathbb{R}^n$, we will denote the Lebesgue measure of E by $|E|$. We will need the notion of a *limsup*-set. Recall that given a sequence of sets E_n, we define the associated *limsup* set,

$$\limsup E_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} E_n.$$

Theorem 1.3 *Let $\psi : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be some function with $\sum_{q=1}^{\infty} q\psi(q) < \infty$. Then,*

$$\left|\left\{x \in \mathbb{R} : \left|x - \frac{p}{q}\right| < \psi(q) \textit{ for infinitely many } (p, q) \in \mathbb{Z} \times \mathbb{N}\right\}\right| = 0,$$

i.e. the set is a null-set with respect to the Lebesgue measure on the real line.

Proof Note first that the set is invariant under translation by integers. Hence, it suffices to prove that the set has Lebesgue measure 0 when intersected with the unit interval [0, 1]. Now, note that this set may be expressed as a *limsup*-set as follows,

$$\left\{x \in [0,1] : \left|x - \frac{p}{q}\right| < \psi(q) \text{ for infinitely many } (p,q) \in \mathbb{Z} \times \mathbb{N}\right\}$$
$$= \bigcap_{N \geq 1} \bigcup_{q \geq N} \bigcup_{p=0}^{q} \left(\frac{p}{q} - \psi(q), \frac{p}{q} + \psi(q)\right) \cap [0,1].$$

In other words, for each $N \in \mathbb{N}$, the set is covered by

$$\bigcup_{q \geq N} \bigcup_{p=0}^{q} \left(\frac{p}{q} - \psi(q), \frac{p}{q} + \psi(q)\right),$$

so using σ-sub-additivity of the Lebesgue measure,

$$\left|\left\{x \in [0,1] : \left|x - \frac{p}{q}\right| < \psi(q) \text{ for infinitely many } (p,q) \in \mathbb{Z} \times \mathbb{N}\right\}\right|$$
$$\leq \left|\left(\bigcup_{q \geq N} \bigcup_{p=0}^{q} \left(\frac{p}{q} - \psi(q), \frac{p}{q} + \psi(q)\right)\right)\right| \leq \sum_{q \geq N} \sum_{p=0}^{q} \left|\left(\frac{p}{q} - \psi(q), \frac{p}{q} + \psi(q)\right)\right|$$
$$= 2 \sum_{q \geq N} (q+1)\psi(q) \leq 4 \sum_{q \geq N} q\psi(q).$$

The latter is the tail of a convergent series, and so will tend to zero as N tends to infinity.

The connoisseur will recognise this as an application of the Borel–Cantelli lemma from probability theory. Again, the result raises more questions. Are the null-sets in fact empty? If not, what makes the elements of these sets so special? And how does one generate the infinitely many good approximants?

The usual strategy is to go via continued fractions, see e.g. [49]. There are many ways to get to these. We will go via an avenue inspired by dynamical systems (for reasons which will become clearer as we progress).

Let $x \in \mathbb{R}$ and define $a_0 = [x]$ and $r_0 = \{x\}$. If $r_0 = 0$, we stop. Otherwise, we see that $1/r_0 > 1$. Let $x_1 = 1/r_0$ and let $a_1 = [x_1]$ and $r_1 = \{x_1\}$. Continuing in this way, we define a (possibly finite) sequence $\{a_n\}$, where $a_0 \in \mathbb{Z}$ and $a_i \in \mathbb{N}$. We define a sequence of rational numbers $\{p_n/q_n\}$ by

$$\frac{p_n}{q_n} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}} = [a_0; a_1, \ldots, a_n]. \tag{1.3}$$

We call the rationals p_n/q_n the *convergents to* x and the integers a_i the *partial quotients of* x. We assume that the procedure and that the following elementary properties are well-known. When $a_0 = 0$ so that $x \in [0, 1)$, we will omit this partial quotient and the semi-colon from the latter notation and write $x = [a_1, a_2, \dots]$.

Proposition 1.4 *The continued fraction algorithm has the following properties:*

(i) *The convergents may be calculated from the following recurrence formulae: Let* $p_{-1} = 1$, $q_{-1} = 0$, $p_0 = a_0$ *and* $q_0 = 1$. *For any* $n \geq 1$,

$$p_n = a_n p_{n-1} + p_{n-2} \quad \textit{and} \quad q_n = a_n q_{n-1} + q_{n-2}.$$

Consequently, $q_n \geq 2^{(n-1)/2}$.

(ii) *For any* $n \geq 0$

$$q_n p_{n-1} - q_{n-1} p_n = (-1)^n,$$

and for any $n \geq 1$,

$$q_n p_{n-2} - q_{n-2} p_n = (-1)^{n-1} a_n$$

(iii) *For an irrational number* x, $x - p_n/q_n$ *is positive if and only if* n *is even.*

(iv) *Any real irrational number* x *has an expansion as a continued fraction. The sequence of convergents of* x *converges to* x, *with the even (resp. odd) order convergents forming a strictly increasing (resp. decreasing) sequence. This expansion is unique, and we write* $x = [a_0; a_1, \dots]$.

(v) *Given a sequence* $\{a_n\}_{n=0}^{\infty}$ *with* $a_0 \in \mathbb{Z}$ *and* $a_i \in \mathbb{N}$ *for* $i \geq 1$, *the sequence* $[a_0; a_1, \dots, a_n]$ *converges to a number having the sequence* $\{a_n\}$ *as its sequence of partial quotients.*

(vi) *The convergents satisfy*

$$\frac{1}{q_n(q_n + q_{n+1})} < \left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

From Proposition 1.4 it is straightforward to construct numbers for which Corollary 1.2 can be improved. Indeed, suppose that we have the a_i for $i \leq n$ given, and let $a_{n+1} = q_n$ where q_n is given by the recursion (i). By (vi), we get

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}} = \frac{1}{q_n(q_n q_n + q_{n-1})} < \frac{1}{q_n^3}. \tag{1.4}$$

By (v), the sequence $\{a_n\}$ defines an irrational number for which the exponent of Corollary 1.2 can be improved to 3 by (1.4). It is easy to modify this construction to produce an uncountable set numbers approximable with any given exponent on the right hand side.

This gives a somewhat satisfactory answer to the questions posed about the null-sets arising from Theorem 1.3. The null-sets are not empty, and the special feature of the elements of the sets is the existence of large partial quotients. Of course, the term 'large partial quotients' should now be quantified, which is where the metrical theory and the use of dynamical systems kicks in. To quantify these notions, we should ask whether there is a typical behaviour of the partial quotients, which is violated for the exceptional numbers.

2 Dynamical Methods

We consider $x \in [0, 1)$ and formalise the continued fraction algorithm in the form of a self-mapping of the unit interval.

Definition 2.1 The *Gauss map* $T : [0, 1) \to [0, 1)$ is defined by

$$Tx = \begin{cases} \{1/x\} & \text{for } x \neq 0 \\ 0 & \text{for } x = 0. \end{cases}$$

In the notation of our description of the continued fraction algorithm, we note that the Gauss map of a number $x \in [0, 1)$ extracts exactly the number $1/r_1$. Applying the Gauss map a second time, we get $T^2 x = 1/r_2$ and so on. It would seem that the Gauss map is an appropriate dynamical description of the continued fractions expansion. All we need is to get the partial quotients out of the r_i. But this can easily be done by defining the axillary function

$$a(x) = \begin{cases} [1/x] & \text{for } x \neq 0 \\ \infty & \text{for } x = 0. \end{cases} \tag{2.1}$$

We now see that

$$a_n(x) = a(T^{n-1}x), \tag{2.2}$$

where $a_n(x)$ denotes the n'th partial quotient in the continued fraction expansion of x, so iterates of the Gauss map are the natural object to study.

Having established that the Gauss map encodes the behaviour of the partial quotients, it is natural to ask for the statistical behaviour of this map – especially as we are interested in typical and a-typical behaviour of the sequence of partial quotients. A tool for this is ergodic theory. We will say that a map $T : [0, 1) \to [0, 1)$ preserves the measure μ if it is a measurable map such that for any measurable set $B \subseteq [0, 1)$, $\mu(T^{-1}B) = \mu(B)$. The Birkhoff (or pointwise) ergodic theorem is the following result (see e.g. [22]).

Theorem 2.2 (The pointwise ergodic theorem) *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and let $T : \Omega \to \Omega$ be a measure preserving transformation. Let $f \in L^1(\Omega)$. Then the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \overline{f}(x)$$

exists for almost every $x \in \Omega$ as well as in $L^1(\Omega)$. If the transformation is ergodic, i.e. $T^{-1}B = B \Rightarrow \mu(B) \in \{0, 1\}$, the function $\overline{f}$ is constant and equal to $\int f d\mu$.

We will not prove the theorem here, but we will apply it to the Gauss map. Our approach is more or less that of [11]. The statement about the map T requires it to be measure preserving, and it is more or less self-evident that the Gauss map does not preserve the Lebesgue measure. However, there is a measure, which is absolutely continuous with respect to Lebesgue measure and with which the Gauss map is ergodic. There are good reasons why this is the correct measure, although it looks slightly mysterious at first sight. For now, we will pull the measure out of a hat and continue to work with it. Later on, we will give some indication of the origins of the measure.

Definition 2.3 Let $\mathcal{B}$ be the Borel σ-algebra in $[0, 1)$. The *Gauss measure* is defined to be the function $\mu : \mathcal{B} \to [0, 1]$ defined by

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+t} dt = \frac{1}{\log 2} \int_0^1 \chi_A(t) \frac{1}{1+t} dt.$$

Theorem 2.4 *The Gauss measure is preserved under the Gauss map, i.e. for any measurable set A, we have $\mu(T^{-1}A) = \mu(A)$.*

Proof We note that it is sufficient to prove that $\mu(T^{-1}[0, y)) = \mu([0, y))$, as we can build any other set from basic set operations on these sets. If one considers the graph of the Gauss map (try drawing it), it is easy to see that

$$T^{-1}([0, y)) = \{x \in [0, 1) : 0 \leq T(x) < y\} = \bigcup_{k=1}^{\infty} \left[\frac{1}{k+y}, \frac{1}{k}\right). \tag{2.3}$$

Thus,

$$\begin{aligned}
\mu\left(T^{-1}[0, y)\right) &= \sum_{k=1}^{\infty} \mu\left(\left[\frac{1}{k+y}, \frac{1}{k}\right)\right) = \sum_{k=1}^{\infty} \frac{1}{\log 2} \int_{1/(k+y)}^{1/k} \frac{1}{1+x} dx \\
&= \sum_{k=1}^{\infty} \frac{1}{\log 2} \left[\log\left(1 + \frac{1}{k}\right) - \log\left(1 + \frac{1}{k+y}\right)\right] \\
&= \sum_{k=1}^{\infty} \frac{1}{\log 2} \log\left(\frac{k+1}{k} \cdot \frac{k+y}{k+y+1}\right).
\end{aligned}$$

This is completely incomprehensible, so we try to get to the same incomprehensible expression from the other side. Cunningly, we make an appropriate partition and get

$$\begin{aligned}\mu([0, y)) &= \frac{1}{\log 2}\int_0^y \frac{1}{1+x}dx = \sum_{k=1}^{\infty}\frac{1}{\log 2}\int_{y/(k+1)}^{y/k}\frac{1}{1+x}dx \\ &= \sum_{k=1}^{\infty}\frac{1}{\log 2}\left[\log\left(1+\frac{y}{k}\right) - \log\left(1+\frac{y}{k+1}\right)\right] \\ &= \sum_{k=1}^{\infty}\frac{1}{\log 2}\log\left(\frac{k+1}{k}\cdot\frac{k+y}{k+y+1}\right).\end{aligned}$$

Luckily, this is the same incomprehensible mess that we have before, so the proof is complete.

Note that the density of the Gauss measure with respect to the Lebesgue measure is continuous, non-negative and in fact invertible. Hence, the two measures are absolutely continuous with respect to each other, and the property of being null or full with respect to one measure automatically implies the same for the other.

We have turned the study of the typical behaviour of continued fractions into a matter of studying the measure preserving system $([0, 1), \mathcal{B}, \mu, T)$, where $\mathcal{B}$ is the Borel σ-algebra, μ is the Gauss measure and T is the Gauss map. We have also seen that the pointwise ergodic theorem is a nice way of studying the almost everywhere behaviour of such maps. It would be desirable if the measure preserving system we have obtained turned out to be ergodic. It turns out that this is in fact the case. We will prove this now.

First, let us see what the Gauss map does to a continued fraction.

Proposition 2.5 *Let $x = [a_1, a_2, \dots] \in [0, 1)$. Then*

$$Tx = T[a_1, a_2, \dots] = [a_2, a_3, \dots].$$

Proof We see that

$$\begin{aligned}T[a_1, a_2, \dots] &= T\left(\cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3+\cdots}}}\right) \\ &= \left\{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3+\cdots}}\right\} = \cfrac{1}{a_2 + \frac{1}{a_3+\cdots}} = [a_2, a_3, \dots]\end{aligned}$$

This is of course obvious from the construction. But in the light of the measure theoretic considerations, it does actually contain information. We define some sets to make life easier.

Definition 2.6 Let $a_1, \dots, a_n \in \mathbb{N}$. Define the *fundamental interval* or *fundamental cylinder*

$$I_n(a_1, \dots, a_n) = \{[a_1, \dots, a_n, b_{n+1}, b_{n+2}, \dots] : b_{n+i} \in \mathbb{N} \text{ for all } i \in \mathbb{N}\}.$$

Note that by Proposition 2.5, the n'th iterate under the Gauss map of any fundamental cylinder I_n is in fact $[0, 1) \setminus \mathbb{Q}$. This reflects the chaotic (or ergodic) nature of the Gauss map. Also note that the cylinders do not include the rational points. This is of little concern to us, as the rationals form a set of measure zero. It does however mean that we have to be extra careful with our bookkeeping.

In the following, let $n \in \mathbb{N}$ and $a_1, \dots, a_n \in \mathbb{N}$ be fixed. Denote by I_n the fundamental interval $I_n(a_1, \dots, a_n)$. We make a few preliminary observations.

Lemma 2.7 *We have $x \in I_n$ if and only if there exists $\theta_n(x) \in (0, 1) \setminus \mathbb{Q}$ such that*

$$x = \cfrac{1}{a_1 + \cfrac{1}{\ddots + \frac{1}{a_n + \theta_n(x)}}}.$$

Proof By definition, $x \in I_n$ if and only if $x = [a_1, \dots, a_n, b_{n+1}, \dots]$. Applying the Gauss map n times, we get

$$\theta_n(x) := T^n x = [b_{n+1}, b_{n+2}, \dots].$$

But this is rational if and only if the sequence of partial quotients b_{n+i} terminates.

The above lemma defines a function on the irrational points in the unit interval $\theta_n : [0, 1) \setminus \mathbb{Q} \to [0, 1) \setminus \mathbb{Q}$.

Lemma 2.8 *Let $u, v \in [0, 1) \setminus \mathbb{Q}$, $u \leq v$. Then*

$$\left|I_n \cap T^{-n}[u, v]\right| = \left|\theta_n^{-1}(v) - \theta_n^{-1}(u)\right|.$$

Proof We see that

$$I_n \cap T^{-n}[u, v] = \{x \in [0, 1) \setminus \mathbb{Q} : x = [a_1, \dots, a_n; \theta]\}$$

where $\theta \in [u, v]$ and $[a_1, \dots, a_n; \theta]$ denotes the continued fraction

$$x = \cfrac{1}{a_1 + \cfrac{1}{\ddots + \frac{1}{a_n + \theta}}}. \tag{2.4}$$

That is, for any $x \in I_n \cap T^{-n}[u, v]$, $\theta_n(x) \in [u, v]$, so

$$I_n \cap T^{-n}[u, v] \subseteq \theta_n^{-1}[u, v].$$

Furthermore, any value of θ in $[u, v]$ inserted in (2.4) will give rise to an element in $I_n \cap T^{-n}[u, v]$, so the converse inclusion holds. Finally, it is an easy exercise left to the reader to see that θ_n^{-1} is monotonic, so this proves the lemma.

It would seem a good idea to find a precise expression for $\theta_n^{-1}(x)$. This may be done from the recursive formulae for the convergents of x.

Lemma 2.9 *We have*

$$\theta_n^{-1}(x) = \frac{p_n + xp_{n-1}}{q_n + xq_{n-1}}.$$

Proof We prove this by induction in n. For $n = 1$, using Proposition 1.4

$$\frac{p_1 + xp_0}{q_1 + xq_0} = \frac{a_1 p_0 + p_{-1} + xp_0}{a_1 q_0 + q_{-1} + xq_0} = \frac{0 + 1 + x \cdot 0}{a_1 + 0 + x} = \frac{1}{a_1 + x}$$

so $\theta_1^{-1}(x) = (p_1 + xp_0)/(q_1 + xq_0)$.

Now, we consider $n + 1$. We let $y = 1/(a_{n+1} + x)$. We know that

$$\theta_{n+1}^{-1}(x) = \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n + \cfrac{1}{a_{n+1}+x}}}} = \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n + y}}} = \theta_n^{-1}(y).$$

By induction hypothesis and Proposition 1.4 again,

$$\theta_n^{-1}(y) = \frac{p_n + yp_{n-1}}{q_n + yq_{n-1}} = \frac{p_n + \left(\frac{1}{a_{n+1}+x}\right) p_{n-1}}{q_n + \left(\frac{1}{a_{n+1}+x}\right) q_{n-1}}$$
$$= \frac{a_{n+1}p_n + p_{n-1} + xp_n}{a_{n+1}q_n + q_{n-1} + xq_n} = \frac{p_{n+1} + xp_n}{q_{n+1} + xq_n}.$$

This completes the proof.

Note that θ_n^{-1} is continuous, so we may extend it to all of $[0, 1]$ and still have Lemma 2.8. We now introduce the so-called Vinogradov notation to make our notation less cumbersome.

Definition 2.10 For two real expressions x and y, we say that $x \ll y$ if there exists a constant $c > 0$ such that $x \leq cy$. If $x \ll y$ and $y \ll x$ we write $x \asymp y$.

Lemma 2.11 *Let* $u, v \in [0, 1)$ *with* $u \leq v$. *Then*

$$\frac{\left|T^{-n}[u, v] \cap I_n\right|}{|I_n|} \asymp |[u, v]|,$$

where the implied constants in $\asymp$ *do not depend on the sequence* (a_n) *defining the* I_n.

Proof We use Lemma 2.8 to obtain

$$\frac{\left|T^{-n}[u,v]\cap I_n\right|}{|I_n|}=\left|\frac{\theta_n^{-1}(v)-\theta_n^{-1}(u)}{\theta_n^{-1}(1)-\theta_n^{-1}(0)}\right|=\left|\frac{\frac{p_n+vp_{n-1}}{q_n+vq_{n-1}}-\frac{p_n+up_{n-1}}{q_n+uq_{n-1}}}{\frac{p_n+p_{n-1}}{q_n+q_{n-1}}-\frac{p_n}{q_n}}\right|$$
$$=(v-u)\left|\frac{q_n(q_n+q_{n-1})}{(q_n+vq_{n-1})(q_n+uq_{n-1})}\right|.$$

The last reduction requires substantial, but completely elementary calculations using Proposition 1.4.

Now, the denominators of the convergents q_n satisfy $q_{n-1}/q_n<1$, so it is easy to see by Proposition 1.4 (i) that

$$\left|\frac{q_n(q_n+q_{n-1})}{(q_n+vq_{n-1})(q_n+uq_{n-1})}\right|\asymp 1.$$

As $|[u,v]|=v-u$, the proof is completed.

Lemma 2.12 *For every $A\in\mathcal{B}$,*

$$\frac{\mu(T^{-n}A\cap I_n)}{\mu(I_n)}\asymp\mu(A).$$

Proof As the Borel σ-algebra is generated by intervals, by Lemma 2.11 for any $A\in\mathcal{B}$,

$$\frac{\left|T^{-n}A\cap I_n\right|}{|I_n|}\asymp|A|\,. \tag{2.5}$$

Also, since $1/2<1/(1+t)\le 1$ for $t\in[0,1)$, we have for any $A\in\mathcal{B}$,

$$\frac{1}{2}\,|A|=\int_A\frac{1}{2}dt\le\int_A\frac{1}{1+t}dt=\mu(A)$$

and

$$\mu(A)=\int_A\frac{1}{1+t}dt\le\int_A 1dt=|A|\,.$$

Hence, $\mu(A)\asymp|A|$, so the Lemma follows from (2.5).

Theorem 2.13 *The Gauss map is ergodic with respect to the Gauss measure.*

Proof Suppose that $T^{-1}A=A$ and that $\mu(A)>0$. It suffices to prove that $\mu(A)=1$. Any Borel set can be generated by the I_n, as these intervals are essentially disjoint with lengths tending to zero. Hence, by generating a set B by I_n's of the same level (up to an arbitrarily small error), Lemma 2.12 implies that

$$\mu(T^{-n}A\cap B)\asymp\mu(A)\mu(B)$$

for any $B \in \mathcal{B}$. On the other hand, as $T^{-1}A = A$,

$$\mu(T^{-n}A \cap B) = \mu(A \cap B)$$

Letting $B = A^c$, we see that $\mu(A \cap B) = 0$, so that $\mu(B) \asymp 0$. This clearly implies that $\mu(B) = \mu(A^c) = 0$, so that $\mu(A) = 1$.

We specify the following corollary of the Pointwise Ergodic Theorem and Theorem 2.13:

Corollary 2.14 *Let f be an integrable function on* $[0, 1)$. *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \frac{1}{\log 2} \int_0^1 \frac{f(t)}{1+t} dt$$

for almost every $x \in [0, 1)$.

Many things follow from the ergodicity of the Gauss map. For instance, almost all numbers have an unbounded sequence of partial quotients, and in fact the arithmetic mean of the partial quotients is infinite almost surely. On the other hand, the geometric mean does have a limiting value almost surely. Calculating the typical frequency of any prescribed partial quotient is an easy exercise in integration, and combining these results with the machinery of continued fractions give a unified way in which to prove many of the classical metrical results in Diophantine approximation. An example is Lévy's theorem.

Theorem 2.15 *For almost every $x \in [0, 1)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log q_n(x) = \frac{\pi^2}{12 \log 2}.$$

We will not derive this theorem here, although we will be appealing to it later. Instead, let us derive a partial converse to Theorem 1.3 due to Khintchine [49].

Theorem 2.16 *Let $\psi : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be some function with $\sum_{q=1}^{\infty} q\psi(q) = \infty$ and with $q\psi(q)$ non-increasing. Then,*

$$\left| \left\{ x \in [0, 1] : \left| x - \frac{p}{q} \right| < \psi(q) \textit{ for infinitely many } (p, q) \in \mathbb{Z} \times \mathbb{N} \right\} \right| = 1.$$

This is the difficult half of Khintchine's theorem, which in its totality consists of Theorem 1.3 and Theorem 2.16. It should be noted that there is an additional assumption on the function ψ. This is strictly needed. We will discuss this later in these notes.

