Reflecting walls and dissipating tiles

Billiards and fractal percolation

Reflecting walls and dissipating tiles

Billiards and fractal percolation

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 1 maart 2013 om 10:00 uur

door

Hendrik DON
wiskundig ingenieur
geboren te Rhenen

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. F.M. Dekking

Copromotor: Dr. C. Kraaikamp

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. F.M. Dekking,	Technische Universiteit Delft, promotor
Dr. C. Kraaikamp,	Technische Universiteit Delft, copromotor
Prof. dr. K. Simon,	Technical University of Budapest, Hongarije
Prof. dr. V. Komornik,	Universite de Strasbourg, Frankrijk
Prof. dr. F.H.J. Redig,	Technische Universiteit Delft
Prof. dr. E.A. Cator,	Radboud Universiteit Nijmegen
Prof. dr. M.S. Keane,	Wesleyan University, Verenigde Staten
Prof. dr. ir. G. Jongbloed,	Technische Universiteit Delft, reservelid

ISBN: 978-94-6182-245-1

Contents

Chapter 1

Introduction

This thesis deals with subsets of intervals and squares. The subsets we consider have in common that they are all constructed by very simple rules. Basically we have two types of rules and they are mentioned in the title of this thesis: reflecting walls and dissipating tiles.

The reflecting walls construction rule is a deterministic one. We study the orbit of a ball moving in an interval or square respectively. Its movement is determined by the fact that the walls are reflecting. In the one-dimensional case, we place walls at the endpoints of an interval and imagine an idealized ball to be bouncing back and forth, see Figure 1.1. By the orbit of the ball we mean the sequence of its landing points. This sequence has some nice properties, in particular we will study how it induces partitions of the interval into subintervals.

In the two-dimensional case, the ball is moving on a rectangle that is surrounded by walls. Such a system is known under the name 'billiard'. From a ball game player's perspective, the relation between the one-dimensional setting and the billiard is immediate. We will show that also from a mathematician's point of view, the billiard is a natural generalization of the one-

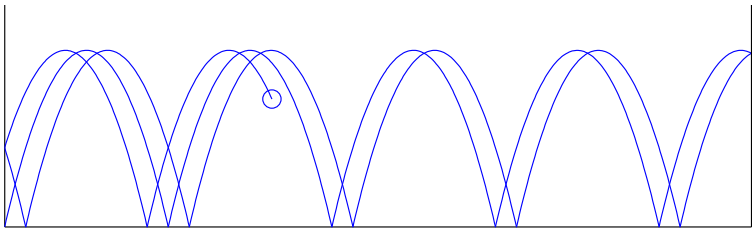

Figure 1.1: A ball bouncing back and forth between two walls.

dimensional case. Comparing Figure 1.1 with Figure 1.2 already lifts a little corner of the veil. In this two-dimensional setting, we again investigate how the orbit of the ball partitions the area where it is moving into smaller pieces.

Our second construction rule involves dissipating tiles. Here the construction procedure is probabilistic. We start with the unit interval or unit square and divide it into some equally sized subintervals or subsquares. Then we randomly select some of them that will be thrown out of our set. In the subintervals or subsquares that are not discarded, we do the same procedure on a smaller scale. Iterating this gives a limiting set that is a random fractal. In the literature these sets are known as Cantor sets or as fractal percolation.

For the one-dimensional case, our main question concerns the algebraic difference of two independent random Cantor sets F_1 and F_2. This is the set of all numbers that can be written as $x - y$, where $x \in F_1$ and $y \in F_2$. In a setting where always the same fraction of subintervals survives, we find conditions for the difference set to contain an interval with positive probability.

In two dimensions, we consider the fractal percolation process, as introduced by Mandelbrot. Suppose we have a square floor tiled by M^2 black squares. For each of them, we flip a (biased) coin that decides whether it is discarded or not. Those tiles that did not dissipate are subdivided into M^2 smaller squares and again we flip the coin for each of them. This is repeated ad infinitum. What does the limiting pattern at the floor look like? Of course this highly depends

on the bias of the coin. It is a classical result that this process exhibits a phase transition: if the retaining probability p is smaller than some critical probability p_c, then the limiting set will be totally disconnected. We only see some black dots here and there. If p is greater than or equal to p_c, then with positive probability the limiting set still connects opposite sides of the square we started with. Over the past decades, it turned out to be a hard problem to find the exact value of p_c. In this thesis we discuss techniques to find better bounds for p_c.

1.1 Reflecting walls

Sequences of numbers have been widely studied in mathematics over a long period of time. Already in the 6th century, the Fibonacci sequence was known to Indian mathematicians. Sequences are not only a powerful tool in describing and understanding the world around us, they can also have quite fascinating mathematical properties.

A classical example is the following. Take an arbitrary real number α and consider the arithmetic progression

$$0, \alpha, 2\alpha, 3\alpha \ldots, n\alpha.$$

For each of the terms in this sequence, calculate the distance to the largest integer smaller than this multiple. This gives a new sequence, with values in $[0, 1]$. It was conjectured by Steinhaus that these numbers partition the interval $[0, 1]$ into subintervals which have at most three different lengths. The first proof was provided by Sós [26]. This result can also be interpreted in terms of rotations on a circle. If we walk on the circumference, each time making steps of the same size, then after n steps the distances between footprints next to each other can have at most three different values. This result is known as the Three Gap Theorem. In the literature, various other statements about this sequence have been proved. For example, if there are three different lengths, then one of

them is the sum of the other two.