Lemma 2.17 *Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers. Suppose that*

$$\sum_{n=1}^{\infty} \tfrac{1}{\alpha_n} = \infty.$$

Then,

$$|\{x \in [0, 1) : a_n(x) > \alpha_n \textit{ for infinitely many } n \in \mathbb{N}\}| = 1.$$

Proof We fix arbitrary $a_1, \ldots, a_n \in \mathbb{N}$ and let I_n denote the fundamental interval corresponding to these partial quotients. We let $E_n = \{x \in [0, 1) : a_n(x) > \alpha_n\}$. We first prove that

$$\frac{\mu(E_{n+1} \cap I_n)}{\mu(I_n)} \gg \frac{1}{\alpha_{n+1} + 1}. \tag{2.6}$$

In fact, all we need to do is to note that

$$E_{n+1} = T^{-n} \left\{x \in [0, 1) : x = [\hat{a}_{n+1}, \ldots] \text{ where } \hat{a}_{n+1} > \alpha_{n+1}\right\}.$$

Thus, by Lemma 2.12,

$$\frac{\mu(E_{n+1} \cap I_n)}{\mu(I_n)} \asymp \mu\left(\left\{x \in [0, 1) : x = [\hat{a}_{n+1}, \ldots] \text{ where } \hat{a}_{n+1} > \alpha_{n+1}\right\}\right).$$

But since the Gauss measure and the Lebesgue measure are absolutely continuous with respect to each other, the above quantity is

$$\asymp \left|\left\{x \in [0, 1) : x = [\hat{a}_{n+1}, \ldots] \text{ where } \hat{a}_{n+1} > \alpha_{n+1}\right\}\right|$$
$$= \sum_{k > \alpha_{n+1}} \left(\frac{1}{k} - \frac{1}{k+1}\right) \gg \frac{1}{\alpha_{n+1} + 1}$$

Repeatedly using (2.6), we get for some universal $C' > 0$,

$$\mu(E_m^c \cap \cdots \cap E_{m+k}^c) \leq \prod_{i=1}^{k} \left(1 - \frac{1}{C'\alpha_{m+i} + 1}\right).$$

This holds for any $m, k \in \mathbb{N}$. To see this, note that we can express the property of being in E_m^c as being in some union of disjoint fundamental intervals. We leave the formalism as an exercise for the interested reader.

Finally,

$$\mu\left(\bigcap_{i=m}^{\infty} E_{m+i}^c\right) \leq \prod_{i=1}^{\infty} \left(1 - \frac{1}{C'\alpha_{m+i} + 1}\right).$$

As $1 - x \leq e^{-x}$ whenever $0 \leq x < 1$, we have

$$\prod_{i=1}^{n}\left(1-\frac{1}{C'\alpha_{m+i}+1}\right) \leq \prod_{i=1}^{n} e^{-\frac{1}{C'\alpha_{m+i}+1}} = e^{-\sum_{i=1}^{n}\frac{1}{C'\alpha_{m+i}+1}}.$$

But this tends to 0, as $\sum 1/\alpha_i$ is assumed to diverge. Thus, the probability that $a_n > \alpha_n$ only occurs finitely many times is zero, which proves the lemma.

We are now ready to prove Khintchine's Theorem.

Proof (*Part II (the divergence case)*) We let N be a fixed integer such that $\log N > \pi^2/12\log 2$ ($N = 4$ will do nicely). By Theorem 2.15, for all but finitely many values of n,

$$\frac{1}{n}\log q_n(x) < \log N \tag{2.7}$$

for almost all $x \in [0, 1)$.

Let $f(q) = q\psi(q)$. Define a function $\phi(n) = N^n f(N^n)$. Since $f(q) = q\psi(q)$ is non-increasing,

$$\sum_{q=N^n}^{N^{n+1}-1} f(q) \leq (N^{n+1} - N^n)f(N^n) = (N-1)\phi(n),$$

so as $\sum f(q)$ diverges, this will also be the case for $\sum \phi(n)$. Therefore, by Lemma 2.17, for almost every $x \in [0, 1)$,

$$a_{n+1}(x) > \frac{1}{\phi(n)}$$

holds for infinitely many n.

Now, we apply our classical estimates:

$$\left|x - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n(x)q_{n+1}(x)} < \frac{1}{a_{n+1}q_n(x)^2} < \frac{\phi(n)}{q_n(x)^2}$$

for infinitely many n for almost all x. But by (2.7), $q_n(x) < N^n$, and as $qf(q)$ is non-increasing, we get

$$\phi(n) = N^n f(N^n) \leq q_n(x) f(q_n(x)).$$

Hence, for infinitely many n,

$$\left|x - \frac{p_n}{q_n}\right| < \frac{q_n(x)f(q_n(x))}{q_n(x)^2} = \psi(q_n(x)).$$

This proves the theorem.

We conclude the second section with an informal discussion of a different – and somewhat more modern – view on the Gauss map. Let us 'decompose' the action of the Gauss map into new maps. To apply the map $x \mapsto \{1/x\}$, we first apply the map $x \mapsto 1/x$ and continue to apply the map $x \mapsto x - 1$ until we finally arrive in the unit interval again. These maps are both Möbius maps given by the matrices

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Recall that the Möbius map associated to a matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is given by

$$x \mapsto \frac{ax+b}{cx+d}.$$

Composition of such maps corresponds to taking products of matrices.

For convenience, we make some sign changes here and there and consider instead the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Together, these matrices generate the group $\mathrm{SL}_2(\mathbb{Z})$.

It is tempting to look for a space on which this group acts naturally, and indeed such a space exists. The hyperbolic plane is such an object. Consider the upper half plane,

$$\mathbb{H} = \{x + iy \in \mathbb{C} : y > 0\}$$

with the Riemannian metric

$$\langle v, w \rangle_z = \frac{1}{y^2}(v \cdot w) \text{ for } z = x + iy.$$

The group $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acts by isometries via Möbius maps on $\mathbb{H}$. The group $\mathrm{SL}_2(\mathbb{Z})$ forms a lattice inside this group of isometries, and so we can consider the Riemann surface $\mathcal{M} = \mathbb{H}/\mathrm{SL}_2(\mathbb{Z})$. This is a surface with three singularities: two points where it is non-smooth and a cusp. Applying the above generators roughly corresponds to crossing the sides of the fundamental domain of the group $\mathrm{SL}_2(\mathbb{Z})$.

Considering the boundary of the hyperbolic plane, $\mathbb{R} \cup \{\infty\}$, we easily see that all rational numbers are identified with the point at infinity under the action of $\mathrm{SL}_2(\mathbb{Z})$. This point in turn becomes the cusp of the surface $\mathcal{M}$.

If one formalises the above discussion, one may prove that the continued fraction of a number corresponds to a geodesic on the surface $\mathcal{M}$. Formalising this is beyond the scope of these notes, but the reader is referred to the paper [46] or the monograph

[22]. The ergodicity of the Gauss map can be seen as an instance of the ergodicity of the geodesic flow on $\mathcal{M}$ (or more generally on surfaces of constant negative curvature). The geodesic flow on $\mathcal{M}$ is a flow in the unit tangent bundle of $\mathcal{M}$, which may be identified with $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. This can be seen as the starting point of the use of homogeneous dynamics in Diophantine approximation, an approach which has become tremendously important in recent years. It also gives an explanation for the origins of the curious density of the Gauss measure, which can be induced on the unit interval from the natural measure on $\mathcal{M}$. In a sense, the hyperbolic measure can be seen as an instance of the hat out of which we previously pulled the Gauss measure.

3 Fractal Geometry – A Crash Course

As we saw in the last section, there is a nice zero–one dichotomy as far as the Lebesgue measure of the set of real numbers with prescribed approximation properties is concerned. However, the null sets obtained in the convergence case of Khintchine's theorem are not empty. One could easily use the machinery of continued fractions to prove that a given set is uncountable, but in fact we may discriminate even more precisely between the sizes of the null sets using the notion of Hausdorff dimension.

Hausdorff dimension was introduced by Felix Hausdorff [27], building on the construction of the Lebesgue measure given by Carathéodory [15]. Carathéodory constructed the Lebesgue measure by approximating a set E by countable covers of simple sets. The simple sets would have a volume, which could be calculated by elementary means. On adding these countably many volumes, Carathéodory would obtain an upper bound on the volume of the set E. To get the Lebesgue measure, one takes the infimum over all such covers. This produces an outer measure for which the Borel sets are measurable.

Hausdorff made the simple but far-reaching observation that if one replaces the usual volume of the sets in the covers by an appropriate function of their diameter, a different measure would be obtained. The usual volume of a hypercube in $\mathbb{R}^n$ is a constant multiple of its diameter raised to the power n, with the constant depending only on n, so this is an entirely natural thing to do. It turns out that an abundance of sets supporting such a measure exist. In particular, with the added flexibility of using different functions, one can discriminate between the sizes of Lebesgue null sets.

Let us be more concrete. For a given countable cover, $\mathcal{C}$ say, of E we consider the following sum sometimes termed the *s-length* of the cover $\mathcal{C}$, given by

$$\ell^s(\mathcal{C}) := \sum_{U \in \mathcal{C}} (\operatorname{diam} U)^s,$$

where $\operatorname{diam} U = \sup\{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in U\}$ is the diameter of U and where $s \geq 0$ is some real number. We will also consider yet another generalisation of the above, also considered by Hausdorff. A *dimension function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous, monotonic function with $f(r) \to 0$ as $r \to 0$. In the above, we may replace $(\text{diam } U)^s$ by $f(\text{diam } U)$ for any such function to obtain the more general notion of f-length.

For clarity of exposition, in the following we consider the special case $f(r) = r^s$ corresponding to the s-length as defined above. This is associated with what is now usually called the Hausdorff dimension but also sometimes called the Hausdorff–Besicovitch dimension. The possibly infinite number $\ell^s(\mathcal{C})$ gives an indication the 's-dimensional volume' of the set E in much the same way Carathéodory would think of it. Taking yet another hint from Carathéodory, the diameter of the sets U in the cover is now restricted to be at most $\delta > 0$.

Let

$$\mathcal{H}^s_\delta(E) := \inf_{\mathcal{C}_\delta} \sum_{U \in \mathcal{C}_\delta} (\text{diam } U)^s = \inf_{\mathcal{C}_\delta} \ell^s(\mathcal{C}_\circ),$$

where the infimum is taken over all covers $\mathcal{C}_\delta$ of E by sets U with diam $U \leq \delta$; such covers are called δ-covers. As δ decreases, $\mathcal{H}^s_\delta$ can only increase as there are fewer U's available, i.e. if $0 < \delta < \delta'$, then

$$\mathcal{H}^s_{\delta'}(E) \leq \mathcal{H}^s_\delta(E).$$

The set function $\mathcal{H}^s_\delta$ is an outer measure on $\mathbb{R}^n$. The limit $\mathcal{H}^s$ (which can be infinite) as $\delta \to 0$, given by

$$\mathcal{H}^s(E) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(E) \ = \sup_{\delta > 0} \mathcal{H}^s_\delta(E) \in [0, \infty], \tag{3.1}$$

is however nicer to work with, as it is a regular outer measure with respect to which the Borel sets are measurable. It is usually called the *Hausdorff s-dimensional measure*. Hausdorff 1-dimensional measure coincides with 1-dimensional Lebesgue measure and in higher dimensions, Hausdorff n-dimensional measure is comparable to n-dimensional Lebesgue measure, i.e.

$$\mathcal{H}^n(E) \asymp |E|,$$

where $|E|$ is the Lebesgue measure of E and the implied constants depend only on n and not on the set E. Thus a set of positive n-dimensional Lebesgue measure has positive Hausdorff n-measure.

As the definition depends only on the diameter of the covering sets, there is no loss of generality in restricting to considering only covers consisting of open, closed or convex sets. Additionally, the resulting measure is clearly invariant under isometries, and scaling affects the measure in a completely natural way: for any $r \geq 0$,

$$\mathcal{H}^s(rE) = r^s \mathcal{H}^s(E).$$

As a function of s, the s-dimensional Hausdorff measure of a fixed set E exhibits an interesting behaviour. For a set E, $\mathcal{H}^s(E)$ is either 0 or ∞, except for possibly one value of s. To see this, the definition of $\mathcal{H}^s_\delta(E)$ implies that there is a δ-cover $\mathcal{C}_\delta$ of E such that

$$\sum_{C\in\mathcal{C}_\delta}(\operatorname{diam} C)^s \leq \mathcal{H}^s_\delta(E)+1 \leq \mathcal{H}^s(E)+1.$$

Suppose that $\mathcal{H}^{s_0}(E)$ is finite and $s = s_0+\varepsilon$, $\varepsilon > 0$. Then for each member C of the cover $\mathcal{C}_\delta$, $(\operatorname{diam} C)^{s_0+\varepsilon} \leq \delta^\varepsilon(\operatorname{diam} C)^{s_0}$, so that the sum

$$\sum_{C\in\mathcal{C}_\delta}(\operatorname{diam} C)^{s_0+\varepsilon} \leq \delta^\varepsilon \sum_{C\in\mathcal{C}_\delta}(\operatorname{diam} C)^{s_0}.$$

Hence

$$\mathcal{H}^{s_0+\varepsilon}_\delta(E) \leq \sum_{C\in\mathcal{C}_\delta}(\operatorname{diam} C)^{s_0+\varepsilon} \leq \delta^\varepsilon \sum_{C\in\mathcal{C}_\delta}(\operatorname{diam} C)^{s_0} \leq \delta^\varepsilon(\mathcal{H}^{s_0}(E)+1),$$

and so

$$0 \leq \mathcal{H}^s(E) = \mathcal{H}^{s_0+\varepsilon}(E) = \lim_{\delta\to 0}\mathcal{H}^{s_0+\varepsilon}_\delta(E) \leq \lim_{\delta\to 0}\delta^\varepsilon(\mathcal{H}^{s_0}(E)+1) = 0.$$

On the other hand suppose $\mathcal{H}^{s_0}(E) > 0$. If for any $\varepsilon > 0$, $\mathcal{H}^{s_0-\varepsilon}(E)$ were finite, then by the above $\mathcal{H}^{s_0}(E) = 0$, a contradiction, whence $\mathcal{H}^{s-\varepsilon}(E) = \infty$.

To summarise, we have obtained a set function $\mathcal{H}^s$ associating to each infinite set $E \subseteq R^n$ an exponent $s_0 \geq 0$ for which

$$\mathcal{H}^s(E) = \begin{cases} \infty, & 0 \leq s < s_0, \\ 0, & s_0 < s < \infty. \end{cases}$$

The critical exponent

$$s_0 = \inf\{s \in [0,\infty) \colon \mathcal{H}^s(E) = 0\} = \sup\{s \in [0,\infty) \colon \mathcal{H}^s(E) = \infty\}, \tag{3.2}$$

where the Hausdorff s-measure crashes is called the *Hausdorff dimension* of the set E and is denoted by $\dim_{\mathrm{H}} E$. It is clear that if $\mathcal{H}^s(E) = 0$ then $\dim_{\mathrm{H}} E \leq s$; and if $\mathcal{H}^s(E) > 0$ then $\dim_{\mathrm{H}} E \geq s$. However, nothing is revealed from the definition about the measure at the critical exponent, and indeed it can take any value in the interval $[0,\infty]$ with ∞ included.

The main properties of Hausdorff dimension for sets in $\mathbb{R}^n$ are:

(i) If $E \subseteq F$ then $\dim_{\mathrm{H}} E \leq \dim_{\mathrm{H}} F$.
(ii) $\dim_{\mathrm{H}} E \leq n$.
(iii) If $|E| > 0$, then $\dim_{\mathrm{H}} E = n$.
(iv) The dimension of a point is 0.

(v) If $\dim_{\mathrm{H}} E < n$, then $|E| = 0$ (however $\dim_{\mathrm{H}} E = n$ does not imply $|E| > 0$).
(vi) $\dim_{\mathrm{H}}(E_1 \times E_2) \geq \dim_{\mathrm{H}} E_1 + \dim_{\mathrm{H}} E_2$.
(vii) $\dim_{\mathrm{H}} \cup_{j=1}^{\infty} E_j = \sup\{\dim_{\mathrm{H}} E_j \colon j \in \mathbb{N}\}$.

It easily follows from the above properties that the Hausdorff dimension of any countable set is 0 and that of any open set in $\mathbb{R}^n$ is n. The nature of the construction of Hausdorff measure ensures that the Hausdorff dimension of a set is unchanged by an invertible transformation which is bi-Lipschitz. This implies that for any set $S \subseteq \mathbb{R} \setminus \{0\}$, $\dim_{\mathrm{H}} S^{-1} = \dim_{\mathrm{H}} S$, where $S^{-1} = \{s^{-1} \colon s \in S\}$. To see this, we split up the positive real axis into intervals $(\frac{1}{n}, \frac{1}{n-1}]$ for $n \geq 2$ together with the intervals $(m, m+1]$ for $m \geq 1$. The negative real axis is similarly decomposed. On each interval, the map $s \mapsto s^{-1}$ is bi-Lipschitz, and so the statement follows by appealing to (vii) above.

Thus on the whole, Hausdorff dimension behaves as a dimension should, although that the natural formula $\dim_{\mathrm{H}}(E_1 \times E_2) = \dim_{\mathrm{H}} E_1 + \dim_{\mathrm{H}} E_2$ does *not* always hold (it does hold for certain sets, e.g., cylinders, such as $E \times I$, where I is an interval: $\dim_{\mathrm{H}}(E \times I) = \dim_{\mathrm{H}} E + \dim_{\mathrm{H}} I = \dim_{\mathrm{H}} E + 1$ by (iii), see [24]).

It is often convenient to restrict the elements in the δ-covers of a set to simpler sets such as balls or cubes. For example, covers consisting of hypercubes

$$H = \{\mathbf{x} \in \mathbb{R}^n \colon |\mathbf{x} - \mathbf{a}|_\infty < \delta\},$$

where $|\mathbf{x}|_\infty = \max\{|x_j| \colon 1 \leq j \leq n\}$ is the *height* of $\mathbf{x} \in \mathbb{R}^n$, centred at $\mathbf{a} \in \mathbb{R}^n$ and with sides of length 2δ are used extensively. While outer measures corresponding to these more convenient restricted covers are not the same as Hausdorff measure, they are comparable and so have the same critical exponent. Thus there is no loss as far as dimension is concerned if the sets U are chosen to be balls or hypercubes.

Of course, the two measures are identical for sets with Hausdorff s-measure which is either 0 or ∞. Such sets are said to obey a '0-∞' law, this being the appropriate analogue of the more familiar '0-1' law in probability. Sets which do not satisfy a 0-∞ law, i.e. sets which satisfy

$$0 < \mathcal{H}^{\dim_{\mathrm{H}} E}(E) < \infty, \tag{3.3}$$

are called *s-sets*; these occur surprisingly often and enjoy some nice properties. One example is the Cantor set which has Hausdorff s-measure 1 when $s = \log 2/\log 3$.

However it seems that s-sets are of minor interest in Diophantine approximation where the sets that arise naturally, such as the set of badly approximable numbers or the set of numbers approximable to a given order (see next section), obey a 0-∞ law. The first steps in this direction were taken by Jarník, who proved that the Hausdorff s-measure of the set of numbers rationally approximable to order v was 0 or ∞. This result turns on an idea related to density of Hausdorff measure.

Lemma 3.1 *Let E be a null set in $\mathbb{R}$ and let $s \in [0, 1]$. Suppose that there exists a constant $K > 0$ such that for any interval (a, b) and $s \in [0, 1]$,*

$$\mathcal{H}^s(E \cap (a,b)) \leq K(b-a)\mathcal{H}^s(E). \tag{3.4}$$

Then $\mathcal{H}^s(E) = 0$ *or* ∞.

Proof Suppose the contrary, i.e. suppose $0 < \mathcal{H}^s(E) < \infty$ and let K be as in the statement of the theorem. Since E is null, there exists a cover of E by open intervals (a_j, b_j) such that

$$\sum_j (b_j - a_j) < \frac{1}{K}.$$

By (3.4),

$$\mathcal{H}^s(E) = \mathcal{H}^s\left(\bigcup_j (a_j, b_j) \cap E\right) \leq K\mathcal{H}^s(E)\sum_j (b_j - a_j) < \mathcal{H}^s(E),$$

a contradiction.

The proof for a general outer measure is essentially the same. The sets we encounter in Diophantine approximation are generally not s-sets and some satisfy this ‘quasi-independence’ property. For instance, it was shown in [14], using a variant of the above lemma, that there is no dimension function such that the associated Hausdorff measure of the set of Liouville numbers (defined below) in an interval is positive and finite. Liouville numbers are those real numbers x for which we for any $v > 0$ can find a rational p/q such that

$$0 < \left| x - \frac{p}{q} \right| < \frac{1}{q^v}.$$

Note that in this case, the existence of a single rational p/q for each $v > 0$ implies the existence of infinitly many. We prove below that this set has Hausdorff dimension 0. The result of [14] shows that even with a general Hausdorff measure $\mathcal{H}^f$, we still cannot get a positive and finite measure.

Unless some general result is available, the Hausdorff dimension $\dim_{\mathrm{H}} E$ of a null set E is usually determined in two steps, with the correct upward inequality $\dim_{\mathrm{H}} E \leq s_0$ and downward inequality $\dim_{\mathrm{H}} E \geq s_0$ being established separately.

For limsup sets, such as the one in Theorems 1.3 and 2.16, the Hausdorff measure version of the Borel–Cantelli lemma is often useful.

Lemma 3.2 *Let* (E_k) *be some sequence of arbitrary sets in* $\mathbb{R}^n$ *and let*

$$E = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in E_k \text{ for infinitely many } k \in \mathbb{N}\}.$$

If for some $s > 0$,

$$\sum_{k=1}^{\infty} \operatorname{diam}(E_k)^s < \infty, \tag{3.5}$$

then $\mathcal{H}^s(E) = 0$ *and* $\dim_H E \le s$.

Proof From the definition, for each $N = 1, 2, \ldots,$

$$E \subseteq \bigcup_{k=N}^{\infty} E_k,$$

so that the family $\mathcal{C}^{(N)} = \{E_k : k \ge N\}$ is a cover for E. By (3.5),

$$\lim_{N\to\infty} \sum_{k=N}^{\infty} \mathrm{diam}(E_k)^s = 0.$$

Hence $\lim_{k\to\infty} \mathrm{diam}(E_k) = 0$ and therefore given $\delta > 0$, $\mathcal{C}^{(N)}$ is a δ-cover of E for N sufficiently large. But

$$\mathcal{H}^s_\delta(E) = \inf_{\mathcal{C}_\delta} \sum_{U\in\mathcal{C}_\circ} (\mathrm{diam}\, U)^s \le \ell^s(\mathcal{C}^{(N)}) = \sum_{k=N}^{\infty} \mathrm{diam}(E_k)^s \to 0$$

as $N \to \infty$. Thus $\mathcal{H}^s_\delta(E) = 0$ and by (3.1), $\mathcal{H}^s(E) = 0$, whence $\dim_{\mathrm{H}} E \le s$.

This was essentially what we did in the last section to prove the easy half of Khintchine's theorem. It follows *mutatis mutandis* from that proof that the Hausdorff dimension of the set

$$\left\{x \in \mathbb{R} : \left|x - \frac{p}{q}\right| < \psi(q) \text{ for infinitely many } (p, q) \in \mathbb{Z} \times \mathbb{N}\right\},$$

is at most s whenever $\sum_{q=1}^{\infty} q\psi(q)^s < \infty$, so that if $\psi(q) = q^{-v}$, the Hausdorff dimension would be at most $2/v$. As with Khintchine's theorem, this is sharp, but the converse inequality is more difficult to prove. We return to it in the final section.

Various methods exist for establishing lower bounds on the Hausdorff dimension of a set. Underlying most of these methods are variants of the so-called mass distribution principle. Of course, the key difficulty is the fact that to get lower bounds, we need to consider all covers rather than just exhibiting a single cover. This is due to the definition of the Hausdorff s-measure as the infimum over all covers of the s-length. The mass distribution principle is the following simple result.

Lemma 3.3 *Let μ be a finite and positive measure supported on a bounded subset E of $\mathbb{R}^n$. Suppose that for some $s \ge 0$, there are strictly positive constants c and δ such that $\mu(B) \le c\,(\mathrm{diam}\, B)^s$ for any ball B in $\mathbb{R}^n$ with $\mathrm{diam}\, B \le \delta$. Then $\mathcal{H}^s(E) \ge \mu(E)/c$. In particular, $\dim_H E \ge s$.*

Proof Let $\{B_k\}$ be a δ-cover of E by balls B_k. Then

$$0 < \mu(E) \leq \mu\left(\bigcup_k B_k\right) \leq \sum_k \mu(B_k) \leq c \sum_k (\operatorname{diam} B)^s.$$

Taking infima over all such covers, we see that $\mathcal{H}^s_\delta(E) \geq \mu(E)/c$, whence on letting $\delta \to 0$,

$$\mathcal{H}^s(E) \geq \frac{\mu(E)}{c} > 0.$$

Thus if E supports a probability measure μ ($\mu(E) = 1$) with $\mu(B) \ll (\operatorname{diam} B)^s$ for all sufficiently small balls B, then $\dim_{\mathrm{H}} E \geq s$.

We now briefly discuss two other notions of dimension, namely box counting dimension and Fourier dimension. Box counting dimension (see e.g. [24]) is somewhat easier to calculate from the empirical side, although it has some serious drawbacks. Given a set $E \subseteq R^n$ and a number $\delta > 0$, let $N_\delta(E)$ denote the least number of closed balls of radius δ needed to cover E. We then define the upper and lower box counting dimensions of E as

$$\underline{\dim}_B(E) = \liminf_{\delta \to 0} \frac{\log N_\delta(E)}{-\log \delta}, \quad \overline{\dim}_B(E) = \limsup_{\delta \to 0} \frac{\log N_\delta(E)}{-\log \delta}.$$

If the values agree, we call this the box counting dimension of E.

The definition is fairly flexible, and changing the counting function to any of a number of related counting functions does not change the value. The above counting function has been chosen as a similar quantity should be familiar to readers experienced with the concept of entropy.

A major drawback of box counting dimension is that it is not associated with any measure. As a consequence of this, it is easy to construct countable sets of positive box counting dimension. In fact, one easily proves that the box counting dimension is unchanged when taking the topological closure of the set in question. Hence, any dense and countable subset of $\mathbb{R}$ has box counting dimension 1. This is problematic if one would like to apply the notion to prove the existence of transcendental numbers with certain properties using metrical methods. Indeed, as the set of algebraic numbers is countable and dense, the statement that a set has positive box counting dimension does not imply that it must contain transcendental numbers. On the other hand, proving that a set has box counting dimension zero is a much stronger statement than the corresponding one for Hausdorff dimension.

We will define one more notion of dimension, namely the Fourier dimension of a set (see [38]). For a measure μ, denote by $\hat{\mu}$ its Fourier transform, i.e.

$$\hat{\mu}(x) = \int e^{-2\pi i \xi \cdot x} d\mu(\xi).$$

We are concerned with positive Radon probability measures, so we suppose that μ is a positive, regular Borel measure with $\mu(\mathbb{R}^n) = 1$. A poor man's version of the uncertainty principle would state, that if the Fourier transform of μ decays as $|x|$

increases, the support of the measure would be 'smeared out', and so be somewhat messy.

The technical definition of the Fourier dimension of a set is as follows: Let $\dim_F(E)$ be the unique number in $[0, n]$ such that for any $s \in (0, \dim_F(E))$, there is a non-zero Radon probability measure μ with $\mathrm{supp}(\mu) \subseteq E$ and with $|\hat{\mu}(x)| \leq |x|^{-s/2}$, and such that for any $s > \dim_F(E)$, no such measure exists.

The types of dimension mentioned here are related as follows for a set $E \subseteq \mathbb{R}^n$:

$$\dim_F(E) \leq \dim_{\mathrm{H}}(E) \leq \underline{\dim}_B(E) \leq \overline{\dim}_B(E). \tag{3.6}$$

Proving the last two inequalities is straightforward, but the first one is difficult, and requires Frostman's Lemma [25], a powerful converse to the mass distribution principle of Lemma 3.3.

We end this section by relating the fractal concepts discussed so far to arithmetical issues. This will motivate the key problem considered when discussing Diophantine approximation on fractal sets.

A classical theorem of Émile Borel [12] states that almost all real numbers with respect to the Lebesgue measure are normal to any integer base $b \geq 2$ (or absolutely normal). In other words, for almost all numbers x, any block of digits occurs in the base b expansion of x with the expected frequency, independently of b. In fact, with reference to the previous section, this can be deduced from the ergodicity of the maps $T_b : [0, 1) \to [0, 1)$ given by $T_b(x) = \{bx\}$ with respect to the Lebesgue measure. The ergodicity of these maps is much easier to prove than for the Gauss map, and the deduction of Borel's result is left as an exercise.

While almost all numbers are normal to any base, the only actual examples known of such numbers are artificial and very technical to even write down (see e.g. [44]). For well-known constants such at π, $\log 2$ or even $\sqrt{2}$, very little is known about their distribution of digits. It is a long-standing conjecture that algebraic irrational numbers should be absolutely normal. In view of this, it seems natural to study which Diophantine properties a number which *fails* to be normal in some way can enjoy. One could hope that this would shed light on the question of the normality (or non-normality) of algebraic numbers.

With these remarks, let us consider a specific set of non-normal numbers which is of interest. Any such set will have Lebesgue measure zero, but in order to get anywhere with our analysis, we will take a particularly structured example. Let

$$\mathcal{C} = \left\{ x \in [0, 1] : x = \sum_{i=1}^{\infty} a_i 3^{-i}, a_i \in \{0, 2\} \right\},$$

i.e. the set of numbers in $[0, 1]$ which can be expressed in base 3 without using the digit 1. Clearly, an element of $\mathcal{C}$ cannot be absolutely normal. Of course, this set is just the well-known ternary Cantor set, and just looking at it would suggest approaching the study of the set by using fractal geometry.

Proposition 3.4 *The set* $\mathcal{C}$ *has* $\dim_H(\mathcal{C}) = \dim_B(\mathcal{C}) = \log 2/\log 3$ *and* $\dim_F(\mathcal{C}) = 0$.

We will calculate the Hausdorff dimension and the box counting dimension here. The result on the Fourier dimension is due to Kahane and Salem [31] and requires more work. However, we will calculate the Fourier transform of a particular measure on $\mathcal{C}$ and show that this does not decay.

For the upper bound on the upper box counting dimension, we will consider the obvious coverings of $\mathcal{C}$ by intervals obtained by fixing the first n coordinates of the elements in $\mathcal{C}$. There are 2^n such intervals, and each has length 3^{-n}. Hence, for $3^{-n} \leq \delta \leq 3^{-n+1}$, we find that $N_\delta(\mathcal{C}) \leq 2^n$. It follows that

$$\overline{\dim}_B(\mathcal{C}) = \limsup_{\delta \to 0} \frac{\log N_\delta(\mathcal{C})}{-\log \delta} \leq \limsup_{n \to \infty} \frac{\log 2^n}{\log 3^{n-1}} = \frac{\log 2}{\log 3}.$$

If we can get the same lower bound on the Hausdorff dimension, applying (3.6) would give us the first two equalities immediately. For this, we will apply the mass distribution principle, so we will need a probability measure.

Initially, we assign the mass 1 to the unit interval. We then divide the mass equally between the two intervals [0, 1/3] and [2/3, 1], so that each has measure 1/2. We continue in this way, at step k dividing the mass of each 'parent' interval equally between the two 'children'. The process converges to a measure supported on the Cantor set (as a sequence of measures in the weak-$*$ topology, but we skip the details). The resulting measure μ is known as the Cantor measure.

To prove that the Cantor measure is good for applying the mass distribution principle, we need an upper estimate on the measure of an interval. Let I be an interval of length <1. Pick an integer $n \geq 0$ such that $3^{-(n+1)} \leq \operatorname{diam}(I) < 3^{-n}$. In the n'th step of the Cantor construction, the minimum gap size is 3^{-n}. Hence, the interval can intersect at most one of the level n intervals, and so, setting $s = \log 2/\log 3$,

$$\mu(I) \leq 2^{-n} = 3^{-ns} \leq 3^s \operatorname{diam}(I)^s = 2 \operatorname{diam}(I)^s.$$

From the mass distribution principle of Lemma 3.3, it immediately follows that $\mathcal{H}^s(\mathcal{C}) \geq 1/2$, whence $\dim_H \mathcal{C} \geq s = \log 2/\log 3$. This completes the proof of the first part of the proposition.

To finish, we will calculate the Fourier transform of the specific Cantor measure μ constructed above. Weak-$*$ convergence of the auxiliary measures imply that for any continuous function f on [0, 1],

$$\int_0^1 f(x) d\mu(x) = \lim_{n \to \infty} 2^{-n} \sum_{a_1,\ldots,a_n \in \{0,2\}} f(a_1 3^{-1} + \cdots + a_n 3^{-n}),$$

so in order to find the Fourier transform of the measure, we need to evaluate the above expression for the function $f_t(x) = e^{-2\pi i t x}$. On inserting, we find that

$$\hat{\mu}(t) = \lim_{n \to \infty} 2^{-n} \sum_{a_1,\ldots,a_n \in \{0,2\}} e^{-2\pi i t (a_1 3^{-1} + \cdots + a_n 3^{-n})}. \tag{3.7}$$

Recalling Euler's formula for the cosine function,

$$\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2},$$

we find that

$$\prod_{k=1}^{n} \cos(2\pi 3^{-k} t) = \prod_{k=1}^{n} \frac{e^{2\pi i 3^{-k} t} + e^{-2\pi i 3^{-k} t}}{2}$$
$$= 2^{-n} \prod_{k=1}^{n} e^{2\pi i 3^{-k} t} \prod_{k=1}^{n} \left(e^{-2\pi i 3^{-k} 2t} + 1 \right).$$

Taking absolute values, the first product becomes of absolute value 1. Expanding the latter product, whatever remains becomes a partial sum from (3.7), so that letting n tend to infinity,

$$|\hat{\mu}(t)| = \left| \prod_{k=1}^{n} \cos(2\pi 3^{-k} t) \right|.$$

It should be clear that this does not decay polynomially with t, so certainly the Cantor measure is no good if we were to believe that the Cantor set had positive Fourier dimension. Of course, this is not the case anyway.

We conclude this chapter with some remarks connecting dynamics, fractals, measures, Diophantine approximation and numeration systems. This requires a bit of functional analysis. We refer the reader to [41] for an excellent textbook on the topic.

As we have seen, to both Diophantine properties and to base b expansions, we may associate dynamical systems on the unit interval. In the former case, this was the Gauss map, and in the latter the base b map. Both of these are ergodic with respect to measures which are absolutely continuous with respect to the Lebesgue measure, so almost all numbers are typical with respect to both of these measures. Our main problem is to take an atypical property from one and prove that this forces the other to be typical.

In terms of measures, invariant sets such as the ternary Cantor set give rise to other preserved measures than the Lebesgue measure (and similarly for sets invariant under the Gauss map). A quick-and-dirty way of constructing such measures is to take a point, look at its backward orbit and take a weak limit of averages of point measures along the orbit. By the Riesz representation theorem, these measures correspond to linear functionals in the unit ball of $C([0, 1])^*$, the dual space to the continuous functions on the interval with the topology of uniform convergence. The Banach-Alaoglu theorem ensures the existence of a limit point, which again by Riesz corresponds to a measure with its support on the orbit closure of the initial point. In this way, one may construct many invariant measures for a continuous transformation of the interval.

The set of invariant measures can easily be shown to be closed (hence compact) and convex. As such, it is spanned by its extremal points, and it is again an easy exercise to prove that these are exactly the measures with respect to which the transformation is ergodic. It then follows from the pointwise ergodic theorem that such measures must be mutually singular. In the cases considered above, the space of ergodic measures has an element which is absolutely continuous with respect to Lebesgue and a whole bunch of 'fractal' measures such as the Cantor measure.

In view of this, it would seem that our problems should boil down to considering nice, convex subsets of the Banach space $C([0, 1])^*$, and subsequently study the support of elements in their intersections, when interpreted as measures by the Riesz representation theorem. However easy this may sound, it really is horrible! Indeed, one can construct a simplex in an infinite dimensional space whose extremal points are dense within it (the Poulsen simplex [43]). Evidently, the existence of such monstrous sets makes life harder for us, but it also puts the study of Diophantine approximation into a much broader context with connections all over mathematics.

4 Higher Dimensional Problems

In the last section, we mentioned a major open problem in Diophantine approximation: *Are algebraic irrational numbers absolutely normal?* In order to approach this problem, we should address the concept of approximation by algebraic numbers. This is a higher dimensional problem, and we will approach it by considering rational approximation in higher dimensional spaces. The added flexibility of having more than one variable allows us to come up with new problems as well as to state analogues of old ones in higher dimension. As it turns out, some of the unsolved problems in one dimension can be resolved in higher dimension, while some problems which naturally live in higher dimensions remain unsolved.

As our starting point, we will derive some elementary results from the geometry of numbers (see [17, 39]). Let $S \subseteq \mathbb{R}^n$ be a centrally symmetric convex set, i.e. a set S such that if $x, y \in S$ then the line segment joining x and y is fully contained in S, and so that if $x \in S$, then $-x \in S$. Of course, such sets need not be Borel, as is easily seen by taking the open unit ball in $\mathbb{R}^2$ together with a non-measurable subset of the unit sphere. However convex sets do belong to the larger class of Lebesgue measurable sets and so have a well defined volume. A first question is, how large this volume can get before we are guaranteed the existence of a point different from the origin with integer coordinates in S. Clearly, 2^n is a lower bound, as is seen by considering the cube. As it turns out, this is best possible.

Theorem 4.1 *Let S be a convex, centrally symmetric body of volume strictly greater than 2^n. Then, S contains a point from $\mathbb{Z}^n \setminus \{0\}$.*

Proof First, consider the set $S' = \frac{1}{2}S$, i.e.

$$S' = \{x \in \mathbb{R}^n : 2x \in S\}.$$

This set has volume strictly greater than 1. We divide the set up into disjoint bits

$$S'_u = \{x \in \mathbb{R}^n : u_i \leq x_i < u_i + 1\} \cap S', \quad \text{where} \quad u = (u_1, \ldots, u_n) \in \mathbb{Z}^n.$$

Now, consider the sets $S''_u = S'_u - u \subseteq [0, 1)^n$. The sum of the volumes of the S''_u is strictly greater than 1, so two of the sets must overlap. Hence, there are distinct points $x', x'' \in S'$ and distinct points $u', u'' \in \mathbb{Z}^n$, such that $x' - x'' = u' - u'' = u \in \mathbb{Z}^n \setminus \{0\}$. But by convexity and central symmetry,

$$\tfrac{1}{2}x' - \tfrac{1}{2}x'' = \tfrac{1}{2}u \in S' = \tfrac{1}{2}S,$$

so that $u \in S$.

Note that if we further assume that S is closed, the inequality of the above theorem can be weakened to $\operatorname{vol}(S) \geq 2^n$ by a simple compactness argument. We can use Theorem 4.1 to provide solutions to systems of Diophantine inequalities.

Theorem 4.2 *Let $(a_{ij}) \in \mathrm{GL}(n, \mathbb{R})$ be some invertible matrix, let $c_1, \ldots, c_n > 0$, and consider the system of inequalities*

$$\left|\sum_{j=1}^{n} a_{1j}x_j\right| \leq c_1$$

$$\left|\sum_{j=1}^{n} a_{ij}x_j\right| < c_i, \quad 2 \leq i \leq n.$$

If $c_1 \cdots c_n \geq \left|\det(a_{ij})\right|$, this system has a non-trivial integer solution.

Proof It is straightforward to verify that the system of inequalities define a centrally symmetric convex set of volume $2^n c_1 \cdots c_n \left|\det(a_{ij})\right|^{-1}$. Thus, if $c_1 \cdots c_n > \left|\det(a_{ij})\right|$, the theorem is immediately implied by Theorem 4.1.

To get the full theorem, we first replace c_1 by $c_1 + \epsilon$ for some arbitrary $\epsilon \in (0, 1)$. By the above argument, there is an integer solution $x^{(\epsilon)} \in \mathbb{Z}^n$ for each ϵ, and furthermore, the corresponding convex sets are all bounded by a constant independent of ϵ. Consequently, there are only finitely many possible integer solutions to the system of equations, so one must occur for all ϵ_k in some sequence with $\epsilon_k \to 0$. This is the point we are looking for.

Now, let $x_1, \ldots, x_n \in \mathbb{R}$, let $N \in \mathbb{N}$ and define the matrix

$$(a_{ij}) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & -x_1 \\ 0 & 1 & \cdots & 0 & -x_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -x_n \end{pmatrix}.$$

Taking $c_1 = N^n$ and $c_2 = \cdots = c_{n+1} = 1/N$, the conditions of Theorem 4.2 are clearly satisfied. We have shown the following extension of Dirichlet's theorem.

Corollary 4.3 *Let $x_1, \dots, x_n \in \mathbb{R}$, let $N \in \mathbb{N}$. There are integers $p_1, \dots, p_n$ and q, $0 < q \leq N^n$ such that*

$$\left| x_i - \frac{p_i}{q} \right| < \frac{1}{qN}, \quad 1 \leq i \leq n.$$

Just as we did in the case of Dirichlet's theorem, we can derive a non-uniform version.

Corollary 4.4 *Let $x_1, \dots, x_n \in \mathbb{R}$. There are infinitely many tuples $(p_1, \dots, p_n) \in \mathbb{Z}^n$ and integers $q \in \mathbb{Z} \setminus \{0\}$, such that*

$$\left| x_i - \frac{p_i}{q} \right| < \frac{1}{q^{1+1/n}}, \quad 1 \leq i \leq n.$$

Considering instead the transpose of the above matrix,

$$(a_{ij}) = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & -x_n & \cdots & -x_2 & -x_1 \end{pmatrix}$$

with $c_1 = \cdots c_n = N$ and $c_{n+1} = N^{-n}$, we get another corollary.

Corollary 4.5 *Let $x_1, \dots, x_n \in \mathbb{R}$, let $N \in \mathbb{N}$. There are integers p and $q_1, \dots, q_n$, $0 < \max\{|q_i|\} \leq N$ such that*

$$|\mathbf{q} \cdot \mathbf{x} - p| < \frac{1}{N^n}.$$

Here and elsewhere, $\mathbf{x}$ denotes the vector with coordinates $(x_1, \dots, x_n)$.

Writing, as is usual in number theory, $\| \cdot \|$ for the distance to the nearest integer (or nearest vector with integer coordinates in sup-norm in higher dimension) and letting $|\mathbf{q}| = \max\{|q_i|\}$, the L^∞-norm of the vector $\mathbf{q}$, we get the following corollary.

Corollary 4.6 *Let $\mathbf{x} \in \mathbb{R}^n$. There are infinitely many $\mathbf{q} \in \mathbb{Z}^n \setminus \{0\}$, such that*

$$\|\mathbf{q} \cdot \mathbf{x}\| < |\mathbf{q}|^{-n}.$$

Several things stand out here. One is the duality between the two forms of Diophantine approximation, the simultaneous approximation and the 'linear form' approximation. Another is a little better hidden. Let $\xi \in \mathbb{R}$ and consider the vector

$\mathbf{x} = (\xi, \xi^2, \ldots, \xi^n)$. Feeding this vector into Corollary 4.6 gives us infinitely many integer polynomials taking small values at ξ.

The curve given by $\Gamma = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i = \xi^i, \xi \in \mathbb{R}\}$ is known as a Veronese curve, and much of the field known as Diophantine approximation on manifolds has its genesis in an attempt to understand the Diophantine properties of points on these curves and so the approximation properties of real numbers by algebraic numbers. This is a natural extension of the usual approximation by rational numbers, which is the case $n = 1$. Indeed, here a linear form has just one variable, so that one considers the quantity $\|q\xi\|$, which on dividing by q gives a rational approximation to the real number ξ.

For completeness, we should mention the two alternative ways of studying algebraic approximation (the book of Bugeaud [13] is an excellent resource). For this purpose, we introduce a little notation. Let $\mathbb{A}_n$ denote the set of real, algebraic numbers of degree at most n. For an integer polynomial P, let $H(P)$ denote the naive height of P, i.e. the maximum among the absolute values of the coefficients of P. Finally, for $\alpha \in \mathbb{A}_n$, let $H(\alpha)$ denote the height of the minimal integer polynomial of α. With these definitions, we introduce two families of Diophantine exponents,

$$w_n(\xi) = \sup\{w > 0 : 0 < |P(\xi)| < H(P)^{-w} \text{ for infinitely many } P \in \mathbb{Z}[X], \deg P \leq n\},$$

and

$$w_n^*(\xi) = \sup\{w > 0 : 0 < |\xi - \alpha| < H(\alpha)^{-w-1} \text{ for infinitely many } \alpha \in \mathbb{A}_n\}.$$

The two exponents were introduced in order to classify the transcendental numbers, a topic which we will not discuss in these notes. The first should be compared with Corollary 4.6, which almost immediately tells us that unless ξ is algebraic, $w_n(\xi) \geq n$ for all $\xi \in \mathbb{R}$. Indeed, we just apply the corollary directly to the vector $(\xi, \xi^2, \ldots, \xi^n)$. The only additional thing to take care of is the fact that $|\mathbf{q}|$ is not equal to $H(P)$, as the latter takes the constant term of P into account where the former does not. However, with ξ restricted to some bounded subset of $\mathbb{R}$, the two are comparable, and the resulting difference in definitions is absorbed by the *supremum* in the definition of $w_n(\xi)$. The two exponents are related, which is to be expected: if a polynomial takes a small value at ξ, it is not too unlikely that there is a root nearby, and conversely if α is an algebraic number close to ξ, then the minimal polynomial of α probably takes a small value at ξ.