1.1.1 Bouncing on an interval

In the setting of the Three Gap Theorem, multiples of an irrational number are rounded down to an integer. Instead of rounding down to an integer, we investigate the sequence obtained by taking the distance to the nearest integer. This leads to results closely connected to the Three Gap Theorem, see Chapter 2 of this thesis. The resulting sequence in $[0, 1/2]$ can be thought of as landing points of an idealized ball bouncing between reflecting walls at 0 and $1/2$. See Figure 3.2 for an illustration of the situation when we put the numbers again on the circumference of a circle. In this setting, we will prove a Four Gap Theorem. The first n numbers in this sequence induce a partition of $[0, 1/2]$ into subintervals having at most four different lengths. All possible relations between these lengths are listed in the main theorem of Chapter 2. The proof provided in this chapter is self-contained. However, as was noted by Komornik, a shorter and more natural proof using the Three Gap Theorem can be given, see Chapter 3.

1.1.2 Bouncing on a billiard

In Chapter 3 we turn our attention to a two-dimensional equivalent of the Four Gap Theorem. Here we study the orbit of a ball on a rectangular billiard that starts from one of the corners. Such an orbit can be obtained by taking a halfline in $\mathbb{R}^2$ that starts in the origin. Taking distances to nearest integers in both coordinates transforms this halfline into a billiard orbit on the square $[0, 1/2]^2$, as is

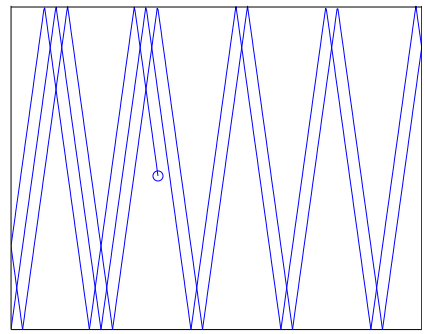

Figure 1.2: The orbit of a billiard ball.

explained in Section 3.3. This insight, combined with the Three Gap Theorem are our main tools to derive some results on billiard orbits. If the billiard orbit is truncated somewhere at the boundary, then it partitions the square into polygons having at most 13 different sizes. This upper bound is sharp.

A special case arises when the shooting angle is rational. In that case the orbit is periodic. After one period, the partition does not change any more. In this regime, the polygons have at most three different sizes. However, within the first period the maximum of thirteen can still be reached. On the other hand also irrational angles can be exceptional, in the sense that for some irrational angles the maximum of thirteen is never reached.

1.2 Dissipating tiles

In this section we introduce fractal sets and discuss a central notion in fractal geometry: the Hausdorff dimension of a fractal set. We introduce Cantor sets and discuss about the algebraic difference of two Cantor sets. In particular we discuss the case when these Cantor sets are random. Next we move on to fractal percolation, where we especially pay attention to the value of the probability parameter at which the phase transition of this process occurs.

1.2.1 Fractal sets

Many phenomena in nature show highly irregular structures. Perfect circles, Platonic solids, straight lines and continuous curves are objects that often give an idealized rather than an accurate description of real-world objects. For this reason, fractal geometry came into the picture in the seventies, not in the last place because of the work of Benoit Mandelbrot [21]. He describes a fractal as a shape for which the degree of irregularity is identical at all scales. Especially those fractal sets that involve randomness are useful to describe the irregular

patterns around us.

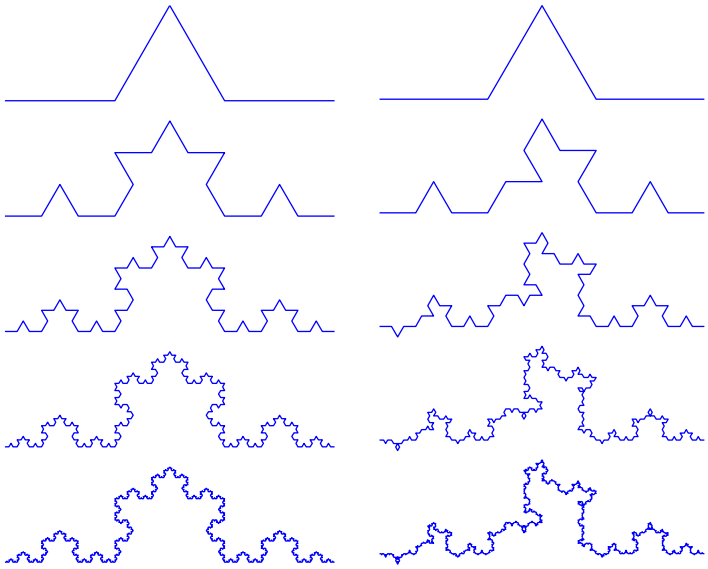

Figure 1.3: Left: construction of the Koch curve. Right: construction of a random Koch curve

Fractal sets can often be obtained by iterating some substitution procedure on smaller and smaller scales. In Figure 1.3, left panel, we see a classical example of this: the Koch curve. It is constructed as follows. We start with the unit interval, and divide it into three line segments of equal length. Then we draw an equilateral triangle, that has the middle segment as its base and points outwards. Finally, we remove the base of the triangle. Repeat this procedure in each of the four resulting segments. This produces a curve that on all scales displays smaller copies of itself.

An important notion in fractal geometry is the Hausdorff dimension of a fractal set F. Informally speaking, the idea here is to cover F by balls of diameter δ. If the number of balls we need to do so is of order δ^{-n}, then n is the Hausdorff

dimension of F. Formalizing this, we define

$$\mathcal{H}^s_\delta(F) := \inf\left\{\sum_{i=1}^{\infty} |U_i|^s : |U_i| \leq \delta \text{ for all } i, F \subseteq \bigcup_{i=1}^{\infty} U_i\right\},$$

where $|U_i|$ stands for the diameter of the set U_i in the Euclidian metric. Letting δ tend to 0 gives the s-dimensional Hausdorff measure of the set F:

$$\mathcal{H}^s(F) := \lim_{\delta \to 0} \mathcal{H}^s_\delta(F).$$

If s is large, then one can construct coverings $\{U_i\}_{i=1}^{\infty}$ by small sets for which the sum $\sum_{i=1}^{\infty} |U_i|^s$ is arbitrarily small. On the other hand, if s is small and we are forced to cover by sets U_i with small diameter, then the sum $\sum_{i=1}^{\infty} |U_i|^s$ will be large. In fact one can prove that there is at most one value of s for which $0 < \mathcal{H}^s(F) < \infty$, and this value is defined to be the Hausdorff dimension $\dim_H(F)$ of the set F:

$$\dim_H(F) := \inf\{s : \mathcal{H}^s(F) = 0\} = \sup\{s : \mathcal{H}^s(F) = \infty\}.$$