We summarise some relations due to Wirsing [48] between the exponents in the following proposition.

Proposition 4.7 *For any n and any ξ, $w_n(\xi) \geq w_n^*(\xi)$. Furthermore, if ξ is not algebraic of degree at most n, the following inequalities hold:*

$$\begin{aligned} w_n^*(\xi) &\geq w_n(\xi) - n + 1 \\ w_n^*(\xi) &\geq \frac{w_n(\xi)+1}{2} \\ w_n^*(\xi) &\geq \frac{w_n(\xi)}{w_n(\xi)-n+1} \\ w_n^*(\xi) &\geq \frac{n}{4} + \frac{\sqrt{n^2+16n-8}}{4}. \end{aligned} \tag{4.1}$$

The two exponents need not be the same.

We will not prove the proposition here. However, there is an interesting point to be made. The relation between the exponents takes us into the world of transference theorems, which underlies the duality between simultaneous and linear forms approximation. We give a very general transference principle, from which many others can be derived (see [16]).

Theorem 4.8 *Consider two systems of l linearly independent linear forms, $(f_k(\mathbf{z}))$ and $(g_k(\mathbf{w}))$, all in l variables. Let $d = |\det(g_k)|$. Suppose the function*

$$\Phi(\mathbf{z}, \mathbf{w}) = \sum_k f_k(\mathbf{z}) g_k(\mathbf{w}),$$

has integer coefficients in all products of variables $z_i w_j$. If the system of inequalities

$$\max |f_k(\mathbf{z})| \leq \lambda$$

can be solved with $\mathbf{z} \in \mathbb{Z}^l \setminus \{0\}$, then so can the system of inequalities

$$\max |g_k(\mathbf{w})| \leq (l-1)(\lambda d)^{1/(l-1)}. \tag{4.2}$$

Proof As the forms (f_k) are linearly independent, the associated homogeneous system of equations only has the zero solution. Hence, for any solution to the first system, $\mathbf{z} \in \mathbb{Z}^n \setminus \{0\}$,

$$0 < \max |f_k(\mathbf{z})| \leq \lambda.$$

Since the right hand side of (4.2) decreases with λ, we may suppose that the last inequality above is actually an equality. Finally, we can permute the forms in order to make sure that the maximum is attained for the last form and change signs to remove the absolute value. In other words, we suppose without loss of generality that $\mathbf{z} \in \mathbb{Z}^l \setminus \{0\}$ is a solution to the initial system of inequalities with

$$\max |f_k(\mathbf{z})| = f_l(\mathbf{z}) = \lambda.$$

Filling these numbers into the expression for Φ, we get a linear form in the variables $\mathbf{w}$, which together with the first $l-1$ forms of the system $(g_k(\mathbf{w}))$ forms a system of linear forms. We may calculate its determinant, which turns out to be $f_l(\mathbf{z})d = \lambda d$. Using Theorem 4.2, the system

$$|\Phi(\mathbf{z}, \mathbf{w})| < 1, \quad |g_k(\mathbf{w})| < (\lambda d)^{1/(l-1)}, 1 \leq k \leq l-1,$$

has a non-zero integer solution $\mathbf{w}$. This certainly gives us the first $l-1$ inequalities of (4.2).

To get the final inequality, $|g_l(\mathbf{w})| < (l-1)(\lambda d)^{1/(l-i)}$, note that $\Phi(\mathbf{z}, \mathbf{w})$ is an integer by assumption, and so must be $= 0$. Hence,

$$|\lambda g_l(\mathbf{w})| = |f_l(\mathbf{z}) g_l(\mathbf{w})| = \left| -\sum_{k=1}^{l-1} f_k(\mathbf{z}) g_k(\mathbf{w}) \right| \leq \lambda(l-1)(\lambda d)^{1/(l-1)},$$

by the triangle inequality. This completes the proof.

This theorem explains why there is a relation between a system of inequalities given by a matrix and that given by its transpose, as seen in the following theorem.

Theorem 4.9 *Let (L_i) denote a system of n linear forms in m variables and let (M_j) denote the transposed system of m linear forms in n variables. Suppose that there is an integer solution $\mathbf{x} \neq 0$ to the inequalities*

$$\|L_i(\mathbf{x})\| \leq C, \quad |x_j| \leq X,$$

where $0 < C < 1 \leq X$. Then the system

$$\|M_j(\mathbf{u})\| \leq D, \quad |u_i| \leq U,$$

has a non-zero integer solution, where

$$D = (l-1) X^{(1-n)/(l-1)} C^{n/(l-1)}, \; U = (l-1) X^{m/(l-1)} C^{(1-m)/(l-1)}, \; l = m+n.$$

Proof We introduce new variables $\mathbf{y} = (y_1, \dots, y_n)$ and $\mathbf{v} = (v_1, \dots, v_m)$ to capture the nearest integers in the systems. Hence, we define two new systems of l linear forms in l variables

$$f_k(\mathbf{x}, \mathbf{y}) = \begin{cases} C^{-1}(L_k(\mathbf{x}) + y_k) & 1 \leq k \leq n \\ X^{-1} x_{k-n} & n < k \leq l \end{cases}$$

and

$$g_k(\mathbf{u}, \mathbf{v}) = \begin{cases} C u_k & 1 \leq k \leq n \\ X(-M_{k-n}(\mathbf{u}) + v_{k-n}) & nk <\leq l \end{cases}.$$

It is easily checked that the conditions of Theorem 4.8 hold true with $d = C^n X^m$. Applying this theorem gives a non-zero integer solution $((u), (v))$.

It remains for us to check that $\mathbf{u} \neq 0$, but this is easy: If $D < 1$ and $\mathbf{u} = 0$, the inequalities resulting from Theorem 4.8 would force $\mathbf{v} = 0$, which contradicts the initial conclusion. Hence, $\mathbf{u} \neq 0$ as required or $D \geq 1$, in which case it is trivial to solve the inequalities.

At this point, let us define some sets which will be of great importance in the next section. The set of badly approximable numbers is the set

$$\mathrm{Bad} = \left\{ x \in \mathbb{R} : \text{For some } C(x) > 0, \left| x - \frac{p}{q} \right| \geq \frac{C(x)}{q^2} \text{ for all } \frac{p}{q} \in \mathbb{Q} \right\}.$$

From Khintchine's theorem, this set is Lebesgue null. From the theory of continued fractions, it is also the set of numbers with bounded partial quotients, so the same conclusion follows immediately from the ergodicity of the Gauss map.

Similarly to the one dimensional case, one can define the sets (also denoted Bad by abuse of notation),

$$\mathrm{Bad} = \left\{ \mathbf{x} \in \mathbb{R}^n : \text{for some } C(\mathbf{x}) > 0,\ \|q\mathbf{x}\| \geq \frac{C(\mathbf{x})}{q^{1/n}} \text{ for all } q \in \mathbb{Z} \setminus \{0\} \right\}.$$

Or the corresponding linear forms version,

$$\mathrm{Bad}^* = \left\{ \mathbf{x} \in \mathbb{R}^n : \text{for some } C(\mathbf{x}) > 0,\ \|\mathbf{q} \cdot \mathbf{x}\| \geq \frac{C(\mathbf{x})}{|\mathbf{q}|^n} \text{ for all } \mathbf{q} \in \mathbb{Z}^n \setminus \{0\} \right\}.$$

Corollary 4.10 *The sets* Bad *and* Bad* *are the same.*

Proof The proof is just an application of Theorem 4.9. For $\mathbf{x} \in \mathbb{R}^n$, define the linear form $M(\mathbf{t}) = \mathbf{x} \cdot \mathbf{t}$ in n variables, and let $L_i(u) = t_i u$ denote the transposed system of n linear forms in 1 variable. Clearly, $\mathbf{x} \in \mathrm{Bad}^*$ if and only if

$$\|M(\mathbf{t}) \max\{|t_i|\}^n \geq c^*, \tag{4.3}$$

where $c^* > 0$ depends only on $\mathbf{x}$ for all non-zero integer vectors $\mathbf{t}$. Similarly, $\mathbf{x} \in \mathrm{Bad}$ if and only if

$$\max\{\|L_i(u)\|\}^n |u| \geq c,$$

where $c > 0$ depends only on $\mathbf{x}$ for all non-zero integers u.

Suppose $\mathbf{x} \in \mathrm{Bad}^*$. We will prove that $\mathbf{x} \in \mathrm{Bad}$, so let $u \neq 0$ be an integer. Let $X = U$ and $C \geq \max\{\|L_i(u)\|\}$ with $0 < C < 1$, as otherwise there is nothing to prove. As $\mathbf{x} \in \mathrm{Bad}^*$, the values of D and U of Theorem 4.9 must satisfy that $DU^n \geq c^*$, as otherwise we would have a contradiction to (4.3). However, with the relations of the theorem,

$$DU^n = nX^{(1-n)/n} C \left(nX^{1/n}\right)^n = n^{n+1} C X^{1/n},$$

so that

$$C^n X = \left(n^{-(n+1)} D U^n\right)^n \geq n^{-n(n+1)} c^{*n}.$$

This shows that if c^* is positive and exists, then $c = n^{-n(n+1)} c^{*n}$ will work as a constant to prove that $\mathbf{x} \in \text{Bad}$. The converse is symmetrical.

Many other nice results follow from the transference technique. It is of interest to note that the above transference inequalities become equalities only at the critical exponent derived from the Dirichlet type theorems, as seen from the above corollary. This is also the case for the inequalities between the exponents of algebraic approximation, where the critical exponent for both variants is n (see (4.1)), where the inequalities become again become equalities. Transference theorems generally reveal less information away from the critical exponent.

It would be natural to conjecture that results similar to the Khintchine theorem should hold true for algebraic approximation or at least in some form for the ambient space containing a given Veronese curve. This is in fact the case, but the lack of a good analogue of continued fractions in higher dimensions is an obstacle for the methods already used to be applicable. We will give a sketch of a geometrical proof of a Khintchine type theorem for a single linear form, which is in a sense stronger than its one-dimensional analogue, as it does not assume the approximation function to be monotonic when the number of variables is at least 3. We will then discuss the monotonicity assumption in one dimension.

For the purposes of the proof, we will need a converse to the Borel–Cantelli lemma, which we state without proof. A lower bound on the measure of such a set may be found using the following lemma (see e.g. [47])

Lemma 4.11 *Let $(\Omega, \mathcal{B}, \mu)$ be a probability space and let E_n be a sequence of events. Suppose that $\sum \mu(E_n) = \infty$. Then,*

$$\mu(\limsup E_n) \geq \limsup_{Q \to \infty} \frac{\left(\sum_{n=1}^{Q} \mu(E_n)\right)^2}{\sum_{n,m=1}^{Q} \mu(E_m \cap E_n)}.$$

In particular, if the events E_n are pairwise independent, $\mu(\limsup E_n) = 1$.

We will consider the set

$$W_n(\psi) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{q} \cdot \mathbf{x}\| < \psi(|\mathbf{q}|) \text{ for infinitely many } \mathbf{q} \in \mathbb{Z}^n\}.$$

In this notation, the set originally considered in the first section would be

$$W_1(\psi) = \{x \in \mathbb{R} : \|qx\| < \psi(|q|) \text{ for infinitely many } q \in \mathbb{Z}\},$$

so it is a natural generalisation. In the case of $W_1(\psi)$, we showed that the set is full provided $\sum \psi(q) = \infty$ (note the change in condition due to the change in definition), and *provided the function $q\psi(q)$ was monotonic*.

Theorem 4.12 *Let $\psi : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be some function, with $q\psi(q)$ monotonic if $m = 1, 2$. Then, the set $W_m(\psi)$ is full if $\sum q^{m-1}\psi(q) = \infty$. If the series converges, the set $W_m(\psi)$ is null.*

We will follow a proof given by Dodson [19]. As in the case of Khintchine's theorem, we will make some restrictions. We will consider only points in the unit square $[0, 1)^m$, which we will think of as a torus by identifying the edges. We will think of $W_m(\psi) \cap [0, 1)^m$ as a *limsup*-set, so for a fixed $\mathbf{q} \in \mathbb{Z}^m$ let

$$E_{\mathbf{q}} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{q} \cdot \mathbf{x}\| < \psi(|\mathbf{q}|)\},$$

so that

$$W_m(\psi) = \limsup E_{\mathbf{q}}.$$

We call the sets $E_{\mathbf{q}}$ resonant sets due to a connection with physics, which we will not explore here.

Lemma 4.13 *For each $\mathbf{q} \in \mathbb{Z}^m$, $|E_{\mathbf{q}}| \asymp \psi(|\mathbf{q}|)$.*

Sketch of proof. We sketch the argument for $m = 2$. This is a simple geometric argument. The set $E_{\mathbf{q}}$ consists of a bunch of parallel strips. Considering only the central lines of these, i.e. the solution curves to $q_1x + q_2y = p$ in the unit square, and matching up the sides of the square to form a torus, we obtain a closed geodesic curve on the torus. The set $E_{\mathbf{q}}$ forms a tubular neighbourhood of this geodesic. Calculating the length (roughly $|\mathbf{q}|$) and width of this strip (roughly $\psi(|\mathbf{q}|)|\mathbf{q}|^{-1}$), we arrive at the conclusion.

Lemma 4.14 *Suppose the vectors $\mathbf{q}, \mathbf{q}' \in \mathbb{Z}^m$ are linearly independent over $\mathbb{R}$. Then the corresponding resonant sets are independent in the sense of probability, i.e. $|E_{\mathbf{q}} \cap E_{\mathbf{q}'}| = |E_{\mathbf{q}}||E_{\mathbf{q}'}|$.*

Sketch of proof. Once more, we give a sketch for $m = 2$. Consider again the central geodesics of the two resonant sets. As the vectors $\mathbf{q}$ and $\mathbf{q}'$ are linearly independent, these tesselate the torus into parallelograms. There will be $|\det(\mathbf{q}, \mathbf{q}')|$ such parallelograms, where $(\mathbf{q}, \mathbf{q}')$ denotes the matrix with columns $\mathbf{q}$ and $\mathbf{q}'$.

Consider now the tubular neighbourhoods and their intersections. These will consist of a union of scaled copies of the parallelograms of the tessellation. Calculating their individual sizes as before will give the required result.

Proof of Theorem 4.12. The convergence statement is easy. The set $W_m(\psi)$ is covered by the set

$$\bigcup_{n \geq k} \bigcup_{|\mathbf{q}| \geq k} E_{\mathbf{q}},$$

so as there are roughly Q^{m-1} vectors $\mathbf{q} \in \mathbb{Z}^m$ with $|\mathbf{q}| = Q$, Lemma 4.13 immediately gives us that

$$\left| \bigcup_{n\geq k} \bigcup_{|\mathbf{q}|\geq k} E_{\mathbf{q}} \right| \ll \sum_{n\geq k} n^{m-1}\psi(n),$$

which is a tail of a convergent series.

To get the divergence half of the statement, we show that a subset has full measure. The key is to pick a sufficiently rich collection of integer vectors for which the thinned out series (the volume sum) still diverges, but for which any pair of vectors is linearly independent. Define the sets

$$S_k = \{\mathbf{q} \in \mathbb{Z}^m : \mathbf{q} \text{ is primitive}, q_m \geq 1, |\mathbf{q}| = k\}.$$

We will consider only vectors in $P = \cup S_k$.

If $\mathbf{q}, \mathbf{q}' \in P$ satisfy a linear dependence, then for some integer v, we must have $\mathbf{q} = v\mathbf{q}'$ (or the converse). It follows that v divides all the coordinates of $\mathbf{q}$, so by primitivity, $v = \pm 1$. But since the last coordinates are positive, we must have $v = 1$, whence $\mathbf{q} = \mathbf{q}'$. In other words, any pair of vectors $\mathbf{q}, \mathbf{q}' \in P$ are linearly independent, and so by Lemma 4.14 we have $|E_{\mathbf{q}} \cap E_{\mathbf{q}'}| = |E_{\mathbf{q}}||E_{\mathbf{q}'}|$.

To apply Lemma 4.11, we must ensure that the volume sum still diverges when restricted to a sum over P. In order to accomplish this, we require an asymptotic formula for the number of elements in S_k. But this is not difficult.

$$\begin{aligned}\#S_k &= \sum_{\substack{|\mathbf{q}|=k, q_m\geq 1\\(q_1,\dots,q_m)=1}} 1 = \sum_{\substack{|\mathbf{q}|=k, q_m\geq 1\\(q_1,\dots,q_m)=h}} \sum_{d|h} \mu(d) = \sum_{d|k} \mu(d) \sum_{\substack{|\mathbf{r}|=k/d\\r_m\geq 1}} 1\\ &= 2^{m-2}(2m-1) \sum_{d|k} \mu(d) \left(\tfrac{k}{d}\right)^{m-1} + \text{error term.}\end{aligned}$$

Here, μ denotes the Möbius function, and we have used the classical fact that $\sum_{d|n} \mu(d)$ is equal to one for $n = 1$ and equal to zero otherwise. For $m = 2$, the main term is $= 3\phi(k)$, where ϕ denotes the Euler ϕ-function. For $m \geq 3$, we have

$$\sum_{d|k} \mu(d) \left(\frac{1}{d}\right)^{m-1} = \prod_{p|k} \left(1 - \frac{1}{p^{m-1}}\right),$$

which lies between $\zeta(m-1)^{-1}$ and 1, where ζ denotes the Riemann ζ-function.

The upshot is that for $m \geq 3$, S_k contains a constant times k^{m-1} elements, and the divergence of the original series implies the divergence of the restricted series without further work. For $m = 2$ we need to average out the irregularities of the Euler function, but using the classical estimate $\sum_{n\leq N} \phi(n) = \frac{3}{\pi^2}N^2 + O(N \log N)$, we may apply Cauchy condensation over 2-adic blocks to get the divergence of the new series. This however requires the monotonicity of the function.

Quite a few remarks should be made at this point. Firstly, the result is valid even more generally than the one stated here. The full Khintchine–Groshev theorem concerns systems of linear forms, and states the set of matrices

$$W_{m,n}(\psi) = \{A \in \mathrm{Mat}_{m\times n}(\mathbb{R}) : \|\mathbf{q}A\| < \psi(|\mathbf{q}|) \text{ for infinitely many } \mathbf{q} \in \mathbb{Z}^m\},$$

is null or full according to the convergence or divergence of the series $\sum q^{m-1}\psi(q)^n$.

Secondly, the divergence assumption is not needed except for in one case. Namely, in the case $m = n = 1$, an explicit counterexample to the classical Khintchine theorem without assumption of monotonicity can be given. However, conjectures do exists, which tell us what to expect. It is natural in this context to impose the restriction that the approximating rationals should be on lowest terms. The Duffin–Schaeffer conjecture [21] states that the set

$$\left\{x \in \mathbb{R} : \left|x - \frac{p}{q}\right| < \psi(q) \text{ for infinitely many coprime } (p,q) \in \mathbb{Z}\times\mathbb{N}\right\},$$

should be null or full according to the convergence or divergence of the series $\sum \phi(q)\psi(q)$.

The convergence half is easy, and in the case of an appropriately monotonic approximation function, the result follows immediately from condensation and Khintchine's theorem. The difficulty is in getting the result for non-monotonic error functions. However, it is known that the set must be either null or full. It is hence tempting to try to apply Lemma 4.11 to get positive measure and proceed to deduce full measure from this law. However, controlling the intersections appears to be beyond the reach of current methods. What is clear is that it is hopeless to control the individual intersections, and the entire sum must be considered at least in very long blocks at a time.

Thirdly, as in the case when the above results give a null set, it is natural to ask what the Hausdorff dimension should be. As in the case of a single number, it is easy to get an upper bound, at least in the case $\psi(q) = q^{-v}$. Just applying a covering argument in the spirit of the convergence case, one finds that in this case, $\dim_{\mathrm{H}}(W_{m,n}(q \mapsto q^{-v})) \leq (m-1)n + (m+n)/(1+v)$. The initial integer comes from the hyperplanes central to the resonant sets, with the fraction at the end being the really interesting component. In fact, this is sharp, and we will return to these types of estimates in the final section of the notes.

Fourthly, we did not answer the question we originally asked. Namely, we were interested in points on the Veronese curves and in a final instance in points in a fractal subset of the Veronese curve. For the full Veronese curves, these things can be done, but with considerably more difficulty. Even the convergence case was not settled before 1989 by Bernik [9] with the divergence case taking another 10 years before being settled by Beresnevich [4]. By contrast, the above results about the ambient space date back to 1938.

As a final remark on the Khintchine–Groshev theorem, in the case when one considers simultaneous approximation or more than one linear form, similar questions can be asked when different rates of approximation are required in the different variables. Again, the questions can be answered under some assumptions on the approximating functions, and whether sets arising in this way are null or full depend once again on the convergence or divergence of a certain series.

We dwell a little on this last point. Consider again Corollary 4.4, this time with $n = 2$. Littlewood suggested multiplying the two inequalities instead of considering them separately, and so to consider

$$\left|x - \frac{p_1}{q}\right| \left|y - \frac{p_2}{q}\right| < \frac{1}{q^3}$$

or more concisely,

$$q\|qx\|\,\|qy\| < 1. \tag{4.4}$$

By Corollary 4.4, this inequality always has infinitely many solutions for any pair (x, y), but it is a little more flexible than the original one. Indeed, one approximation could be pretty bad indeed, just as long as the other one is very good, and the inequality would still hold.

From the theory of continued fractions, we know that many numbers x exist which have $q\,\|qx\| > C > 0$ for all q (the badly approximable numbers, or equivalently those with bounded partial quotients), and similarly we know that there are badly approximable pairs, i.e. pairs for which Corollary 4.4 cannot be improved beyond a positive constant (we will prove this and much more in the next section). Littlewood asked whether there are pairs such that (4.4) cannot be improved beyond a constant. Due to the added flexibility, he conjectured that this should not be the case, so that for any pair (x, y),

$$\liminf_{q\to\infty} q\|qx\|\,\|qy\| = 0, \tag{4.5}$$

where the liminf is taken over positive integers q.

Equation (4.5) is the Littlewood conjecture. Littlewood apparently did not think that it should be too difficult, and set it as an exercise to his students in the thirties. To date, it is an important unsolved problem in Diophantine approximation. This is a case, where many of the known results are metric. Probably the most famous among them is the result of Einsiedler, Katok and Lindenstrauss [23], which states that the set of exceptions (x, y) to (4.5) must lie in a countable union of sets of box counting dimension zero. From the elementary properties of Hausdorff dimension together with (3.6), it follows that both the Hausdorff dimension and the Fourier dimension are also equal to zero.

In fact, their approach follows the approach to continued fractions via the geodesic flow outlined in the first section. There is no good analogue of continued fractions in higher dimension, but an analogue of the geodesic flow on $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ is certainly constructible. In the classical, one-dimensional case, the geodesic flow is given by the action of the diagonal subgroup of $\mathrm{SL}_2(\mathbb{R})$, and badly approximable numbers correspond to geodesics which remain in a compact subset of the space $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. The approach of Einsiedler, Katok and Lindenstrauss works instead with the diagonal subgroup of $\mathrm{SL}_3(\mathbb{R})$, acting on $\mathrm{SL}_3(\mathbb{R})/\mathrm{SL}_3(\mathbb{Z})$.

This action is a two-parameter flow and its dynamics is very complicated. Nonetheless, the three authors manage to prove many things about the simplex of preserved measures described in the final section, and in particular about the extremal points. They show that the only possible ergodic measures are the Haar measure and measures with respect to which every one-parameter subgroup of the diagonal group acts with zero entropy. Readers acquainted with the notion of entropy should be able to see how this will have an impact on the box counting dimension of the support of the measure.

The relation between the flow and the Littlewood conjecture is a little technical, but briefly the pair (x, y) satisfies the Littlewood conjecture if and only if the orbit of the point

$$\begin{pmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & 0 & 1 \end{pmatrix} \mathrm{SL}_3(\mathbb{Z})$$

is unbounded under the action of the semigroup

$$A^+ = \left\{ \begin{pmatrix} e^{-s-t} & 0 & 0 \\ 0 & e^s & 0 \\ 0 & 0 & e^t \end{pmatrix} : s, t \in \mathbb{R}_+ \right\}.$$

Hence, the set of exceptions can be embedded into a set of points in $\mathrm{SL}_3(\mathbb{R}) / \mathrm{SL}_3(\mathbb{Z})$ with bounded A^+-orbits, and the result can be deduced from the dynamical statement. It is very impressive work. We will say something non-trivial but somewhat easier in the final section of these notes.

5 Badly Approximable Elements

In this section, we will discuss badly approximable numbers, and in fact do so in higher dimensions. A consequence of our main result is Jarník's theorem [29]: the set of badly approximable numbers has maximal Hausdorff dimension. However we will prove much more, including the result that badly approximable numbers form a set of maximal dimension inside the Cantor set. The latter relates digital properties with Diophantine properties, and although it does not resolve the question of absolute normality of algebraic irrational numbers, it does provide information on which Diophantine properties a number failing spectacularly at being normal to some base can have.

The work presented in this section originated in an unfortunately failed attempt to resolve the Schmidt conjecture with Thorn and Velani [36]. We did solve other problems in the process, though. In order to present this, we need some new sets. For $i, j \geq 0$ with $i + j = 1$, denote by $\mathrm{Bad}(i, j)$ the set of (i, j)–badly approximable pairs $(x_1, x_2) \in \mathbb{R}^2$; that is $(x_1, x_2) \in \mathrm{Bad}(i, j)$ if there exists a positive constant $c(x_1, x_2)$ such that for all $q \in \mathbb{N}$

$$\max\{\|qx_1\|^{1/i}, \|qx_2\|^{1/j}\} > c(x_1, x_2)\, q^{-1}.$$

In the case $i = j = 1/2$, the set is simply the standard set of badly approximable pairs or equivalently as we saw in the last section the set of badly approximable linear forms in two variables. If $i = 0$ we identify the set $\mathrm{Bad}(0, 1)$ with $\mathbb{R} \times \mathrm{Bad}$ where Bad is the set of badly approximable numbers. That is, $\mathrm{Bad}(0, 1)$ consists of pairs (x_1, x_2) with $x_1 \in \mathbb{R}$ and $x_2 \in \mathrm{Bad}$. The roles of x_1 and x_2 are reversed if $j = 0$. In full generality, Schmidt's conjecture states that $\mathrm{Bad}(i, j) \cap \mathrm{Bad}(i', j') \neq \emptyset$. It is a simple exercise to show that if Schmidt's conjecture is false for some pairs (i, j) and (i', j') then Littlewood's conjecture in simultaneous Diophantine approximation is true.

The Schmidt conjecture was recently settled in the affirmative by Badziahin, Pollington and Velani [3], who established a stronger version. An [1] subsequently proved an even stronger result, which we remark on towards the end of this section.

We will set up a scary generalisation of the sets $\mathrm{Bad}(i, j)$. For the purposes of these notes, we will consider general metric spaces. The examples to keep in mind are nice fractal subsets of Euclidean space, such as the Cantor set or the Sierpiński gasket. Let (X, d) be the product space of t metric spaces (X_i, d_i) and let (Ω, d) be a compact subspace of X which contains the support of a non-atomic finite measure m.

Let $\mathcal{R} = \{R_\alpha \subseteq X : \alpha \in J\}$ be a family of subsets R_α of X indexed by an infinite, countable set J. Thus, each resonant set R_α can be split into its t components $R_{\alpha,i} \subset (X_i, d_i)$. Let $\beta : J \to \mathbb{R}_+ : \alpha \to \beta_\alpha$ be a positive function on J and assume that the number of $\alpha \in J$ with β_α bounded from above is finite. We think of these as the resonant sets similar to the central lines of the resonant neighbourhoods considered in the last section.

For each $1 \leq i \leq t$, let $\rho_i : \mathbb{R}_+ \to \mathbb{R}_+ : r \to \rho_i(r)$ be a real, positive function such that $\rho_i(r) \to 0$ as $r \to \infty$ and that ρ_i is decreasing for r large enough. Furthermore, assume that $\rho_1(r) \geq \rho_2(r) \geq \cdots \geq \rho_t(r)$ for r large – the ordering is irrelevant. Given a resonant set R_α, let

$$F_\alpha(\rho_1, \ldots, \rho_t) = \{x \in X \colon d_i(x_i, R_{\alpha,i}) \leq \rho_i(\beta_\alpha) \quad \text{for all } \ 1 \leq i \leq t\}$$

denote the 'rectangular' $(\rho_1, \ldots, \rho_t)$–neighbourhood of R_α. For a real number $c > 0$, we will define the scaled rectangle,

$$cF_\alpha(\rho_1, \ldots, \rho_t) = \{x \in X \colon d_i(x_i, R_{\alpha,i}) \leq c\rho_i(\beta_\alpha) \quad \text{for all } \ 1 \leq i \leq t\},$$

and similarly for other rectangular regions throughout this section. Consider the set

$$\begin{aligned} &\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t) \\ &\qquad = \{x \in \Omega : \exists\, c(x) > 0 \ \text{ s.t. } x \notin c(x)\, F_\alpha(\rho_1, \ldots, \rho_t) \ \text{ for all } \ \alpha \in J\}. \end{aligned}$$

Thus, $x \in \mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t)$ if there exists a constant $c(x) > 0$ such that for all $\alpha \in J$,

$$d_i(x_i, R_{\alpha,i}) \ \geq \ c(k)\, \rho_i(\beta_\alpha) \quad \text{for some } 1 \leq i \leq t.$$

We wish to find a suitably general framework which gives a lower bound for the Hausdorff dimension of $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t)$. Without loss of generality we shall assume that $\sup_{\alpha \in J} \rho_i(\beta_\alpha)$ is finite for each i – otherwise $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t) = \emptyset$ and there is nothing to prove.

Given $l_1, \ldots, l_t \in \mathbb{R}_+$ and $c \in \Omega$ let

$$F(c; l_1, \ldots, l_t) = \{x \in X \ : \ d_i(x_i, c_i) \leq l_i \quad \text{for all } \ 1 \leq i \leq t\}$$

denote the closed 'rectangle' centred at c with 'sidelengths' determined by $l_1, \ldots, l_t$. Also, for any $k > 1$ and $n \in \mathbb{N}$, let F_n denote any generic rectangle intersected with Ω, i.e. a set of the form $F(c; \rho_1(k^n), \ldots, \rho_t(k^n)) \cap \Omega$ in Ω centred at a point c in Ω. As before, $B(c, r)$ is a closed ball with centre c and radius r. The following conditions on the measure m and the functions ρ_i will play a central role in our general framework.

(A) There exists a strictly positive constant δ such that for any $c \in \Omega$

$$\liminf_{r \to 0} \frac{\log m(B(c, r))}{\log r} = \delta.$$

It is easily verified from the Mass distribution principle of Lemma 3.3 that if the measure m supported on Ω is of this type, then $\dim \Omega \geq \delta$ *and so* $\dim X \geq \delta$.

(B) *For $k > 1$ sufficiently large, any integer $n \geq 1$ and any $i \in \{1, \ldots, t\}$,*

$$\lambda_i^l(k) \leq \frac{\rho_i(k^n)}{\rho_i(k^{n+1})} \leq \lambda_i^u(k),$$

where λ_i^l and λ_i^u are lower and upper bounds depending only on k but not on n,such that $\lambda_i^l(k) \to \infty$ as $k \to \infty$.

(C) *There exist constants $0 < a \leq 1 \leq b$ and $l_0 > 0$ such that*

$$a \leq \frac{m(F(c; l_1, \ldots, l_t))}{m(F(c'; l_1, \ldots, l_t))} \leq b$$

for any $c, c' \in \Omega$ and any $l_1, \ldots, l_t \leq l_0$. This condition implies that rectangles of the same size centred at points of Ω have comparable m-measure.

(D) *There exist strictly positive constants D and l_0 such that*

$$\frac{m(2\,F(c; l_1, \ldots, l_t))}{m(F(c; l_1, \ldots, l_t))} \leq D$$

for any $c \in \Omega$ and any $l_1, \ldots, l_t \leq l_0$. This condition simply says that the measure m is 'doubling' with respect to rectangles.

(E) *For $k > 1$ sufficiently large and any integer $n \geq 1$*

$$\frac{m(F_n)}{m(F_{n+1})} \geq \lambda(k),$$

where λ is a function depending only on k such that $\lambda(k) \to \infty$ as $k \to \infty$.

In terms of achieving a lower bound for $\dim \mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \dots \rho_t)$, the above four conditions are rather natural. The following final condition is in some sense the only genuine technical condition and is not particularly restrictive.

We should state at this point that if m is a product measure of measures satisfying the decay condition that there exist strictly positive constants δ and r_0 such that for $c \in \Omega$ and $r \leq r_0$

$$a\, r^{\delta} \leq \quad m(B(c, r)) \quad \leq b\, r^{\delta}\ , \tag{5.1}$$

where $0 < a \leq 1 \leq b$ are constants independent of the ball, then the product measure satisfies all conditions above. This is extremely useful, and missing digit sets have this property, as do all regular Cantor sets.

Theorem 5.1 *Let (X, d) be the Cartesian product space of the metric spaces $(X_1, d_1), \dots, (X_t, d_t)$ and let (Ω, d, m) be a compact measure subspace of X. Let the measure m and the functions ρ_i satisfy conditions (A) to (E). For $k \geq k_0 > 1$, suppose there exists some $\theta \in \mathbb{R}_+$ so that for $n \geq 1$ and any rectangle F_n there exists a disjoint collection $\mathcal{C}(\theta F_n)$ of rectangles $2\theta F_{n+1}$ contained within θF_n satisfying*

$$\#\mathcal{C}(\theta F_n) \geq \kappa_1 \frac{m(\theta F_n)}{m(\theta F_{n+1})} \tag{5.2}$$

and

$$\begin{aligned}\#\left\{2\theta F_{n+1} \subset \mathcal{C}(\theta F_n) : \min_{\alpha \in J(n+1)} d_i(c_i, R_{\alpha,i}) \leq 2\theta \rho_i(k^{n+1}) \ \textit{for any}\ 1 \leq i \leq t\right\}& \\ \leq \kappa_2 \frac{m(\theta F_n)}{m(\theta F_{n+1})}&.\end{aligned} \tag{5.3}$$

where $0 < \kappa_2 < \kappa_1$ are absolute constants independent of k and n. Furthermore, suppose $\dim_H(\cup_{\alpha \in J} R_\alpha) < \delta$. *Then*

$$\dim_H \mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \dots, \rho_t) \geq \delta.$$

The statement of the theorem with all its assumptions is pretty bad, and the proof is in fact a rather dull affair. We give a short sketch. Fixing $k \geq k_0$, the conditions of the theorem give us a way to construct a Cantor type set inside $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \dots, \rho_t)$. Namely, we begin with a rectangle θF_1. In this and any subsequent step, we take out the collection $\mathcal{C}(\theta F_n)$, which is pretty big due to (5.2). The points in the rectangles from (5.3) will have some difficulties lying in the set $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \dots, \rho_t)$, as they are fairly close to a resonant set from $J(n+1)$. Hence, we discard them and retain a collection $\mathcal{F}_{n+1}(\theta F_n)$ of closed rectangles. The assumption (5.3) tells us that a positive proportion of the collection will remain, and we continue in this way to get a Cantor set constructed from rectangles. From the construction, this set is contained in $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t)$, and in fact the unspecified constant in the definition of the set can in all cases be chosen to be $c(k) = \min_{1\le i\le t}(\theta/\lambda_i^u(k))$. We call the set $\mathbf{K}_{c(k)}$.

We now construct a probability measure on the Cantor set recursively. For any rectangle θF_n in $\mathcal{F}_n$ we attach a weight $\mu(\theta F_n)$ which is defined recursively as follows: for $n = 1$,

$$\mu(\theta F_1) = \frac{1}{\#\mathcal{F}_1} = 1$$

and for $n \ge 2$,

$$\mu(\theta F_n) = \frac{1}{\#\mathcal{F}_n(\theta F_{n-1})}\,\mu(\theta F_{n-1}) \quad (F_n \subset F_{n-1}).$$

This procedure thus defines inductively a mass on any rectangle used in the construction of $\mathbf{K}_{c(k)}$. In fact a lot more is true: μ can be further extended to all Borel subsets A of Ω to determine $\mu(A)$ so that μ constructed as above actually defines a measure supported on $\mathbf{K}_{c(k)}$. The probability measure μ constructed above is supported on $\mathbf{K}_{c(k)}$ and for any Borel subset A of Ω

$$\mu(A) = \inf \sum_{F\in\mathcal{F}} \mu(F),$$

where the infimum is taken over all coverings $\mathcal{F}$ of A by rectangles $F \in \{\mathcal{F}_n : n \ge 1\}$.

The mass distribution principle of Lemma 3.3 can then be applied to this measure to find that $\dim \mathbf{K}_{c(k)} \ge \delta - 2\epsilon(k)$, where $\epsilon(k)$ tends to zero as k tends to infinity. To conclude, we let k do this, and so have constructed a subset of $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots, \rho_t)$ whose dimension is lower bounded by δ. All the technical assumptions occur naturally in the process of constructing the set and applying the mass distribution principle.

The remarks preceding the statement of Theorem 5.1 immediately gives us the following version, which is of more use to us.

Theorem 5.2 *For $1 \le i \le t$, let (X_i, d_i) be a metric space and (Ω_i, d_i, m_i) be a compact measure subspace of X_i where the measure m_i satisfies (5.1) with exponent δ_i. Let (X, d) be the product space of the spaces (X_i, d_i) and let (Ω, d, m) be the product measure space of the measure spaces (Ω_i, d_i, m_i). Let the functions ρ_i satisfy condition (B). For $k \ge k_0 > 1$, suppose there exists some $\theta \in \mathbb{R}_+$ so that for $n \ge 1$ and any rectangle F_n there exists a disjoint collection $\mathcal{C}(\theta F_n)$ of rectangles $2\theta F_{n+1}$ contained within θF_n satisfying*

$$\#\mathcal{C}(\theta F_n) \ \geq \ \kappa_1 \prod_{i=1}^{t} \left(\frac{\rho_i(k^n)}{\rho_i(k^{n+1})} \right)^{\delta_i} \tag{5.4}$$

and

$$\#\left\{ 2\theta F_{n+1} \subset \mathcal{C}(\theta F_n) : \min_{\alpha \in J(n+1)} d_i(c_i, R_{\alpha,i}) \leq 2\theta \rho_i(k^{n+1}) \textit{ for any } 1 \leq i \leq t \right\}$$
$$\leq \kappa_2 \prod_{i=1}^{t} \left(\frac{\rho_i(k^n)}{\rho_i(k^{n+1})} \right)^{\delta_i}, \tag{5.5}$$

where $0 < \kappa_2 < \kappa_1$ *are absolute constants independent of* k *and* n*. Furthermore, suppose* $\dim_H(\cup_{\alpha \in J} R_\alpha) < \sum_{i=1}^{t} \delta_i$*. Then*

$$\dim_H \mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \ldots \rho_t) = \sum_{i=1}^{t} \delta_i.$$

Note that while Theorem 5.1 will only give the lower bound on the Hausdorff dimension in the last equation, the upper bound is a consequence of the assumptions. Indeed, any set satisfying (5.1) will have Hausdorff dimension equal to δ, and for these particular nice sets, the Cartesian product satisfies the expected dimensional relation, so that the ambient space in the above result is of Hausdorff dimension $\sum_{i=1}^{t} \delta_i$.

The interest in Theorem 5.1 is not in its proof, but in its applications. In the original paper, in addition to the study of $\mathrm{Bad}(i, j)$ and similar sets, the theorem was applied to approximation of complex numbers by ratios of Gaussian integers, to approximation of p-adic numbers, to function fields over a finite field, to problems in complex dynamics and to limit sets of Kleinian groups. More applications have occurred since then.

Initially, we use it to prove Jarník's theorem. Let $I = [0, 1]$ and consider the set

$$\mathrm{Bad}_I = \left\{ x \in [0, 1] : \left| x - \frac{p}{q} \right| > c(x)/q^2 \text{ for all rationals } \frac{p}{q} \right\}.$$

This is the classical set Bad of badly approximable numbers restricted to the unit interval. Clearly, it can be expressed in the form $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho)$ with $\rho(r) = r^{-2}$ and

$$X = \Omega = [0, 1], \quad J = \{(p, q) \in \mathbb{N} \times \mathbb{N} \backslash \{0\} : p \leq q\},$$

$$\alpha = (p, q) \in J, \quad \beta_\alpha = q, \quad R_\alpha = \frac{p}{q}.$$

The metric d is of course the standard Euclidean metric; $d(x, y) := |x - y|$. Thus in this basic example, the resonant sets R_α are simply rational points p/q. With reference to our framework, let the measure m be one–dimensional Lebesgue measure on I. Thus, $\delta = 1$ and all the many conditions are easily checked.

We show that the conditions of Theorem 5.1 are satisfied for this basic example. The existence of the collection $\mathcal{C}(\theta B_n)$, where B_n is an arbitrary closed interval of length $2\,k^{-2n}$ follows immediately from the following simple observation. For any two distinct rationals p/q and p'/q' with $k^n \le q, q' < k^{n+1}$ we have that

$$\left|\frac{p}{q} - \frac{p'}{q'}\right| \ge \frac{1}{qq'} > k^{-2n-2}. \tag{5.6}$$

Thus, any interval θB_n with $\theta := \frac{1}{2}k^{-2}$ contains at most one rational p/q with $k^n \le q < k^{n+1}$. Let $\mathcal{C}(\theta B_n)$ denote the collection of intervals $2\theta B_{n+1}$ obtained by subdividing θB_n into intervals of length $2k^{-2n-4}$ starting from the left hand side of θB_n. Clearly

$$\#\mathcal{C}(\theta B_n) \;\ge\; [k^2/2] \;>\; k^2/4 \;=\; \text{r.h.s. of (5.2) with } \kappa_1 = 1/4.$$

Also, in view of the above observation, for k sufficiently large

$$\text{l.h.s. of (5.3)} \;\le\; 1 \;<\; k^2/8 \;=\; \text{r.h.s. of (5.3) with } \kappa_2 = 1/8.$$

The upshot of this is that Theorem 5.1 implies that $\dim_H \mathrm{Bad}_I \ge 1$. In turn, since Bad_I is a subset of $\mathbb{R}$, this implies that $\dim_H \mathrm{Bad}_I = 1$.

The key feature exploited to check the conditions on the collection is the fact that rational numbers are well spaced. In higher dimensions, the appropriate analogue is the following lemma, the idea of which goes back to Davenport.

Lemma 5.3 *Let $n \ge 1$ be an integer and $k > 1$ be a real number. Let $E \subseteq \mathbb{R}^n$ be a convex set of n–dimensional Lebesgue measure*

$$|E| \le \frac{1}{n!k^{-(n+1)}}.$$

Suppose that E contains $n+1$ rational points $(p_i^{(1)}/q_i, \ldots, p_i^{(n)}/q_i)$ with $1 \le q_i < k$, where $0 \le i \le n$. Then these rational points lie in some hyperplane.

Proof Suppose to the contrary that this is not the case. In that case, the rational points $(p_i^{(1)}/q_i, \ldots, p_i^{(n)}/q_i)$ where $0 \le i \le n$ are distinct. Consider the n–dimensional simplex Δ subtended by them, i.e. an interval when $n = 1$, a triangle when $n = 2$, a tetrahedron when $n = 3$ and so on. Clearly, Δ is a subset of E since E is convex. The volume $|\Delta|$ of the simplex times n factorial is equal to the absolute value of the determinant

$$\det = \begin{vmatrix} 1 & p_0^{(1)}/q_0 & \cdots & p_0^{(n)}/q_0 \\ 1 & p_1^{(1)}/q_1 & \cdots & p_1^{(n)}/q_1 \\ \vdots & \vdots & & \vdots \\ 1 & p_n^{(1)}/q_n & \cdots & p_n^{(n)}/q_n \end{vmatrix}.$$

As this determinant is not zero, it follows from the assumption made on the q_i that

$$n! \times |\Delta| = |\det| \geq \frac{1}{q_0 q_1 \cdots q_n} > k^{-(n+1)}.$$

Consequently, $|\Delta| > (n!)^{-1} k^{-(n+1)} \geq |E|$. This contradicts the fact that $\Delta \subseteq E$.

Of course, in one dimension this is exactly the spacing estimate used in the proof of Jarník's result above.

Lemma 5.3 serves to ensure that not too many rectangles are bad for the application of Theorem 5.1, but we need some way of ensuring that there are enough rectangles to begin with. Lemma 5.5 below accomplishes this and is proved using the following simple covering lemma.