The Hausdorff dimension of the Koch curve can be easily found by using that it is equal to its so-called similarity dimension. If an object consists of N (nearly) disjoint copies of itself that are scaled down by a factor r, then the Hausdorff dimension of this object equals $\log(N)/\log(r)$. As a consequence, the Koch curve has Hausdorff dimension $\log(4)/\log(3) \approx 1.26$.

The above construction of the Koch curve is deterministic. One could also change the procedure to get a *random fractal*. Each time we draw an equilateral triangle that has the middle segment as its base, we flip a fair coin to decide whether it points outwards or inwards. This gives a completely different behavior, as is seen in the right part of Figure 1.3. We lose the strictly self-similar structure, but nevertheless we still have copies of the same random object at all scales. Such random fractals are often more suitable to model real-life phenomena and their unpredictable fluctuations than the deterministic ones. For

example this random Koch curve looks more like a real coastline than the deterministic version. In this thesis random fractals play a central role in Chapter 4 and Chapter 5.

1.2.2 Algebraic differences of Cantor sets

The algebraic difference of a two sets of real numbers A and B is given by

$$A - B = \{a - b : a \in A, b \in B\}.$$

An interesting case arises if we take A and B to be fractal sets. Let us see what happens if we for A and B take the triadic Cantor set. This set is obtained by repeatedly removing open middle thirds of a collection of intervals, starting with the unit interval. A way to make this precise is the following: let $C_0 = [0, 1]$ and define

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3} \right),$$

and now the Cantor set C is equal to the intersection of all its approximants:

$$C = \bigcap_{n=0}^{\infty} C_n.$$

Alternatively, one could think of the Cantor set as those points on the deterministic Koch curve that are on its baseline, so in the left part of Figure 1.3 one can also see approximants of the Cantor set. The Hausdorff dimension of the Cantor set can be easily found: $\dim_H(C) = \log(2)/\log(3) \approx 0.63$.

The Cantor set is full of holes and has zero Lebesgue measure, but nevertheless it has the remarkable property that any real number between -1 and 1 can be represented as the difference between two numbers in the Cantor set. In other words,

$$C - C = [-1, 1]. \tag{1.1}$$

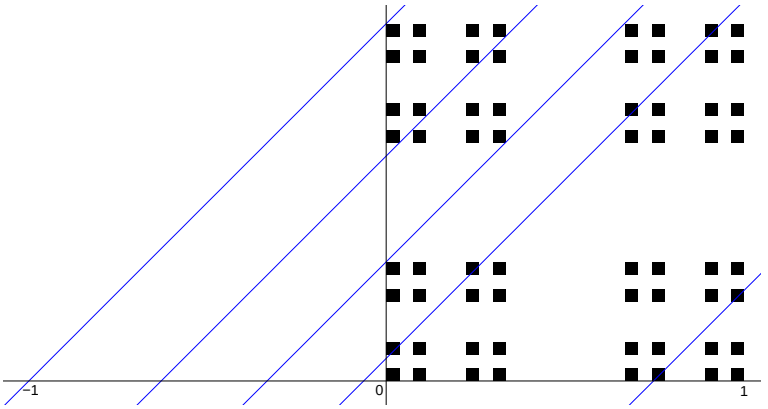

Figure 1.4: Projecting $C_3 \times C_3$ under a $45°$ degree angle gives the full interval $[-1, 1]$.

There are several ways to prove this. One of them uses the ternary characterization of the Cantor set: a real number from $[0, 1]$ belongs to the Cantor set if and only if it has a ternary expansion that contains only digits 0 and 2. There is also a geometric proof, that served as the basis for the results in Chapter 4 of this thesis. Consider the line in $\mathbb{R}^2$ through the point (a, b) that has slope 1. Points (x, y) on this line satisfy the equation $x - a = y - b$. Hence also the point $(a - b, 0)$ lies on this line. This means that the algebraic difference $C - C$ can be found by projecting the Cartesian product $C \times C$ on the x-axis under a $45°$ angle. Figure 1.4 shows $C_3 \times C_3$ and the way of projection on the x-axis.

The set $C \times C$ is known as Cantor dust, a term introduced by Mandelbrot. Any line with slope 1 that intersects the x-axis between -1 and 1 has non-empty intersection with $C_0 \times C_0$, which is just the unit square $[0, 1]$. It is easy to see that it also has to intersect at least one of the four corner squares, that together build $C_1 \times C_1$. By self-similarity of the set, it then also intersects $C_2 \times C_2$ and repeating this shows that the line intersects $C_n \times C_n$ for all n, and therefore it also intersects $C \times C$, proving (1.1).

The situation gets a lot more complicated if we introduce randomness here.

There are several ideas to construct random Cantor sets. The size and the position of the surviving intervals can be made stochastic, as was done by Larsson in his thesis [17, 18]. Another option is to randomize the selection of intervals that survive. Dekking and Kuijvenhoven studied M-adic random Cantor sets in which each of the M subintervals survives independent of all others [8].