Lemma 5.4 *Let (X, d) be the Cartesian product space of the metric spaces (X_1, d_1), $\ldots$, (X_t, d_t) and $\mathcal{F}$ be a finite collection of 'rectangles' $F = F(c; l_1, \ldots, l_t)$ with $c \in X$ and $l_1, \ldots, l_t$ fixed. Then there exists a disjoint sub-collection $\{F_m\}$ such that*

$$\bigcup_{F \in \mathcal{F}} F \subset \bigcup_m 3F_m.$$

Proof Let S denote the set of centres c of the rectangles in $\mathcal{F}$. Choose $c(1) \in S$ and for $k \geq 1$,

$$c(k+1) \in S \setminus \bigcup_{m=1}^{k} 2\,F(c(m); l_1, \ldots, l_t)$$

as long as $S \setminus \bigcup_{m=1}^{k} 2\,F(c(m); l_1, \ldots, l_t) \neq \emptyset$. Since $\#S$ is finite, the process terminates and there exists $k_1 \leq \#S$ such that

$$S \subset \bigcup_{m=1}^{k_1} 2\,F(c(m); l_1, \ldots, l_t).$$

By construction, any rectangle $F(c; l_1, \ldots, l_t)$ in the original collection $\mathcal{F}$ is contained in some rectangle $3\,F(c(m); l_1, \ldots, l_t)$ and since $d_i(c_i(m), c_i(n)) > 2l_i$ for each $1 \leq i \leq t$ the chosen rectangles $F(c(m); l_1, \ldots, l_t)$ are clearly disjoint.

Lemma 5.5 *Let (X, d) be the Cartesian product of the metric spaces (X_1, d_1), …, (X_t, d_t) and let (Ω, d, m) be a compact measure subspace of X. Let the measure m and the functions ρ_i satisfy conditions (B) to (D). Let k be sufficiently large. Then* *for any $\theta \in \mathbb{R}_+$ and for any rectangle F_n $(n \geq 1)$ there exists a disjoint collection $\mathcal{C}(\theta F_n)$ of rectangles $2\theta F_{n+1}$ contained within θF_n satisfying (5.2) of Theorem 5.1.*

Proof Begin by choosing k large enough so that for any $i \in \{1, \ldots, t\}$,

$$\frac{\rho_i(k^n)}{\rho_i(k^{n+1})} \geq 4. \tag{5.7}$$

That this is possible follows from the fact that $\lambda_i^l(k) \to \infty$ as $k \to \infty$ (condition (B)). Take an arbitrary rectangle F_n and let $l_i(n) := \theta \rho_i(k^n)$. Thus $\theta F_n := F(c; l_1(n), \ldots, l_t(n))$. Consider the rectangle $T_n \subset \theta F_n$ where

$$T_n := F(c; l_1(n) - 2l_1(n+1), \ldots, l_t(n) - 2l_t(n+1)) .$$

Note that in view of (5.7) we have that $T_n \supset \frac{1}{2}\theta F_n$. Now, cover T_n by rectangles $2\theta F_{n+1}$ with centres in $\Omega \cap T_n$. By construction, these rectangles are contained in θF_n and in view of the Lemma 5.4 there exists a disjoint sub-collection $\mathcal{C}(\theta F_n)$ such that

$$T_n \subset \bigcup_{2\theta F_{n+1} \subset \mathcal{C}(\theta F_n)} 6\theta F_{n+1} .$$

Using the fact that rectangles of the same size centred at points of Ω have comparable m measure (condition (C)), it follows that

$$a\, m\left(\frac{1}{2}\theta F_n\right) \leq m(T_n) \leq \#\mathcal{C}(\theta F_n)\, b\, m(6\theta F_{n+1}) .$$

Using the fact that the measure m is doubling on rectangles (condition (D)), so that $m(\frac{1}{2}\theta F_n) \geq D^{-1} m(\theta F_n)$ and $m(6\theta F_{n+1}) \leq m(8\theta F_{n+1}) \leq D^3 m(\theta F_{n+1})$, it follows that

$$\#\mathcal{C}(\theta F_n) \geq \frac{a}{bD^4} \frac{m(\theta F_n)}{m(\theta F_{n+1})}.$$

With Theorem 5.1 and the lemmas, we can intersect the sets $\mathrm{Bad}(i_1, \ldots, i_n)$ with nice fractals. We begin with the case when $i_1 = \cdots = i_n = 1/n$.

Let Ω be a compact subset of $\mathbb{R}^n$ which supports a non-atomic, finite measure m. Let $\mathcal{L}$ denote a generic hyperplane of $\mathbb{R}^n$ and let $\mathcal{L}^{(\epsilon)}$ denote its ϵ-neighbourhood. We say that m is *absolutely α–decaying* if there exist strictly positive constants C, α, r_0 such that for any hyperplane $\mathcal{L}$, any $\epsilon > 0$, any $x \in \Omega$ and any $r < r_0$,

$$m\left(B(x, r) \cap \mathcal{L}^{(\epsilon)}\right) \leq C \left(\frac{\epsilon}{r}\right)^{\alpha} m(B(x, r)) .$$

This is a quantitative way of saying that the support of the measure does not concentrate on any hyperplane, and so a way of quantifying the statement that the set Ω

is sufficiently spread out in $\mathbb{R}^n$. If $\Omega \subset \mathbb{R}$ is supporting a measure m which satisfies (5.1), then it is relatively straightforward to show that m is absolutely δ–decaying. However, in higher dimensions one need not imply the other.

Inside the set Ω, we will define sets of weighted badly approximable numbers as follows. For $0 \leq i_1, \ldots, i_n \leq 1$ with $i_1 + \cdots + i_n = 1$, let

$$\begin{aligned}&\mathrm{Bad}_\Omega(i_1, \ldots, i_n)\\ &= \left\{x \in \Omega : \max_{1\leq j\leq n}\{\|qx_i\|^{1/i_j} > c(x)q^{-1} \text{ for some } c(x) > 0, \text{ for all } q \in \ltimes\right\}.\end{aligned}$$

When $i_1 = \cdots = i_n = 1/n$, we will for brevity denote this set by $\mathrm{Bad}_\Omega(n)$.

Theorem 5.6 *Let Ω be a compact subset of $\mathbb{R}^n$ which supports a measure m satisfying condition* (5.1) *and which in addition is absolutely α–decaying for some $\alpha > 0$. Then*

$$\dim_H \mathrm{Bad}_\Omega(n) = \dim_H \Omega\ .$$

Proof The set $\mathrm{Bad}_\Omega(n)$ can be expressed in the form $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho)$ with $\rho(r) = r^{-(1+\frac{1}{n})}$ and

$$X = (\mathbb{R}^n, d)\ ,\quad J = \{((p_1, \ldots, p_n), q) \in \mathbb{N}^n \times \mathbb{N}\backslash\{0\}\}\ ,$$

$$\alpha = ((p_1, \ldots, p_n), q) \in J\ ,\quad \beta_\alpha = q\ ,\quad R_\alpha = (p_1/q, \ldots, p_n/q)\ .$$

Here d is standard sup metric on $\mathbb{R}^n$; $d(x, y) = \max\{d(x_1, y_1), \ldots, d(x_n, y_n)\}$. Thus balls $B(c, r)$ in $\mathbb{R}^n$ are genuinely cubes of sidelength $2r$.

We show that the conditions of Theorem 5.1 are satisfied. Clearly the function ρ satisfies condition (B) and we are given that the measure m supported on Ω satisfies condition (A) Conditions (C), (D) and (E) also follow from (5.1). Since the resonant sets $\mathcal{R}_\alpha$ are all points, the condition $\dim_H(\cup_{\alpha\in J} R_\alpha) < \delta$ is satisfied by properties (iv) and (vii) of Hausdorff dimension. We need to establish the existence of the disjoint collection $\mathcal{C}(\theta B_n)$ of balls (cubes) $2\theta B_{n+1}$ where B_n is an arbitrary ball of radius $k^{-n(1+\frac{1}{n})}$ with centre in Ω. In view of Lemma 5.5, there exists a disjoint collection $\mathcal{C}(\theta B_n)$ such that

$$\#\mathcal{C}(\theta B_n)\ \geq\ \kappa_1\ k^{(1+\frac{1}{n})\delta}; \tag{5.8}$$

i.e. (5.2) of Theorem 5.1 holds. We now verify that (5.3) is satisfied for any such collection.

We consider two cases.

Case 1: $n = 1$. The trivial spacing argument of (5.6) shows that any interval θB_n with $\theta := \frac{1}{2}k^{-2}$ contains at most one rational p/q with $k^n \leq q < k^{n+1}$; i.e. $\alpha \in J(n+1)$. Thus, for k sufficiently large

$$\text{l.h.s. of (5.3)} \ \leq \ 1 \ < \ \frac{1}{2} \times \ \text{r.h.s. of (5.8)} \ .$$

Hence (5.3) is trivially satisfied and Theorem 5.1 implies the desired result. As a special case, we have shown that the badly approximable numbers in the ternary Cantor set form a set of maximal dimension.

Case 2: $n \geq 2$. We will prove the theorem in the case that $n = 2$. There are no difficulties and no new ideas are required in extending the proof to higher dimensions. One just needs to apply Lemma 5.3 in higher dimensions.

Suppose that there are three or more rational points $(p_1/q, p_2/q)$ with $k^n \leq q < k^{n+1}$ lying within the ball/square θB_n. Now put $\theta = 2^{-1}(2k^3)^{-1/2}$. Then Lemma 5.3 implies that the rational points must lie on a line $\mathcal{L}$ passing through θB_n. Setting $\epsilon = 8\theta k^{-(n+1)\frac{3}{2}}$, it follows that

$$\begin{aligned}\text{l.h.s. of (5.3)} &\leq \# \{2\theta B_{n+1} \subset \mathcal{C}(\theta B_n) : 2\theta B_{n+1} \cap \mathcal{L} \neq \emptyset\} \\ &\leq \# \left\{2\theta B_{n+1} \subset \mathcal{C}(\theta B_n) : 2\theta B_{n+1} \subset \mathcal{L}^{(\epsilon)}\right\} .\end{aligned}$$

Using that the balls $2\theta_{n+1}$ are disjoint and that the measure m is absolutely α-decaying, this is

$$\leq \frac{m(\theta B_n \cap \mathcal{L}^{(\epsilon)})}{m(2\theta B_{n+1})} \leq a^{-1} b C 8^{\alpha} 2^{-\delta} k^{\frac{2}{3}(\delta - \alpha)}.$$

On choosing k large enough, this becomes $\leq \frac{1}{2} \times$ r.h.s. of (5.8). Hence (5.3) is satisfied and Theorem 5.1 implies the desired result.

We now prove the result for general values of i_j, but under a more restrictive assumption on the underlying fractal.

Theorem 5.7 *For $1 \leq j \leq n$, let Ω_j be a compact subset of $\mathbb{R}$ which supports a measure m_j satisfying (5.1) with exponent δ_j. Let Ω denote the product set $\Omega_1 \times \cdots \times \Omega_n$. Then, for any n–tuple $(i_1, \ldots, i_n)$ with $i_j \geq 0$ and $\sum_{j=1}^n i_j = 1$,*

$$\dim_H \mathrm{Bad}_{\Omega}(i_1, \ldots, i_n) = \dim_H \Omega \ .$$

A simple application of the above theorem leads to following result.

Corollary 5.8 *Let K_1 and K_2 be regular Cantor subsets of $\mathbb{R}$. Then*

$$\dim_H \left((K_1 \times K_2) \cap \mathrm{Bad}(i, j)\right) \ = \ \dim_H (K_1 \times K_2) \ = \ \dim_H K_1 \ + \ \dim_H K_2 \ .$$

Proof of Theorem 5.7. We shall restrict our attention to the case $n = 2$ and leave it for the reader to extend this to higher dimensions.

A relatively straightforward argument shows that $m := m_1 \times m_2$ is absolutely α–decaying on Ω with $\alpha := \min\{\delta_1, \delta_2\}$. In fact, more generally for $2 \leq j \leq n$, if each m_j is absolutely α_j–decaying on Ω_j, then $m := m_1 \times \ldots \times m_n$ is absolutely α–decaying on $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with $\alpha = \min\{\alpha_1, \cdots, \alpha_n\}$.

Now let us write $\mathrm{Bad}(i, j)$ for $\mathrm{Bad}(i_1, i_2)$ and without loss of generality assume that $i < j$. The case $i = j$ is already covered by Theorem 4 since m is absolutely α–decaying on Ω and clearly satisfies (5.1). The set $\mathrm{Bad}_\Omega(i, j)$ can be expressed in the form $\mathrm{Bad}^*(\mathcal{R}, \beta, \rho_1, \rho_2)$ with $\rho_1(r) = r^{-(1+i)}$, $\rho_2(r) = r^{-(1+j)}$ and

$$X = \mathbb{R}^2\ , \quad \Omega = \Omega_1 \times \Omega_2\ , \quad J = \{((p_1, p_2), q) \in \mathbb{N}^2 \times \mathbb{N}\backslash\{0\}\}\ ,$$

$$\alpha = ((p_1, p_2), q) \in J\ , \quad \beta_\alpha = q\ , \quad R_\alpha = (p_1/q, p_2/q)\ .$$

The functions ρ_1, ρ_2 satisfy condition (B) and the measures m_1, m_2 satisfy (5.1). Also note that $\dim_H(\cup_{\alpha \in J} R_\alpha) = 0$ since the union in question is countable. We need to establish the existence of the collection $\mathcal{C}(\theta F_n)$, where each F_n is an arbitrary closed rectangle of size $2k^{-n(1+i)} \times 2k^{-n(1+j)}$ with centre c in Ω. By Lemma 5.5, there exists a disjoint collection $\mathcal{C}(\theta F_n)$ of rectangles $2\theta F_{n+1} \subset \theta F_n$ such that

$$\#\mathcal{C}(\theta F_n) \ \geq\ \kappa_1\, k^{(1+i)\delta_1} k^{(1+j)\delta_2}; \tag{5.9}$$

i.e. (5.4) of Theorem 5.2 is satisfied. We now verify that (5.5) is satisfied for any such collection. With $\theta = 2^{-1}(2k^3)^{-1/2}$, the Lemma 5.3 implies that

$$\text{l.h.s. of } \ (5.5)\ \leq\ \#\{2\theta F_{n+1} \subset \mathcal{C}(\theta F_n) : 2\theta F_{n+1} \cap \mathcal{L} \neq \emptyset\}\ , \tag{5.10}$$

where $\mathcal{L}$ is a line passing through θF_n. Consider the thickening $T(\mathcal{L})$ of $\mathcal{L}$ obtained by placing rectangles $4\theta F_{n+1}$ centred at points of $\mathcal{L}$; that is, by 'sliding' a rectangle $4\theta F_{n+1}$, centred at a point of $\mathcal{L}$, along $\mathcal{L}$. Then, since the rectangles $2\theta F_{n+1} \subset \mathcal{C}(\theta F_n)$ are disjoint,

$$\begin{aligned} &\#\{2\theta F_{n+1} \subset \mathcal{C}(\theta F_n) : 2\theta F_{n+1} \cap \mathcal{L} \neq \emptyset\} \\ &\qquad\leq \#\{2\theta F_{n+1} \subset \mathcal{C}(\theta F_n) : 2\theta F_{n+1} \subset T(\mathcal{L})\} \\ &\qquad\leq \frac{m(T(L) \cap \theta F_n)}{m(2\theta F_{n+1})}. \end{aligned} \tag{5.11}$$

Without loss of generality we can assume that $\mathcal{L}$ passes through the centre of θF_n. To see this, suppose that $m(T(\mathcal{L}) \cap \theta F_n) \neq 0$ since otherwise there is nothing to prove. Then, there exists a point $x \in T(\mathcal{L}) \cap \theta F_n \cap \Omega$ such that

$$T(\mathcal{L}) \cap \theta F_n \subset 2\theta F'_n \cap T'(\mathcal{L}')\ .$$

Here F'_n is the rectangle of size $k^{-n(1+i)} \times k^{-n(1+j)}$ centred at x, $\mathcal{L}'$ is the line parallel to $\mathcal{L}$ passing through x and $T'(L')$ is the thickening obtained by 'sliding' a rectangle $8\theta F_{n+1}$ centred at x, along $\mathcal{L}'$. Then the following argument works just as well on $2\theta F'_n \cap T'(\mathcal{L}')$.

Let Δ denote the slope of the line $\mathcal{L}$ and assume that $\Delta \geq 0$. The case $\Delta < 0$ can be dealt with similarly. By moving the rectangle θF_n to the origin, straightforward geometric considerations lead to the following facts:

(F1)
$$T(\mathcal{L}) = \mathcal{L}^{(\epsilon)} \text{ where } \epsilon = \frac{4\theta\left(k^{-(n+1)(1+j)} + \Delta k^{-(n+1)(1+i)}\right)}{\sqrt{1+\Delta^2}},$$

(F2) $T(\mathcal{L}) \cap \theta F_n \subset F(c; l_1, l_2)$ where $F(c; l_1, l_2)$ is the rectangle with the same centre c as F_n and of size $2l_1 \times 2l_2$ with
$$l_1 = \frac{\theta}{\Delta}\left(k^{-n(1+j)} + 4k^{-(n+1)(1+j)} + \Delta k^{-(n+1)(1+i)}\right) \text{ and } l_2 = \theta k^{-n(1+j)}.$$

The asymmetrical shape of the sliding rectangle adds tremendously to the technical calculations from now on. However, we can in fact estimate the right hand side of (5.11) by considering two cases, depending on the magnitude of Δ. Throughout, let a_i, b_i denote the constants associated with the measure m_i and condition (5.1) and let
$$\varpi = 3\left(\frac{4b_1b_2}{\kappa_1\, a_1a_2 2^{\delta_1+\delta_2}}\right)^{1/\delta_1}.$$

Case 1: $\Delta \geq \varpi k^{-n(1+j)}/k^{-n(1+i)}$. In view of (F2) above, we trivially have that
$$m(\theta F_n \cap T(\mathcal{L})) \leq m(F(c; l_1, l_2)) \leq b_1\, b_2\, l_1^{\delta_1}\, l_2^{\delta_2}.$$

It follows that

$$\begin{aligned}
\frac{m(T(L)\cap\theta F_n)}{m(2\theta F_{n+1})} &\leq \frac{b_1b_2 l_1^{\delta_1} l_2^{\delta_2}}{a_1a_2(2\theta)^{\delta_1+\delta_2}\, k^{-(n+1)(1+j)\delta_1}\, k^{-(n+1)(1+i)\delta_2}} \\
&\leq \frac{b_1b_2}{a_1a_2 2^{\delta_1+\delta_2}}\left(\frac{1}{\varpi} + \frac{1}{\varpi k^{1+j}} + \frac{1}{k^{1+i}}\right)^{\delta_1} k^{(1+j)\delta_1+(1+i)\delta_2} \\
&\leq \frac{b_1b_2}{a_1a_2 2^{\delta_1+\delta_2}}\left(\frac{3}{\varpi}\right)^{\delta_1} k^{(1+j)\delta_1+(1+i)\delta_2} = \frac{\kappa_1}{4} k^{(1+j)\delta_1+(1+i)\delta_2}.
\end{aligned}$$

Case 2: $0 \leq \Delta < \varpi k^{-n(1+j)}/k^{-n(1+i)}$. By Lemma 5.4, there exists a collection $\mathcal{B}_n$ of disjoint balls B_n with centres in $\theta F_n \cap \Omega$ and radii $\theta k^{-n(1+j)}$ such that
$$\theta F_n \cap \Omega \subset \bigcup_{B_n \in \mathcal{B}_n} 3B_n.$$

Since $i < j$, it is easily verified that the disjoint collection $\mathcal{B}_n$ is contained in $2\theta F_n$ and thus $\#\mathcal{B}_n \leq m(2\theta F_n)/m(B_n)$. It follows that

$$m(\theta F_n \cap T(\mathcal{L})) \leq m\left(\cup_{B_n \in \mathcal{B}_n} 3B_n \cap\ T(\mathcal{L})\right) \leq \#\mathcal{B}_n\ m(3B_n \cap T(\mathcal{L})).$$

Applying (F1) and subsequently the fact that m is absolutely α-decaying, this is

$$\leq \frac{m(2\theta F_n)}{m(B_n)} m\left(3B_n \cap \mathcal{L}^{(\epsilon)}\right) \leq m(2\theta F_n)\,\frac{m(3B_n)}{m(B_n)}\left(\frac{\epsilon}{3\theta k^{-n(i+j)}}\right)^{\alpha}$$

Now notice that

$$\frac{\epsilon}{3\theta k^{-n(i+j)}}\ \leq\ \frac{4}{3}\,(k^{-(1+j)} + \varpi k^{-(1+i)})\ .$$

Hence, for k sufficiently large we have

$$\frac{m(T(\mathcal{L}) \cap \theta F_n)}{m(2\theta F_{n+1})}\ \leq\ \frac{\kappa_1}{4}\ k^{(1+j)\delta_1} k^{(1+i)\delta_2}\ .$$

On combining the above two cases, we have

$$\text{l.h.s. of (5.5)} \leq \frac{m(T(\mathcal{L}) \cap \theta F_n)}{m(2\theta F_{n+1})} \leq \frac{\kappa_1}{4}\ k^{(1+j)\delta_1} k^{(1+i)\delta_2} = \frac{1}{4} \times \text{ l.h.s. of (5.9)}.$$

Hence (5.5) is satisfied and Theorem 5.2 implies the desired result.

We give a few remarks on the results above. Firstly, an alternative approach using homogeneous dynamics is also known due to Kleinbock and Weiss [33]. This approach is less versatile, as it only works in real, Euclidean space, but the conditions on the measure are slightly less restrictive, so in this respect their result is stronger.

Secondly, it would be really nice if the approach could give a result on numbers badly approximable by algebraic numbers. Unfortunately, the known spacing estimates for algebraic numbers are not good enough to get the naive approach to work (more on this in the next section). One could hope that proving a result on badly approximable vectors on the Veronese curves would work, but unfortunately Lemma 5.3 does not allow us to control the direction of the hyperplane containing the rational points. If this hyperplane is close to tangential to the curve, the approach used above will not work, so more input is needed. In fact, the problem has now been resolved for the full Veronese curve in 2 dimensions by Badziahin and Velani [2] and in higher dimensions by Beresnevich [5]. To the knowledge of the author, the problem of intersecting the sets with fractals remain unsolved.

Thirdly, our failure in proving Schmidt's conjecture with this approach was due to the fact that we could not construct a measure on $\mathrm{Bad}(i, j)$ satisfying the conditions of Theorem 5.1. Since the publication of the paper, the conjecture has been settled, and in fact An [1] proved that the sets $\mathrm{Bad}(i, j)$ are winning for the so-called Schmidt game [45]. This implies that they are stable under countable intersection. We proceed with a discussion of Schmidt games in one dimension and leave it as an exercise to extend this to higher dimensions.