Also in Chapter 4 of this thesis, we study M-adic random Cantor sets with randomly selected intervals. Our construction procedure is the following: start with the unit interval and divide it into M subintervals of length $1/M$. Now we either discard all subintervals, leaving us the empty set, or a fixed fraction of subintervals (m out of M) survives. The selection of the m surviving subintervals is uniformly at random. In each surviving subinterval, repeat this process. Now intervals are not selected independently any more, and therefore this class is called *correlated fractal percolation*.

We investigate the question under which conditions the algebraic difference of two independent realizations F_1 and F_2 of correlated fractal percolation almost surely contains an interval and when not. Palis [23] conjectured that 'generically' it should be true that the algebraic difference of two Cantor sets contains an interval if the sum of their Hausdorff dimensions exceeds 1. We prove that for correlated fractal percolation a strong version of this conjecture holds: $F_1 - F_2$ contains an interval almost surely if and only if $\dim_H(F_1) + \dim_H(F_2) > 1$. The proof involves a thorough study of 45° degree projections of randomized Cantor dust.

1.2.3 Fractal percolation and its critical value

The topic that is covered in Chapter 5 of this thesis is the two-dimensional version of random Cantor sets, called fractal percolation or Mandelbrot percolation. Here we start with a square and repeatedly divide into subsquares and select some that survive and some that are discarded. This model has received quite some attention, mainly at the end of the twentieth century. Real world

phenomena for which it has been proposed as a model include for example turbulence and porous media.

Fractal percolation is constructed as follows. Choose a retaining probability $p \in (0, 1)$ and choose a natural number $M \geq 2$ as division base. Let K_0 be the unit square. Divide K_0 into M^2 subsquares of equal size, and let each of them survive with probability p, independent of the $M^2 - 1$ other subsquares. Let K_1 be the set consisting of all subsquares that survived. Similarly, we obtain K_2 by dividing the squares in K_1 into M^2 smaller squares that again survive with probability p, independent of all others. Repeating this procedure gives a sequence $K_0, K_1, \ldots$ of random sets. In Figure 1.5 a realization of $K_1, \ldots K_8$ is shown for $M = 2$ and $p = 0.75$.

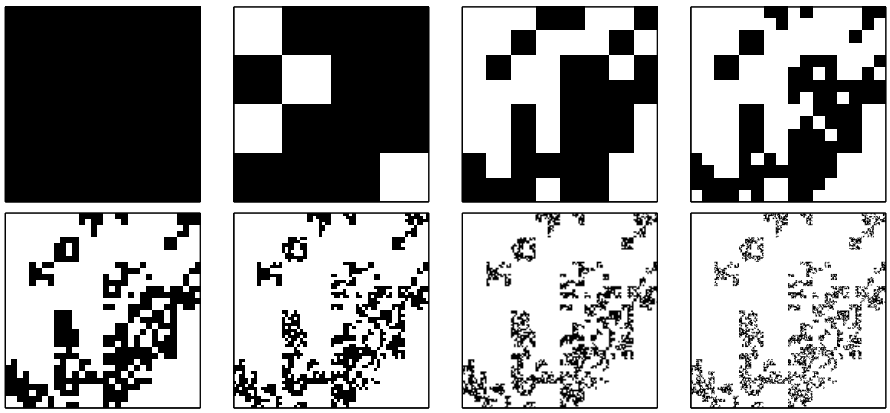

Figure 1.5: A realization of the random sets $K_1, \ldots K_8$, with base $M = 2$ and survival probability $p = 0.75$. In this thesis we investigate the connectivity of the liiming set.

The sequence $(K_n)_{n=0}^{\infty}$ is monotone decreasing and therefore it converges to a limit set $K = \bigcap_{n=0}^{\infty} K_n$. Conditioned on the event that K is non-empty, its Hausdorff dimension $\dim_H(K)$ is almost surely equal to $\log(pM^2)/\log(M)$.

Our primary interests lie in the connectivity properties of the limiting set K. We say that K percolates if it contains a connected component intersecting both the

left side and the right side of the unit square. Define the percolation function $\theta(M,p)$ as the probability that K percolates if we use base M and survival probability p. Then obviously $\theta(M,p)$ is increasing in p and $\theta(M,0) = 0$ and $\theta(M,1) = 1$. So we can define a critical value as follows:

$$p_c(M) := \inf\{p : \theta(M,p) > 0\}.$$

There is a pretty large difference between the subcritical behavior and the (super)critical behavior. The function $\theta(M,p)$ is known to be discontinuous at p_c, see Figure 1.6. Chayes, Chayes and Durrett [4] prove that in the subcritical regime $p < p_c$ the set K is almost surely "dust-like", that is to say, the largest connected component is a point. If $p \geq p_c$, then there is a positive probability that K percolates. So, the process exhibits a first order phase transition at p_c.

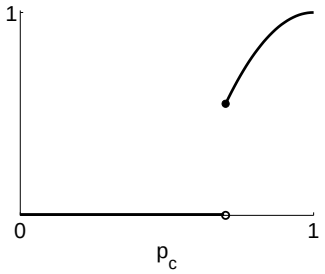

Figure 1.6: Rough sketch of $\theta(M,p)$.

The problem to find p_c turned out to be quite a hard one. The first non-trivial lower bound is $p_c(M) > 1/M^2$, since the surviving squares form a branching process with mean pM^2. This dies out almost surely for $p \leq 1/M^2$, such that K is empty almost surely. Another branching process argument (see Chapter 5) enabled Chayes, Chayes and Durrett to prove that $p_c(M) > 1/\sqrt{M}$. For the case $M = 2$, White [31] sharpened this to $p_c(2) \geq 0.810$ by adding countably many straight line segments to the set K in order to simplify its connectivity structure.

To establish upper bounds, the first idea (which works for $M \geq 3$ and is again due to Chayes et al) was the following. We call a realization of K_1 *good* if in the first construction step at least $M^2 - 1$ subsquares survive. A realization of K_n can be obtained by taking a realization of K_1 and then replacing each square that survived independent of all others by a scaled realization of K_{n-1}. Define

a realization of K_n to be good if at least $M^2 - 1$ of these scaled realizations are good.

It is not hard to see that K_n percolates if it is a good realization. The argument roughly goes as follows: two neighboring good squares share an edge, and along this edge they both have M subsquares, making M pairs. At least one of these pairs consists of two good squares, since $M \geq 3$. Consequently, a connection crossing this edge is preserved. This is sufficient to have K_n percolating for all n, and hence also K percolates. Filling in the details in this reasoning leads to the following result: $p_c(M) \leq p^*(M)$ for $M \geq 3$, where $p^*(M)$ is the infimum over p for which the polynomial

$$(px)^{M^2} + M^2(px)^{M^2-1}(1 - px)$$

has a fixed point in the half open interval $(0, 1]$. An upper bound for $M = 2$ can be obtained by a coupling with the case $M = 4$. For $M = 3$ this leads to $p_c(3) < 0.993$. Dekking and Meester [9] described the problem using multi-valued substitutions and sharpened the upper bound to $p_c(3) < 0.991$, which was improved in the thesis of van der Wal to $p_c(3) < 0.965$.

In Chapter 5, we combine several insights of the aforementioned work. In particular, we extend the idea to classify realizations, where we do not restrict ourselves to good and bad realizations, but we map realizations to a finite alphabet. The choice of the alphabet is inspired by van der Wal's work. Together with a coupling with site percolation, this permits us for all M to construct a principally computable sequence of lower bounds for $p_c(M)$ that converges to $p_c(M)$. We also show how to compute upper bounds using similar ideas. This leads to the following numerical results:

$$0.881 < p_c(2) < 0.993, \quad 0.784 < p_c(3) < 0.940, \quad \text{and} \quad p_c(4) < 0.972.$$

1.3 List of publications

The research for this thesis has led to the following publications:

1. Don, H. – *On the distribution of the distances of multiples of an irrational number to the nearest integer*, Acta Arithmetica 139 (2009), 253–264.
2. Dekking, F.M. and Don, H. – *Correlated fractal percolation and the Palis conjecture*, Journal of Statistical Physics. 139 (2010), no. 2, 307-325.
3. Don, H. – *Polygons in billiard orbits*, Journal of Number Theory 132 (2012), no. 6, 1151-1163.
4. Don, H. – *New methods to determine the critical probability in fractal percolation*, submitted to Random Structures and Algorithms (2012). (arXiv:1210.4150)

Chapter 2

On the distribution of distances of multiples of an irrational number to the nearest integer

2.1 Introduction

Take an arbitrary irrational number α and compute for the first n multiples the distance to the nearest integer. What can we say about the distribution of this sequence in the interval $[0, 1/2]$? In this chapter we study the partition of the interval $[0, 1/2]$ induced by this sequence. The main result (Theorem 2.2) states that this sequence divides the interval in subintervals which can take at most four different lengths. This result is strongly related to the Three Gap Theorem,

which states that for α irrational and $n \in \mathbb{N}$, the numbers

$$\{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{n\alpha\} \tag{2.1}$$

divide the interval $[0, 1]$ in subintervals of at most three different lengths. Here $\{x\} = x - \lfloor x \rfloor = x \bmod 1$ is the fractional part of x. The Three Gap Theorem was originally a conjecture of H. Steinhaus. Proofs were offered by various authors, for example by Sós [26], Świerckowski [29], Surányi [28], Slater [25] and van Ravenstein [24].

We start with Theorem 2.1, a variation on the Three Gap Theorem, which states that if we divide the interval $[0, 1]$ in subintervals by the numbers

$$\{\alpha\}, \{-\alpha\}, \{2\alpha\}, \{-2\alpha\}, \ldots, \{n\alpha\}, \{-n\alpha\} \tag{2.2}$$

then the subintervals again have at most three different lengths. We give an elementary proof for this theorem.

From Theorem 2.1 we extract the main result, Theorem 2.2. This 'Four Gap Theorem' gives an analogous statement about the distances to the nearest integers of the multiples of α: the numbers

$$||\alpha||, ||2\alpha||, ||3\alpha||, \ldots, ||n\alpha|| \tag{2.3}$$

divide the interval $[0, \frac{1}{2}]$ in subintervals of at least two and at most four different lengths, where $||x||$ denotes the distance from x to the nearest integer. Here the number four is the best possible. We also derive some properties of the lengths of the subintervals in which $[0, \frac{1}{2}]$ is divided.

2.2 A variation on the three gap theorem

If we consider not only the fractional parts of the positive multiples of an irrational number α, but also of the negative multiples, we have the following result:

Theorem 2.1. *Let α be an irrational number between 0 and 1, and let $n \in \mathbb{N}, n \geq 1$. For the first n numbers in the sequence*

$$S_\alpha : \{\alpha\}, \{-\alpha\}, \{2\alpha\}, \{-2\alpha\}, \{3\alpha\}, \{-3\alpha\}, \ldots \tag{2.4}$$

the following assertions hold:

1. *They divide the interval $[0, 1]$ in subintervals of either two or three different lengths, $l_1 > l_2 (> l_3)$. If we have three different lengths, $l_1 > l_2 > l_3$, then $l_1 = l_2 + l_3$.*
2. *By adding the $(n + 1)$th element of the sequence S_α to the partition of $[0, 1]$, one of the subintervals of length l_1 is divided in a subinterval of length l_2 and a subinterval of length $l_1 - l_2$.*

Before proving the assertions we make some preparations by collecting observations that will be helpful in proving the assertions. Note that it makes no difference in Theorem 2.1 if we consider the open interval $\big(0, 1\big)$.