Definition 5.9 Let $F \subseteq \mathbb{R}$, and let $\alpha, \beta \in (0, 1)$. The *Schmidt game* is played by two players, Black and White, according to the following rules:

1. Black picks a closed interval B_1 of length r.
2. White picks a closed interval $W_1 \subseteq B_1$ of length αr.
3. Black picks a closed interval $B_2 \subseteq W_1$ of length $\beta\alpha r$.
4. And so on…

By Banach's fixpoint theorem, $\cap_i B_i$ consists of a single point, x say. If $x \in F$, White wins the game. Otherwise, Black wins.

By requiring that the initial ball is chosen in a particular way, we can use the above for bounded sets without breaking the game.

We are concerned with winning strategies for the game. In particular, we will prove that if White has a winning strategy, then the set F is large.

Definition 5.10 A set F is said to be (α, β)-winning if White can always win the Schmidt game with these parameters. A set F is said to be α-winning if it is (α, β)-winning for any $\beta \in (0, 1)$.

We will prove the following theorems.

Theorem 5.11 *An α-winning set F has Hausdorff dimension* 1.

Theorem 5.12 *If (F_i) is a sequence of α-winning sets, then $\cap_i F_i$ is also α-winning.*

Proof of Theorem 5.11. We suppose without loss of generality that $r = 1$. For ease of computation, we will also assume that $\beta = 1/N$ for an integer $N > 1$. Given an interval W_k in the game, we may partition this set into N (essentially) disjoint intervals. We will restrict the possible choices that Black can make by requiring that she picks one of these. By requiring that β was slightly smaller than $1/N$, we could ensure that the intervals were properly disjoint. We will continue our calculations with this assumption, even though we should strictly speaking add a little more technicality to the setup.

Given that White plays according to a winning strategy, we find disjoint paths through the game depending on the choices made by Black. In other words, for each possible resulting element in F, we find a unique sequence with elements in $\{0, \dots, N-1\}$ and *vice versa*, for each such sequence, we obtain a resulting element. Hence, this particular subset of F may be mapped onto the unit interval by thinking of the sequence from the game as a sequence of digits in the base N-expansion of a number between 0 and 1. To sum up, we have constructed a surjective function

$$g : F^* \to [0, 1],$$

where $F^* \subseteq F$. We extend this to a function on arbitrary subsets of $\mathbb{R}$ by setting $g(A) = g(A \cap F^*)$.

Now, let $\{U_i\}$ be a cover of F^* with U_i having diameter ρ_i. Then, with $\mathcal{L}$ denoting the outer Lebesgue measure,

$$\sum_{i=1}^{\infty} \mathcal{L}(g(U_i)) \geq \mathcal{L}\left(\bigcup_{i=1}^{\infty} g(U_i)\right) \geq \mathcal{L}([0, 1]) = 1.$$

Let $\omega > 0$ be so small that any interval of length $\omega(\alpha\beta)^k$ intersects at most two of the generation k intervals chosen by Black, i.e. any of the intervals $B_k(j_1, \ldots, j_k)$ where $j_i \in \{0, \ldots, N-1\}$. Even in higher dimensions, $\omega = 2/\sqrt{3} - 1$ will do nicely. Finally, define integers

$$k_i = \left[\frac{\log(2\omega^{-1}\rho_i)}{\log \alpha\beta}\right].$$

If ρ_i is sufficiently small, then $k_i > 0$ and $\rho_i < \omega(\alpha\beta)^{k_i}$. Hence, the interval U_i intersects at most two of the generation k_i-intervals chosen by Black. The image of such an interval under g is evidently an interval of length N^{-k_i}, so since there are no more than two of them, $\mathcal{L}(g(U_i)) \leq 2N^{-k_i}$. Summing up over i, we find that

$$1 \leq \sum_{i=1}^{\infty} \mathcal{L}(g(U_i)) \leq \sum_{i=1}^{\infty} 2N^{-k_i} \leq K \sum_{i=1}^{\infty} \rho_i^{\frac{\log N}{|\log(\alpha\beta)|}},$$

where $K > 0$ is explicitly computable in terms of N, α, β and ω. Nonetheless, we have obtained a positive lower bound on the $\frac{\log N}{|\log(\alpha\beta)|}$-length of an arbitrary cover of F^* with small enough sets. It follows that

$$\dim_{\mathrm{H}}(F) \geq \frac{\log N}{|\log(\alpha\beta)|} = \frac{|\log \beta|}{|\log \alpha| + |\log \beta|}.$$

The result now follows on letting $\beta \to 0$.

Proof of Theorem 5.12. White plays according to different strategies at different stages of the game. Explicitly, for α and β fixed, in the first, third, fifth etc. move, White plays according to a $(\alpha, \alpha\beta\alpha; E_1)$-winning strategy, i.e. a strategy for the $(\alpha, \alpha\beta\alpha; E_1)$ for which White is guaranteed to win the game. Since $\rho(B_{l+1}) = \alpha\beta\alpha\rho(B_{l-1})$, this is a valid strategy, and hence the resulting $x \in E_1$. Along the second, sixth, tenth etc. move, White plays according to a $(\alpha, \alpha(\beta\alpha)^3; E_2)$-winning strategy. This is equally valid, and ensures that $x \in E_2$.

In general, in the k'th move with $k \equiv 2^{l-1} \pmod 2)^l$, White moves as if he was playing the $(\alpha, \alpha(\beta\alpha)^{2^l-1}; E_l)$-game. This ensures that the resulting element x is an element of E_l for any l.

A positional strategy is a strategy which may be chosen by looking only at the present state of the game without taking previous moves into account. In An's proof, the strategy chosen by White is not positional (at least it appears not to be). This is a little annoying, as it was shown by Schmidt that any winning set admits a positional strategy. Of course, this proof depends on the well-ordering principle, and so ultimately on the axiom of choice. Describing a positional winning strategy may hence not be that easy.

6 Well Approximable Elements

In this section, we return to where we started, namely to Khintchine's theorem and to fractals arising from continued fractions. We will address three problems. The first is the problem of the size of the exceptional sets in Khintchine's theorem. This will be resolved using a technique due to Beresnevich and Velani known as the mass transference principle. The second is concerned with the ternary Cantor set and the Diophantine properties of elements in it. We will discuss the possibility of getting a Khintchine type theorem for this set and also give a quick-and-dirty argument, stating that most numbers in the set are not ridiculously well approximable by algebraic numbers. Finally, we will remark on some fractal properties which can be used in the study of Littlewood's conjecture.

Initially, we begin with a discussion of the exceptional sets arising from Khintchine's theorem. The Hausdorff dimension of the null sets in the case of convergence was originally calculated by Jarník [30] and independently by Besicovitch [10]. Various new methods were introduced during the last century, with the notion of ubiquitous systems being a key concept in recent years. Ubiquity was introduced (or at least named) by Dodson, Rynne and Vickers [20] and put in a very general form by Beresnevich, Dickinson and Velani [8]. A complete discussion of ubiquity will not be given here, but the reader is strongly encouraged to look up the paper Beresnevich, Dickinson and Velani.

Even more recently (this century), it was observed by Beresnevich and Velani [7] that under relatively mild assumptions on a *limsup* set, one may transfer a zero–one law for such a set to a zero–infinity law for Hausdorff measures, at least in the case of full measure. In the cases considered in these notes, the converse case of measure zero is easy. Note that the measure zero case is not always the easiest! One can cook up problems where the convergence case of a Khintchine type theorem is the difficult part. The sets considered in these notes however all fall within the category where divergence is the difficult problem.

At the heart of the observation of Beresnevich and Velani is the following theorem, usually called the mass transference principle. To state it, we will need a little notation. For a ball $B = B(x, r) \subseteq \mathbb{R}^n$ and a dimension function f, we define

$$B^f = B(x, f(r)^{1/n}),$$

the ball with the same centre but with its radius adjusted according to the dimension function and the dimension of the ambient space. As usual, for $f(r) = r^k$, we denote B^f by B^k.

Theorem 6.1 *Let $\{B_i\}$ be a sequence of balls in $\mathbb{R}^n$ with $r(B_i) \to 0$ as $i \to \infty$. Let f be a dimension function such that $r^{-n}f(r)$ is monotonic. Suppose that*

$$\mathcal{H}^n(B \cap \limsup B_i^f) = \mathcal{H}^n(B),$$

for any ball $B \subseteq \mathbb{R}^n$. Then, for any ball $B \subseteq \mathbb{R}^n$,

$$\mathcal{H}^f(B \cap \limsup B_i^n) = \mathcal{H}^f(B),$$

Notice that the first requirement just says that the set $\limsup B_i^f$ is full with respect to Lebesgue measure, as $\mathcal{H}^n$ is comparable with the Lebesgue measure. Note also, that if $r^{-n}f(r) \to \infty$ as $r \to 0$, $\mathcal{H}^f(B) = \infty$, so a re-statement of the conclusion of the theorem would be as follows: Suppose that a *limsup* set of balls is full with respect to Lebesgue measure. Then a *limsup* set of appropriately scaled balls is of infinite Hausdorff measure.

The mass transference principle is valid in a more general setting of certain metric spaces. As was the case in the framework of badly approximable sets, the metric space must support a natural measure, which in this case should be a Hausdorff measure. This is the case for the Cantor set with the Hausdorff log 2/log 3-measure, and the mass transference principle is exactly the same if one reads log 2/log 3 for n everywhere.

Finally, the reader will note that the mass transference principle in the present form only works for *limsup* sets of balls. If one were to consider linear forms approximation as we did in Sect. 4, the *limsup* set would be built from tubular neighbourhoods of hyperplanes, and for the more general setting of systems of linear forms from tubular neighbourhoods of lower dimensional affine subspaces. This can be overcome by a slicing technique, also developed by Beresnevich and Velani [6].

We will not go into details on the higher dimensional variant here, nor will we prove the mass transference principle. Instead, we will deduce the original Jarník–Besicovitch theorem from Khintchine's theorem.

Theorem 6.2 *Let f be a dimension function and let $\psi : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be some function with $q^2 f(\psi(q))$ decreasing. Then,*

$$\mathcal{H}^f\left\{x \in [0,1] : \left|x - \frac{p}{q}\right| < \psi(q) \textit{ for infinitely many } (p,q) \in \mathbb{Z} \times \mathbb{N}\right\},$$

is zero or infinity according to whether the series $\sum qf(\psi(q))$ converges or diverges.

Proof The convergence half is the usual covering argument, which we omit. For the divergence part, we apply the mass transference principle. The balls are indexed by rational numbers, with $B_{p/q} = B(p/q, \psi(q))$ and $k = 1$, so the *limsup* set of the conclusion of Theorem 6.1 is just the set from the statement of the theorem. Hence, it suffices to prove that $\limsup B_{p/q}^f$ is full with respect to Lebesgue measure, provided the series in question diverges. But this is just the statement of the original Khintchine's theorem.

An easy corollary of this statement tells us, that for $\psi(q) = q^{-v}$, the upper bound of $2/v$ obtained on the Hausdorff dimension of the above set in Sect. 3 is sharp. In fact, in Khintchine's theorem, the requirement that $q^2\psi(q)$ is monotonic can be relaxed substantially to the requirement that $\psi(q)$ is monotonic, which in turn gives us the Hausdorff measure at the critical dimension (it is infinite) by the above argument. The latter result could be deduced directly from Dirichlet's theorem, as the set of all numbers evidently is full, but only for the special case of the approximating function $\psi(q) = q^{-v}$.

We leave it for the reader to explore applications of the mass transference principle (there are many). The point we want to make is that once a zero–one law for a natural measure is known (Lebesgue in the case of the real numbers), it is usually a straightforward matter of applying the mass transference principle to get a Hausdorff measure variant of the known result. In other words, it is natural to look for a zero–one law for the natural measure and deduce the remainder of the metrical theory from this result.

We now consider the ternary Cantor set. Recall that the natural measure μ constructed in Sect. 3 on this set has the nice decay property, that for $\delta = \log 2/\log 3$ and $c_1, c_2 > 0$,

$$c_1 r^\delta \leq \mu([c-r, c+r]) \leq c_2 r^\delta \tag{6.1}$$

for all $c \in \mathcal{C}$ and $r > 0$ small enough. We used this property in the preceeding section as well. It is easy to see that any non-atomic measure supported on $\mathcal{C}$ satisfying hypothesis (6.1) must also satisfy

$$\mu([c-\epsilon r, c+\epsilon r]) \leq c_3 \epsilon^\delta \mu([c-r, c+r]), \tag{6.2}$$

for some $c_3 > 0$, whenever r and ϵ are small and $c \in \mathbb{R}$. The inequality in (6.2) is the statement that the measure is absolutely δ-decaying, which was also used in the preceeding section. In fact, this measure can also be seen to be the restriction of the Hausdorff δ-measure to $\mathcal{C}$, so we are within the framework where the mass transference principle can be applied.

Levesley, Salp and Velani [37] proved a zero–one law (and deduced the corresponding statement for Hausdorff measures) for the set

$$W_{\mathcal{C}} = \left\{x \in \mathcal{C} : \left|x - \frac{p}{3^n}\right| < \psi(3^n) \text{ for infinitely many } (p, n) \in \mathbb{Z} \times \mathbb{N}\right\}.$$

Note the restriction on the approximating rationals. They are all rationals whose denominator is a power of 3, and so are the endpoints in the usual construction of the set. This is of course not satisfactory, as there are other rationals in the Cantor set, e.g. $1/4$. Nevertheless, we do not at present know a full zero–one law for approximation of elements in the Cantor set by rationals in the Cantor set. Their result is the following.

Theorem 6.3

$$\mu(W_{\mathcal{C}}) = \begin{cases} 0, & \sum_{n=1}^{\infty} (3^n \psi(3^n))^\delta < \infty, \\ 1, & \sum_{n=1}^{\infty} (3^n \psi(3^n))^\delta = \infty. \end{cases}$$

We make a few comments on the proof, as it is a good model for many proofs of zero–one laws for *limsup* sets. The convergence part is usually proved by a covering argument as we have seen. For the divergence part, we give an outline of the method used by Levesley, Salp and Velani. From Sect. 4, recall the converse to the Borel–

Cantelli lemma given in Lemma 4.11. By this lemma, it would suffice to prove that the sets forming the *limsup* set are pairwise independent to ensure full measure. However in general, we have no reason to suspect that these sets are pairwise independent, which is always a problem. However, if they satisfy the weaker condition of quasi-pairwise independence (see below), the first part of the lemma will give positive measure. This can subsequently be inflated in a number of ways. One possibility is to look for an underlying invariance for an ergodic transformation. Another is to apply the following local density condition.

Lemma 6.4 *Let μ be a finite, doubling Borel measure supported on a compact set $X \subseteq R^k$, and let $E \subseteq X$ be a Borel set. Suppose that there are constants $r_0, c > 0$, such that for any ball $B = B(x, r)$ with $x \in X$ and $r < r_0$,*

$$\mu(E \cap B) \geq c\mu(B).$$

Then E is full in X with respect to μ.

This result is a consequence of the Lebesgue density theorem.

In order to prove a zero–one law, one now attempts to verify the conditions of Lemma 6.4, with E being the *limsup* set, using Lemma 4.11. In other words, for the *limsup* set

$$\Lambda = \limsup E_n,$$

there is a constant $c > 0$, such that for any sufficiently small ball B centred in X,

$$\frac{\mu((E_m \cap B) \cap (E_n \cap B))}{\mu(B)} \leq c \frac{\mu(E_m \cap B)}{\mu(B)} \frac{\mu(E_n \cap B)}{\mu(B)} \tag{6.3}$$

whenever $m \neq n$. The probability measure used in Lemma 4.11 is the normalised restriction of μ to B. Just inserting the above estimate proves that $\mu(\Lambda \cap B) \geq c^{-1}\mu(B)$, whence Λ is full within X. In other words, it suffices to prove local pairwise quasi-independence of events in the above sense.

For the result of Levesley, Salp and Velani, one considers the subset of $W_{\mathcal{C}}(\psi)$ which is the *limsup* set of the sets

$$E_n = \bigcup_{\substack{0 \leq p \leq 3^n \\ 3 \nmid p}} B\left(\frac{p}{3^n}, \psi(3^n)\right) \cap \mathcal{C}.$$

After making some preliminary reductions, it is then possible to prove (6.3) by splitting up into the cases when m and n are pretty close (in which case intersection on the left hand side is empty) and the case when they are pretty far apart, where some clever counting arguments and the specific form of the measure is needed. The point we want to make is not in the details, but rather in the methods applied.

Of course, applying the mass transference principle immediately gives a condition for the Hausdorff measure of the set $W_{\mathcal{C}}$ to be infinite. Combining this with a covering argument for the convergence case, Levesley, Salp and Velani obtained the following theorem in full.

Theorem 6.5 *Let f be a dimension function with $r^{-\delta}f(r)$ monotonic. Then,*

$$\mathcal{H}^f(W_{\mathcal{C}}) = \begin{cases} 0, & \sum_{n=1}^{\infty} 3^{\delta n} f(\psi(3^n)) < \infty. \\ \mathcal{H}^f(\mathcal{C}), & \sum_{n=1}^{\infty} 3^{\delta n} f(\psi(3^n)) = \infty. \end{cases}$$

We now consider the approximation of elements of $\mathcal{C}$ by elements of $\mathbb{A}_n$, where the quality of approximation is measured in terms of the height of the approximating number. The present argument is from [35]. Of course, we cannot hope to get a Khintchine type result by the methods above, as we do not expect there to be any algebraic irrational elements in $\mathcal{C}$. In fact, we can say very little, and we are only able to get a convergence result. We proceed to give a quick argument, which is not best possible, but relatively short.

Let $\psi : \mathbb{R}_{\geq 1} \to \mathbb{R}_+$. We define the set

$$\mathcal{K}_n^*(\psi; \mathcal{C}) = \{x \in \mathcal{C} : |x - \alpha| < \psi(H(\alpha)) \text{ for infinitely many } \alpha \in \mathbb{A}_n\}. \tag{6.4}$$

Theorem 6.6 *Let $\mathcal{C}$ be the ternary Cantor set and let $\delta = \log 2/\log 3$. Suppose that $\psi : \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 0}$ satisfies either*

$$\sum_{r=1}^{\infty} r^{2n\delta - 1}\psi(r)^{\delta} < \infty \textit{ and } \psi \textit{ is non-increasing} \quad \textit{or} \quad \sum_{r=1}^{\infty} r^n \psi(r)^{\delta} < \infty.$$

Then

$$\mu\left(\mathcal{K}_n^*(\psi; \mathcal{C})\right) = 0.$$

The result is almost surely not sharp. We first prove that the convergence of the first series ensures that the measure is zero. This is by far the most difficult part of the proof. We will use a bound on the distance between algebraic numbers, which is in a sense best possible. If α and β are distinct real algebraic numbers of degree at most n, then

$$|\alpha - \beta| \geq c_4 H(\alpha)^{-n} H(\beta)^{-n}, \tag{6.5}$$

where the constant $c_4 > 0$ depends solely on n. The result can be found in Bugeaud's book [13] as a special case of corollary A.2, where the explicit form of the constant c_4 is also given. It generalises (5.6), which is the same estimate for rational numbers between 0 and 1. In the case of rational numbers, the spacing distribution is much more well-behaved than for real algebraic numbers of higher degree, and for this reason, Theorem 6.6 is almost certainly not as sharp as it could be. Nonetheless, as remarked in [13], the estimate in (5.6) is in some sense best possible.

If for some $k \in \mathbb{N}$, $2^k \le H(\alpha), H(\beta) < 2^{k+1}$, (6.5) implies that $|\alpha - \beta| > \frac{1}{2}c_4 2^{-2n(k+1)}$. Consequently, for distinct real algebraic numbers α_i with $2^k \le H(\alpha_i) < 2^{k+1}$, the intervals $[\alpha_i - \frac{1}{4}c_4 2^{-2n(k+1)}, \alpha_i + \frac{1}{4}c_4 2^{-2n(k+1)}]$ are disjoint.

Let $k \in \mathbb{N}$. We will show that as $k \to \infty$,

$$\max_{2^k \le r < 2^{k+1}} \frac{\psi(r)}{4^{-1}c_4 2^{-2n(k+1)}} = o(1). \tag{6.6}$$

In other words, the ratio tends to 0 as k tends to infinity. Indeed, suppose to the contrary that there is a $c_5 > 0$ and a strictly increasing sequence $\{k_i\}_{i=1}^{\infty} \subseteq \mathbb{N}$ such that for any $i \in \mathbb{N}$

$$\max_{2^{k_i} \le r < 2^{k_i+1}} \frac{\psi(r)}{4^{-1}c_4 2^{-2n(k_i+1)}} > c_5.$$

By the convergence assumption of the theorem together with Cauchy's condensation criterion and the monotonicity of ψ,

$$\sum_{k=1}^{\infty} 2^{2n(k+1)\delta}\psi(2^k)^{\delta} = 2^{2n\delta} \sum_{k=1}^{\infty} \left(2^{2kn}\psi(2^k)\right)^{\delta} < \infty.$$

On the other hand, as ψ is non-increasing,

$$\begin{aligned} \sum_{k=1}^{\infty} 2^{2n(k+1)\delta}\psi(2^k)^{\delta} &\ge 4^{-\delta}c_4^{\delta} \sum_{i=1}^{\infty} \left(\max_{2^{k_i} \le r < 2^{k_i+1}} \frac{\psi(r)}{4^{-1}c_4 2^{-2n(k_i+1)}} \right)^{\delta} \\ &\ge 4^{-\delta}c_4^{\delta}c_5^{\delta} \sum_{i=1}^{\infty} 1 = \infty, \end{aligned}$$

which is the desired contradiction.