First note that for $x \in \mathbb{R}\backslash\mathbb{Z}$ we have $\{-x\} = 1-\{x\}$, so the partition induced by the first $2n$ terms of the sequence S_α is symmetric with respect to $\frac{1}{2}$. This also means that without loss of generality we may assume that $\alpha < \frac{1}{2}$. Sometimes α will be called the *step size*.

It will prove useful to introduce some notation and definitions. For $n \geq 1$, $S_\alpha(n)$ denotes the nth term of S_α. For each $k \in \mathbb{N}, k \geq 1$ let $n_k(\alpha)$ be the unique integer for which:

$$n_k(\alpha)\alpha < k < (n_k(\alpha) + 1)\alpha. \tag{2.5}$$

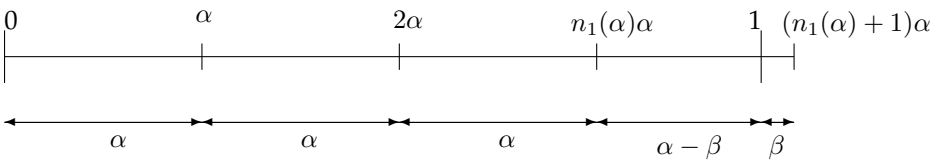

Figure 2.1: The first steps in the partition process.

Since α is irrational, k can never be a multiple of α. Define β by

$$\beta := (n_1(\alpha) + 1)\alpha - 1. \tag{2.6}$$

Note that $\beta = \{(n_1(\alpha) + 1)\alpha\}$. Figure 2.1 illustrates these definitions in case $n_1(\alpha) = 3$.

Definition 2.1. *For $k \in \mathbb{N}, k \geq 1$ the kth cycle of the sequence S_α consists of all those fractional parts $\{m\alpha\}, m \in \mathbb{Z}$ for which $k - 1 < |m\alpha| < k$, or equivalently $n_{k-1}(\alpha) < |m| \leq n_k(\alpha)$.*

Observe that a cycle consists of at least four partition points, because we assumed that $\alpha < \frac{1}{2}$. We are going to use this observation later. The next definition concerns intervals which are partitioned in the same way.

Definition 2.2. *For $0 \leq a, b < 1$, $y \leq \min\{1 - a, 1 - b\}$ and $n \in \mathbb{N}$ we write $\big(a, a + y\big)(n) \simeq \big(b, b + y\big)(n)$ if for all $x \in \big(0, y\big)$ the following equivalence holds:*

$$\exists k_1 \in \mathbb{Z}, |k_1| \leq n \text{ such that } a + x = \{k_1\alpha\}$$

$$\Longleftrightarrow$$

$$\exists k_2 \in \mathbb{Z}, |k_2| \leq n \text{ such that } b + x = \{k_2\alpha\}.$$

Note that $\simeq$ is an equivalence relation on the class of partitioned open subin-

tervals of $\big(0,1\big)$. If we replace $b+x$ by $b+y-x$ in Definition 2.2, we get an equivalence for an interval and the mirror image of an other interval. If two intervals satisfy this adjusted definition, we will write $\big(a,a+y\big)(n) \overset{m}{\simeq} \big(b,b+y\big)(n)$.

Now let us investigate what happens in the interval $\big(0,\alpha\big)$. Note that S_α is a sequence in the *open* interval $\big(0,1\big)$. Therefore also here we investigate which values we get in the *open* interval $\big(0,\alpha\big)$. For $k \in \mathbb{N}, k \geq 1$ the interval $\big(k, k+\alpha\big)$ contains exactly one positive multiple of α and the interval $\big(-k,-k+\alpha\big)$ contains exactly one negative multiple of α. Hence, in each cycle we get two values in $\big(0,\alpha\big)$, one of them being the fractional part of a positive multiple of α and the other being the fractional part of a negative multiple of α. The first cycle is the only exception, since there is no positive multiple of α in $\big(0,\alpha\big)$.

The first positive multiple of α for which the fractional part is in $\big(0,\alpha\big)$ is $(n_1(\alpha)+1)\alpha = 1+\beta$, which gives β as a first hit in $\big(0,\alpha\big)$. Because $1+\beta$ is a positive multiple of α, for $k \in \mathbb{N}$ the number $k+k\beta$ is also a positive multiple of α. The fractional parts of these numbers are fractional parts of multiples of β. As long as $k\beta < \alpha$ this gives hits in $\big(0,\alpha\big)$. As soon as $k\beta$ exceeds α, i.e. when $k = \lfloor \alpha/\beta \rfloor + 1$, we leave the interval $\big(0,\alpha\big)$, but in that case we already had hit the value $k\beta - \alpha$. This is exactly how it continues all the time: each next hit in $\big(0,\alpha\big)$ is shifted β in positive direction and as soon as we leave the interval, we come back modulo α. Hence, for every positive integer k, the kth hit in $\big(0,\alpha\big)$ by the fractional part of a positive multiple of α is $k\beta$ mod α.

The first negative multiple of α for which the fractional part is in $\big(0,\alpha\big)$ is $-n_1(\alpha)\alpha$, giving the value $\{-n_1(\alpha)\alpha\} = 1 - \{n_1(\alpha)\alpha\} = 1 - n_1(\alpha)\alpha = \alpha - \beta$. Each next hit in $\big(0,\alpha\big)$ is shifted β to the left until $\alpha - k\beta$ dives under 0. In that case we leave $\big(0,\alpha\big)$, but the previous hit was $\alpha - k\beta + \alpha$, which is in $\big(0,\alpha\big)$. Hence, the kth hit in $\big(0,\alpha\big)$ by the fractional part of a negative multiple of α is $\alpha - (k\beta \text{ mod } \alpha)$.

By noting that the hits by fractional parts of positive and negative multiples of

α are alternating we see that in $\big(0,\alpha\big)$ we get the following sequence of hits:

$$\alpha-(\beta \bmod \alpha), \beta \bmod \alpha, \alpha-(2\beta \bmod \alpha), 2\beta \bmod \alpha,$$
$$\alpha-(3\beta \bmod \alpha), 3\beta \bmod \alpha, \ldots \tag{2.7}$$

By multiplying each term by $1/\alpha$ we get

$$1-\Big(\frac{\beta}{\alpha} \bmod 1\Big), \frac{\beta}{\alpha} \bmod 1, 1-\Big(\frac{2\beta}{\alpha} \bmod 1\Big), \frac{2\beta}{\alpha} \bmod 1,$$
$$1-\Big(\frac{3\beta}{\alpha} \bmod 1\Big), \frac{3\beta}{\alpha} \bmod 1, \ldots \tag{2.8}$$

By defining $\tilde{\alpha} := 1-\dfrac{\beta}{\alpha}$, we can rewrite this as

$$\{\tilde{\alpha}\}, \{-\tilde{\alpha}\}, \{2\tilde{\alpha}\}, \{-2\tilde{\alpha}\}, \{3\tilde{\alpha}\}, \{-3\tilde{\alpha}\}, \ldots \tag{2.9}$$

Hence, (2.7) is a scaled version of the sequence S_α (with a different irrational step size). That means that the partition of the subinterval $\big(0,\alpha\big)$ has exactly the same structure and properties as the partition of $\big(0,1\big)$. The same self-similarity holds for the subintervals $\big(\alpha,2\alpha\big),\ldots,\big((n_{1/2}(\alpha)-1)\alpha, n_{1/2}(\alpha)\alpha\big)$, where $n_{1/2}(\alpha)\alpha$ is the largest multiple of α smaller than $1/2$. In these subintervals we get the same sequence (2.7), but now shifted by a multiple of α to the corresponding positions in the subinterval. By using symmetry we also find the same structure of lengths for the intervals $\big(1-n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)-1)\alpha\big),\ldots,\big(1-\alpha,1\big)$. These intervals are mirror images of the subintervals

$$\big(0,\alpha\big),\ldots,\big((n_{1/2}(\alpha)-1)\alpha, n_{1/2}(\alpha)\alpha\big).$$

Each cycle of S_α gives two hits in each of those intervals. We conclude that for all $k\in\mathbb{N}, k\geq 1$

$$\big(0,\alpha\big)(n_k(\alpha)) \simeq \ldots \simeq \big((n_{1/2}(\alpha)-1)\alpha, n_{1/2}(\alpha)\alpha\big)(n_k(\alpha)) \overset{m}{\simeq} \tag{2.10}$$
$$\big(1-n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)-1)\alpha\big)(n_k(\alpha)) \simeq \ldots \simeq \big(1-\alpha,1\big)(n_k(\alpha)).$$

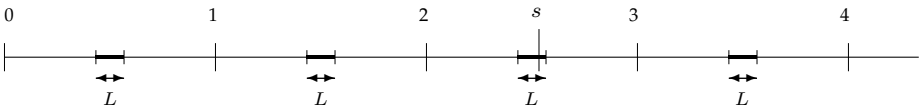

Figure 2.2: Each interval between two integers consists of two parts of length $n_{1/2}(\alpha)\alpha$ and a part of length L (bold). Fractional parts of numbers in the bold intervals are in I_m. The lengths of the bold parts to the left of s add up to α.

The only part which is not yet considered is the middle part of $\big(0,1\big)$: the interval $\big(n_{1/2}(\alpha)\alpha, 1-n_{1/2}(\alpha)\alpha\big)$, which will be denoted by I_m and its length by L. Denote the smallest positive multiple of α for which the fractional part is in I_m by s, see Figure 2.2. The Lebesque measure of the set

$$\{x \in (0,s) : \{x\} \notin I_m\} \tag{2.11}$$

is a multiple of $n_{1/2}(\alpha)\alpha$. Hence the Lebesque measure of the set

$$\{x \in (0,s) : \{x\} \in I_m\} \tag{2.12}$$

must also be a multiple of α. From the fact that s is the *smallest* number for which the measure of this set is a multiple of α it follows that its measure is exactly α. This implies that the first element of S_α which is in I_m is given by

$$\{s\} = n_{1/2}(\alpha)\alpha + (\alpha \bmod L). \tag{2.13}$$

For each next multiple of α giving a hit in I_m, a similar argument applies, but now the measure of the set in (2.12) increases with α for each next hit. We conclude that the positive multiples of α give the following sequence of hits in I_m:

$$\big(n_{1/2}(\alpha)\alpha + (k\alpha \bmod L)\big)_{k=1}^{\infty} \tag{2.14}$$

By symmetry we see that by adding the negative multiples of α too, we find

the following sequence of hits in I_m:

$$\begin{aligned} &n_{1/2}(\alpha)\alpha + (\alpha \bmod L), 1 - n_{1/2}(\alpha)\alpha - (\alpha \bmod L), \\ &n_{1/2}(\alpha)\alpha + (2\alpha \bmod L), 1 - n_{1/2}(\alpha)\alpha - (2\alpha \bmod L), \\ &n_{1/2}(\alpha)\alpha + (3\alpha \bmod L), 1 - n_{1/2}(\alpha)\alpha - (3\alpha \bmod L), \ldots, \end{aligned} \tag{2.15}$$

where the alternating order follows from the fact that the successor of $\{k\alpha\}$ in S_α is $\{-k\alpha\}$.