Consider the sets

$$E_k = \bigcup_{\substack{\alpha \in \mathbb{A}_n \\ 2^k \le H(\alpha) < 2^{k+1}}} [\alpha - \psi(H(\alpha)), \alpha + \psi(H(\alpha))].$$

Clearly, for k large enough

$$\begin{aligned} \mu(E_k) &\le \sum_{\substack{\alpha \in \mathbb{A}_n \\ 2^k \le H(\alpha) < 2^{k+1}}} \mu([\alpha - \psi(H(\alpha)), \alpha + \psi(H(\alpha))]) \\ &\le c_3 c_4^{\delta} 4^{-\delta} 2^{2n(k+1)\delta} \psi(2^k)^{\delta} \\ &\qquad \sum_{\substack{\alpha \in \mathbb{A}_n \\ 2^k \le H(\alpha) < 2^{k+1}}} \mu([\alpha - \tfrac{1}{4}c_4 2^{-2n(k+1)}, \alpha + \tfrac{1}{4}c_4 2^{-2n(k+1)}]), \end{aligned}$$

where we have used (6.2) and (6.6). The intervals in the final sum are disjoint. Hence, the sum of their measure is bounded from above by the measure of K, which is equal to 1. We have shown that for $k \geq k_0$,

$$\mu(E_k) \leq c_3 c_4^\delta 4^{-\delta} 2^{2n(k+1)\delta} \psi(2^k)^\delta.$$

To complete the proof of this case, we note that $\mathcal{K}_n^*(\psi; K)$ is the set of points falling in infinitely many of the E_k. But

$$\sum_{k=k_0}^{\infty} \mu(E_k) \leq c_3 c_4^\delta 4^{-\delta} \sum_{k=k_0}^{\infty} 2^{2n(k+1)\delta} \psi(2^k)^\delta = c_3 c_4^\delta 4^{-\delta} 2^{2n\delta} \sum_{k=k_0}^{\infty} 2^{2nk\delta} \psi(2^k)^\delta.$$

Using Cauchy's condensation criterion and the convergence assumption of the theorem, the latter series converges. Hence, the Borel–Cantelli lemma implies the theorem.

To show that the convergence of the second series is sufficient to ensure zero measure, we note that

$$\#\{\alpha \in \mathbb{A}_n : \alpha \in [0,1], H(\alpha) = H\} \leq n(n+1)(2H+1)^n. \tag{6.7}$$

By (6.1), for any such α, we have $\mu([\alpha - \psi(H); \alpha + \psi(H)]) \leq c_6 \psi(H)^\delta$ for some $c_6 > 0$. Elements of $\mathcal{K}_n^*(\psi; K)$ fall in infinitely many of these intervals, and as

$$\sum_{H=1}^{\infty} \sum_{\substack{\alpha \in \mathbb{A}_n \\ \alpha \in [0,1] \\ H(\alpha) = H}} \mu([\alpha - \psi(H); \alpha + \psi(H)]) \leq n(n+1) c_6 \sum_{H=1}^{\infty} (2H+1)^n \psi(H)^\delta,$$

which converges by assumption, the measure of $\mathcal{K}_n^*(\psi; K)$ is zero by the Borel–Cantelli lemma. □

It is possible to prove a stronger result using homogeneous dynamics. This was done by Kleinbock, Lindenstrauss and Weiss [34], but the present result has the advantage of being relatively simple to prove.

The final thing, which we will touch upon in these notes, is a result on the Littlewood conjecture, which uses the Fourier dimension, which we defined in Sect. 3, but did not use for anything. We will sketch a proof of the following result, which is a partial result of [28].

Theorem 6.7 *Let* $\{\alpha_i\} \subseteq \mathrm{Bad}$ *be a countable set of badly approximable numbers. The set of* $\beta \in \mathrm{Bad}$ *for which all pairs* (α_i, β) *satisfy the Littlewood conjecture is of Hausdorff dimension* 1.

For a single α_i, this result was also proven by Pollington and Velani [42] by a similar, but slightly more complicated method. It could also be deduced from homogeneous dynamics, but the present method is different, and in fact gives a stronger result. However, it falls short of anything near the seminal result of Einsiedler, Katok and Lindenstrauss [23].

Note that unless both α and β are badly approximable, the Littlewood conjecture is trivially satisfied. Indeed, if α is not badly approximable, there is a sequence q_n such that

$$q_n\|q_n\alpha\| = q_n^2\left|\alpha - \frac{p_n}{q_n}\right| \to 0.$$

Brutally estimating $\|q_n\beta\| \leq 1/2$, we find that

$$q_n\|q_n\alpha\|\|q_n\beta\| \to 0,$$

so that the pair (α, β) satisfies the Littlewood conjecture. Hence, this problem naturally lives on a set of measure zero, namely Bad $\times$ Bad.

The key tool in proving Theorem 6.7 is a result on the discrepancy of certain sequences, which holds true for almost all α with respect to a certain measure introduced by Kaufman [32].

Kaufman's measure μ_M is a measure supported on the set of real numbers with partial quotients bounded above by M. To be explicit, for each real number $\alpha \in [0, 1)$, let

$$\alpha = [a_1, a_2, \dots] = \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots}}}$$

be the simple continued fraction expansion of α. For $M \geq 3$, let

$$F_M = \{\alpha \in [0, 1) : a_i(\alpha) \leq M \text{ for all } i \in \mathbb{N}\}. \tag{6.8}$$

Recall that the set of badly approximable numbers consists exactly of the numbers for which the partial quotients form a bounded sequence, so that

$$\text{Bad} = \bigcup_{M=1}^{\infty} F_M.$$

Kaufman proved that the set F_M supports a measure μ_M satisfying a number of nice properties. For our purposes, we need the following two properties.

(i) For any $s < \dim_H(F_M)$, there are positive constants $c, l > 0$ such that for any interval $I \subseteq [0, 1)$ of length $|I| \leq l$,

$$\mu_M(I) \leq c\,|I|^s.$$

(ii) For any M, there are positive constants $c, \eta > 0$ such that the Fourier transform $\hat{\mu}_M$ of the Kaufman measure μ_M satisfies

$$\hat{\mu}_M(u) \leq c\,|u|^{-\eta}.$$

The first property allows us to connect the Kaufman measure with the Hausdorff dimension of the set F_M via the mass distribution principle of Lemma 3.3. The second property provides a positive lower bound on the Fourier dimension of the set F_M, but for our purposes the property is used only in computations.

The second key tool is the notion of discrepancy from the theory of uniform distribution. The discrepancy of a sequence in [0, 1) measures how uniformly distributed a sequence is in the interval. Specifically, the discrepancy of the sequence (x_n) is defined as

$$D_N(x_n) = \sup_{I \subseteq [0,1]} \left| \sum_{n=1}^{N} \chi_I(x_n) - N\,|I| \right|,$$

where I is an interval and χ_I is the corresponding characteristic function. A sequence (x_n) is uniformly distributed if $D_N(x_n) = o(N)$.

Our key result is the following discrepancy estimate, which implies Theorem 6.7

Theorem 6.8 *Let μ_M be a Kaufman measure and assume that for positive integers $u < v$ we have*

$$\sum_{n,m=u}^{v} |a_n - a_m|^{-\eta} \ll \frac{1}{\log v} \sum_{n=u}^{v} \psi_n$$

where (ψ_n) is a sequence of non-negative numbers and $\eta > 0$ is the constant from property (ii) of the Kaufman measure. Then for μ_M-almost every $x \in [0, 1]$ we have

$$D_N(a_n x) \ll (N \log(N)^2 + \Psi_N)^{1/2} \log(N \log(N)^2 + \Psi_N)^{3/2+\varepsilon} + \max_{n \leq N} \psi_n$$

where $\Psi_N = \psi_1 + \cdots + \psi_N$.

We will need a probabilistic lemma which can be found in [26].

Lemma 6.9 *Let (X, μ) be a measure space with $\mu(X) < \infty$. Let $F(n, m, x),\ n, m \geq 0$ be μ-measurable functions and let ϕ_n be a sequence of real numbers such that $|F(n-1, n, x)| \leq \phi_n$ for $n \in \mathbb{N}$. Let $\Phi_N = \phi_1 + \cdots + \phi_N$ and assume that $\Phi_N \to \infty$. Suppose that for $0 \leq u < v$ we have*

$$\int_X |F(u, v, x)|^2 \, d\mu \ll \sum_{n=u}^{v} \phi_n.$$

Then for μ-almost all x, we have

$$F(0, N, x) \ll \Phi_N^{1/2} \log(\Phi_N)^{3/2+\varepsilon} + \max_{n \leq N} \phi_n.$$

We will also need the classical Erdős–Turán inequality which can be found in [40].

Theorem 6.10 *For any positive integer K and any sequence $(x_n) \subseteq [0, 1)$,*

$$D_N(x_n) \leq \frac{N}{K+1} + 3\sum_{k=1}^{K} \frac{1}{k} \left| \sum_{n=1}^{N} e(kx_n) \right|,$$

where as usual $e(x) = \exp(2\pi i x)$.

Proof of Theorem 6.8. Suppose $M \geq 3$ and for integers $0 \leq u < v$ let

$$F(u, v, x) = \sum_{h=1}^{v} \frac{1}{h} \left| \sum_{n=u}^{v} e(ha_n x) \right|.$$

Theorem 6.10 with $K = N$ tells us that

$$D_N(a_n x) \ll F(0, N, x).$$

Integrating with respect to $d\mu_M(x)$ and applying the Cauchy–Schwarz inequality gives

$$\begin{aligned}\int |F(u, v, x)|^2 \, d\mu_M &\leq \sum_{h,k=1}^{v} \frac{1}{hk} \int \left| \sum_{n=u}^{v} e(ha_n x) \right|^2 d\mu_M \\ &= \sum_{h,k=1}^{v} \frac{1}{hk} \left(v - u + 1 + \sum_{\substack{n,m=u \\ n \neq m}}^{v} \hat{\mu}_M(h(a_n - a_m)) \right).\end{aligned}$$

Finally using property (ii) of the Kaufman measure we have

$$\begin{aligned}\int |F(u, v, x)|^2 \, d\mu_M &\ll \sum_{h,k=1}^{v} \frac{1}{hk} \left(v - u + 1 + h^{-\eta} \sum_{\substack{n,m=u \\ n \neq m}}^{v} |a_n - a_m|^{-\eta} \right) \\ &\ll \sum_{n=u}^{v} [\log(n)^2 + \psi_n].\end{aligned}$$

Since $F(n-1, n, x) \ll \log(n)^2 + \psi_n$ for all $n \geq 1$, the theorem then follows from Lemma 6.9.

For sequences which grow sufficiently rapidly, the theorem has a corollary with a much cleaner statement. We will say that an increasing sequence of positive integers (a_n) is *lacunary* if there is a $c > 1$ such that for any n, $a_{n+1}/a_n > c$. Applying this inductively, we see that the sequence must grow at least as fast as some geometric sequence.

Corollary 6.11 *Let $\nu > 0$, let μ be a Kaufman measure and (a_n) a lacunary sequence of integers. For μ-almost every $x \in [0, 1]$ we have $D_N(a_n x) \ll N^{1/2} (\log N)^{5/2+\nu}$.*

Proof We apply again Theorem 6.8. Using lacunarity of the sequence (a_n), we see that

$$\sum_{n,m=1}^{\infty} |a_n - a_m|^{-\eta} < \infty.$$

Consequently, we can absorb all occurrences of Ψ_N as well as the final term $\max_{n \le N} \psi_n$ in the discrepancy estimate of Theorem 6.8 into the implied constant. It follows that

$$D_N(a_n x) \ll (N \log(N)^2)^{1/2} \log(N \log(N)^2)^{3/2+\varepsilon} \ll N^{1/2} (\log N)^{5/2+\nu}$$

for μ-almost every x, where ν can be made as small as desired by picking ε small enough.

Proof of Theorem 6.7. Let G denote the set of numbers $\beta \in \text{Bad}$ for which there is an i, such that

$$\liminf q \,\|q\alpha_i\|\, \|q\beta\| > 0. \tag{6.9}$$

Suppose, contrary to what we are to prove, that $\dim_H G < 1$. Pick an $M \ge 3$ such that $\dim_H F_M > \dim_H G$ (this can be done in light of Jarník's theorem). Let $\mu = \mu_M$ denote the Kaufman measure on F_M.

Consider first one of the α_i, and let (q_k) denote the sequence of denominators of convergents in the simple continued fraction expansion of α_i. In the following, we will use the various parts of Proposition 1.4 many times to deduce results about this sequence and its relation to α_i.

The sequence q_k is lacunary. Hence, by Corollary 6.11, for $\mu-$almost every x,

$$D_N(q_n x) \ll N^{1/2} (\log N)^{5/2+\nu}.$$

Let $\psi(N) = N^{-1/2+\epsilon}$ for some $\epsilon > 0$ and consider the interval

$$I_N = [-\psi(N), \psi(N)].$$

By the definition of discrepancy, for every $\gamma \in [0, 1]$ and μ-almost every β

$$|\#\{k \le N : \{q_k \beta\} \in I_N\} - 2N\psi(N)| \ll N^{1/2} (\log N)^{5/2+\nu}.$$

Hence,

$$\begin{aligned} \#\{k \le N : \{q_k \beta\} \in I_N\} &\ge 2N\psi(N) - KN^{1/2} (\log N)^{5/2+\nu} \\ &= 2N^{1/2+\epsilon} - KN^{1/2} (\log N)^{5/2+\nu}, \end{aligned}$$

where $K > 0$ is the implied constant from Corollary 6.11. Next let N_h denote the increasing sequences defined by

$$N_h = \min\left\{N \in \mathbb{N} : \#\{k \le N : \{q_k\beta\} \in I_N\} = h\right\}.$$

Since each q_{N_h} is a denominator of a convergent to α_i,

$$q_{N_h} \left\|q_{N_h}\alpha_i\right\| \le 1.$$

Hence,

$$q_{N_h} \left\|q_{N_h}\alpha_i\right\| \left\|q_{N_h}\beta\right\| \le \left\|q_{N_h}\beta\right\| \le \left(N_h^{\gamma}\right)^{-1/2+\epsilon}.$$

This establishes our claim and shows that the exceptional set $E_i \subseteq F_M$ for which (6.9) holds has $\mu(E_i) = 0$.

To conclude, let E be the set of $\beta \in F_M$ for which there is an i or a j such that either (6.4) or (6.5) is not satisfied. Then,

$$E = \bigcup_i E_i \cup \bigcup_j E'_j,$$

and therefore $\mu(E) = 0$.

Finally $\mu(G)$ is maximal, so consider the trace measure $\tilde{\mu}$ of μ on G, defined by $\tilde{\mu}(X) = \mu(X \cap G)$. It follows from property (i) of Kaufman's measures that μ is a mass distribution on $[0, 1)$, and since G is full, $\tilde{\mu}$ inherits the decay property of (i) from μ. By the mass distribution principle it then follows that $\dim_H(G) = \dim_H(F_M) > \dim_H(G)$, which contradicts our original assumption. Therefore we conclude that $\dim_H(G) = 1$.

The proof of Theorem 6.7 in fact tells us that something stronger than the Littlewood conjecture holds for the pairs (α_i, β). Indeed, we can work a little more with the inequalities obtained and get a speed of convergence along the sequence (q_{N_h}). Further results using the full force of the uniform distribution of the sequence $(q_n\beta)$ can be found in the original paper [28], where we also prove similar results for the related p-adic and mixed Littlewood conjectures. However, it is beyond the scope of these notes to discuss these topics, and for the clarity of the exposition we have restricted ourselves to results on the original conjecture.

7 Concluding Remarks

These notes are far from being a complete description of the state-of-the-art in metric Diophantine approximation. Recent developments in metric Diophantine approximation on manifolds has barely been touched upon, and the relation with homogeneous dynamics which has led to spectacular advances in the theory has only been superficially described. The selection of results reflect the tastes and expertises of the author, and much is left out.

Nonetheless, it is hoped that the reader has caught a glimpse of the richness and beauty of the metric theory of Diophantine approximation and has acquired a taste for more. Certainly, there are problems and literature enough to last a lifetime of research.

Acknowledgements I acknowledge the support of the Danish Natural Science Research Council. I am grateful to the organisers and participants of the Summer School 'Diophantine Analysis' in Würzburg 2014. Finally, I warmly thank Kalle Leppälä, Steffen Højris Pedersen and Morten Hein Tiljeset for their comments on various drafts of this manuscript.

References

1. J. An, Badziahin-Pollington-Velani's theorem and Schmidt's game. Bull. Lond. Math. Soc. **45**(4), 721–733 (2013)
2. D. Badziahin, S. Velani, Badly approximable points on planar curves and a problem of Davenport. Math. Ann. **359**(3–4), 969–1023 (2014)
3. D. Badziahin, A. Pollington, S. Velani, On a problem in simultaneous Diophantine approximation: Schmidt's conjecture. Ann. of Math. (2) **174**(3), 1837–1883 (2011)
4. V. Beresnevich, On approximation of real numbers by real algebraic numbers. Acta Arith. **90**(2), 97–112 (1999)
5. V. Beresnevich, Badly approximable points on manifolds. Invent. Math. **202**(3), 1199–1240 (2015)
6. V. Beresnevich, S. Velani, Schmidt's theorem, Hausdorff measures, and slicing. Int. Math. Res. Not. **24** (2006). (Art. ID 48794)
7. V. Beresnevich, S. Velani, A mass transference principle and the Duffin-Schaeffer conjecture for Hausdorff measures. Ann. Math. (2) **164**(3), 971–992 (2006)
8. V. Beresnevich, D. Dickinson, S. Velani, Measure theoretic laws for lim sup sets. Mem. Amer. Math. Soc. **179**(846) (2006). (x+91)
9. V.I. Bernik, The exact order of approximating zero by values of integral polynomials. Acta Arith. **53**(1), 17–28 (1989)
10. A.S. Besicovitch, Sets of fractional dimension (IV); on rational approximation to real numbers. J. Lond. Math. Soc. **9**, 126–131 (1934)
11. P. Billingsley, *Ergodic Theory and Information* (Robert E. Krieger Publishing Co., Huntington, 1978)
12. É. Borel, Les probabilités dénombrables et leurs applications arithmétiques. Rend. Circ. Math. Palermo **27**, 247–271 (1909)
13. Y. Bugeaud, *Approximation by Algebraic Numbers* (Cambridge University Press, Cambridge, 2004)
14. Y. Bugeaud, M.M. Dodson, S. Kristensen, Zero-infinity laws in Diophantine approximation. Q. J. Math. **56**(3), 311–320 (2005)
15. C. Carathéodory, *Über das lineare Mass von Punktmengen, eine Verallgemeinerung des Längenbegriffs*, Gött. Nachr. 404–426 (2014)
16. J.W.S. Cassels, *An Introduction to Diophantine Approximation* (Cambridge University Press, New York, 1957)
17. J.W.S. Cassels, *An Introduction to the Geometry of Numbers* (Springer-Verlag, Berlin, 1997)
18. L.G.P. Dirichlet, *Verallgemeinerung eines Satzes aus der Lehre von den Kettenbrüchen nebst einige Anwendungen auf die Theorie der Zahlen*, S.-B. Preuss. Akad Wiss. 93–95 (1842)

19. M.M. Dodson, *Geometric and probabilistic ideas in the metric theory of Diophantine approximations*. Uspekhi Mat. Nauk **48**(5)(293), 77–106 (1993)
20. M.M. Dodson, B.P. Rynne, J.A.G. Vickers, Diophantine approximation and a lower bound for Hausdorff dimension. Mathematika **37**(1), 59–73 (1990)
21. R.J. Duffin, A.C. Schaeffer, Khintchine's problem in metric Diophantine approximation. Duke Math. J. **8**, 243–255 (1941)
22. M. Einsiedler, T. Ward, *Ergodic Theory with a View Towards Number Theory* (Springer, London, 2011)
23. M. Einsiedler, A. Katok, E. Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture*. Ann. Math. (2) **164**(2), 513–560 (2006)
24. K. Falconer, *Fractal Geometry, Mathematical Foundations and Applications* (Wiley, Hoboken, 2003)
25. O. Frostman, Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions. Meddel. Lunds Univ. Math. Sem **3**, 1–118 (1935)
26. G. Harman, *Metric Number Theory* (The Clarendon Press, Oxford University Press, New York, 1998)
27. F. Hausdorff, Dimension und äusseres Mass. Math. Ann. **79**, 157–179 (1919)
28. A. Haynes, J.L. Jensen, S. Kristensen, Metrical musings on Littlewood and friends. Proc. Amer. Math. Soc. **142**(2), 457–466 (2014)
29. V. Jarník, *Zur metrischen Theorie der diophantischen Approximationen*, Prace Mat.-Fiz. 91–106 (1928–1929)
30. V. Jarník, Diophantischen Approximationen und Hausdorffsches Mass. Mat. Sbornik **36**, 91–106 (1929)
31. J.-P. Kahane, R. Salem, *Ensembles parfaits et séries trigonométriques* (Hermann, Paris, 1994)
32. R. Kaufman, Continued fractions and Fourier transforms. Mathematika **27**(2), 262–267 (1980)
33. D. Kleinbock, B. Weiss, Badly approximable vectors on fractals. Israel J. Math. **149**, 137–170 (2005)
34. D. Kleinbock, E. Lindenstrauss, B. Weiss, *On fractal measures and Diophantine approximation*, Selecta Math. (N.S.) **10**(4), 479–523 (2004)
35. S. Kristensen, *Approximating numbers with missing digits by algebraic numbers*. Proc. Edinb. Math. Soc. (2) **49**(3), 657–666 (2006)
36. S. Kristensen, R. Thorn, S. Velani, Diophantine approximation and badly approximable sets. Adv. Math. **203**(1), 132–169 (2006)
37. J. Levesley, C. Salp, S. Velani, On a problem of K. Mahler: Diophantine approximation and Cantor sets. Math. Ann. **338**(1), 97–118 (2007)
38. P. Mattila, *Geometry of Sets and Measures in Euclidean Spaces* (Cambridge University Press, Cambridge, 1995)
39. H. Minkowski, Geometrie der Zahlen, Leipzig and Berlin (1896)
40. H. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis* (American Mathematical Society, Providence, 1994)
41. G.K. Pedersen, *Analysis Now* (Springer, New York, 1989)
42. A.D. Pollington, S. Velani, On a problem in simultaneous Diophantine approximation: Littlewood's conjecture. Acta Math. **185**(2), 287–306 (2000)
43. E.T. Poulsen, A simplex with dense extreme points. Ann. Inst. Fourier. Grenoble **11**, 83–87 (1961)
44. W.M. Schmidt, *Über die Normalität von Zahlen zu verschiedenen Basen*. Acta Arith. **7**, 299–309 (1961/1962)
45. W.M. Schmidt, On badly approximable numbers and certain games. Trans. Am. Math. Soc. **123**, 178–199 (1966)
46. C. Series, *The modular surface and continued fractions*. J. London Math. Soc. (2) **31**(1), 69–80 (1985)
47. V.G. Sprindžuk, *Metric Theory of Diophantine Approximations* (Wiley, London, 1979)
48. E. Wirsing, Approximation mit Algebraischen Zahlen beschränkten Grades. J. Reine Angew. Math. **206**, 67–77 (1961)
49. A. Ya, *Khintchine* (Dover Publications Inc, Mineola, NY, Continued fractions, 1997)

http://www.springer.com/978-3-319-48816-5

Diophantine Analysis
Course Notes from a Summer School
Steuding, J. (Ed.)
2016, XI, 232 p. 34 illus., 31 illus. in color., Hardcover
ISBN: 978-3-319-48816-5
A product of Birkhäuser Basel