Subtract $n_{1/2}(\alpha)\alpha$ to get

$$\begin{aligned} &\alpha \bmod L, L - (\alpha \bmod L), 2\alpha \bmod L, \\ &L - (2\alpha \bmod L), 3\alpha \bmod L, L - (3\alpha \bmod L), \ldots \end{aligned} \tag{2.16}$$

Multiplying by $1/L$ yields

$$\begin{aligned} &\frac{\alpha}{L} \bmod 1, 1 - \Big(\frac{\alpha}{L} \bmod 1\Big), \frac{2\alpha}{L} \bmod 1, \\ &1 - \Big(\frac{2\alpha}{L} \bmod 1\Big), \frac{3\alpha}{L} \bmod 1, 1 - \Big(\frac{3\alpha}{L} \bmod 1\Big), \ldots \end{aligned} \tag{2.17}$$

This is exactly S_α, with step size α/L. It follows that (2.15) is a scaled and translated version of the sequence S_α with a different step size.

The next step is to find the relation between the behavior of the partition process in I_m and its complement. The intervals $\big(0, L\big)$ and I_m have the same length (by definition of L) and the distance between their left endpoints is a multiple of α. From this we can conclude that in each cycle a value $x \in \big(0, L\big)$ is hit if and only if in the same cycle the point $x + n_{1/2}(\alpha)$ is hit in I_m. This reasoning is also valid when $\big(0, L\big)$ and I_m are not disjoint (which is possible when $L > \alpha$). By noting that $\big(0, L\big)(n_1(\alpha)) \simeq I_m(n_1(\alpha))$ and using induction

on k it follows that $\forall k \in \mathbb{N}, k \geq 1$:

$$\big(0, L\big)(n_k(\alpha)) \simeq I_m(n_k(\alpha)). \tag{2.18}$$

In words: after each complete cycle the two intervals $\big(0, L\big)$ and I_m are partitioned in an equivalent way in the sense of Definition 2.2.

Proof of Theorem 2.1
To prove Theorem 2.1, we use induction on the cycle number k. Note that if the theorem holds for n, then to go to $n+1$ it suffices to check the second assertion of the theorem. We can see this as follows. If we had three lengths, then one of the longest subintervals is divided in two existing lengths, so we get nothing new. If we had two lengths, then we get one new length, being the difference of the two existing lengths. These remarks show that the 'at most three' part of the first assertion and the requirement $l_1 = l_2 + l_3$ in case of three lengths are not violated. The 'at least two' part of the first assertion of the theorem follows from the irrationality of α. If only one length is remaining, the interval $[0, 1]$ must be divided in equal parts. But in this case α would be a rational number.

-Step 1- The first step in our induction argument is to show that during the first cycle (containing the first $2n_1(\alpha)$ terms of S_α) always one of the longest subintervals is divided in two intervals of which one has the second length occurring before the division. The first number in the sequence S_α is $\{\alpha\}$, so after adding this first number the interval $\big(0, 1\big)$ is divided in two subintervals, one of length α and one of length $1 - \alpha$, where the latter is the longest in view of our assumption that $\alpha < 1/2$. So now this longest subinterval should be divided in a part of length α (the second length) and a remaining part. Because the second hit is $\{-\alpha\} = 1 - \alpha$ this is indeed the case. The process continues in the same way, each time reducing the length of the middle subinterval by α, until the length of the middle subinterval becomes smaller than α. Now this middle subinterval has length $\alpha - \beta$, by definition of β.

At this point we have two different lengths: α and $\alpha - \beta$. The situation is illus-

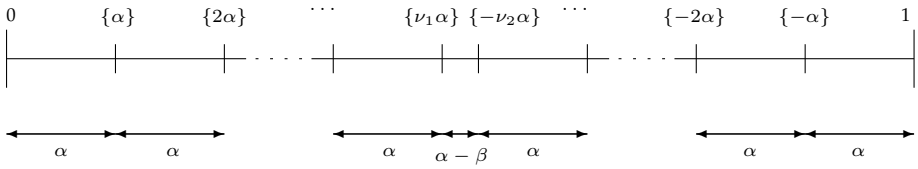

Figure 2.3: Halfway the first cycle: either $\nu_1 = \nu_2$ or $\nu_1 = \nu_2 + 1$.

trated by Figure 4.5. We now distinguish two cases.
If $\nu_1 = \nu_2$, then the next hit will be $\{(\nu_1 + 1)\alpha\}$, dividing an interval of length α in a part of length $\alpha - \beta$ (which was the second length) and a part of length β (a new length). Now we have three different lengths and the sum of the two smallest equals the largest, as required. The next hit now is $\{-(\nu_1 + 1)\alpha\}$ and again this divides an interval of length α in a part of length $\alpha - \beta$ and a part of length β. The partition process continues in this way as long as we are in the first cycle.
If $\nu_1 = \nu_2 + 1$, then the next hit will be $\{-\nu_1\alpha\}$ and also in this case all intervals of length α will successively be divided in a part of length $\alpha - \beta$ and a part of length β.
Hence we conclude that the theorem is valid for the first cycle.

-*Step 2*- The next step in the induction argument is to show that if the theorem holds in the first k cycles, then the theorem also holds in the next cycle. To prove this we use the observations made before, which state that the behavior of the partition process in each of the intervals

$$\begin{aligned}\big(0,\alpha\big),\ \ldots,\ \big((n_{1/2}(\alpha)-1)\alpha, n_{1/2}(\alpha)\alpha\big),\ \big(n_{1/2}(\alpha)\alpha, 1-n_{1/2}(\alpha)\alpha\big),\\ \big(1-n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)-1)\alpha\big),\ \ldots,\ \big(1-\alpha,1\big)\end{aligned} \tag{2.19}$$

has after rescaling the same properties as the behavior in the entire interval $\big(0,1\big)$. From now on we will call these intervals *elementary intervals*.

A crucial remark is that all boundaries (except 0 and 1) of the elementary intervals belong to the first cycle of S_α. This implies that (at any point in one of the next cycles) the subintervals in which $\big(0,1\big)$ is divided can only intersect *one* of the elementary intervals. This guarantees that to find all lengths of subintervals in $\big(0,1\big)$, it suffices to find all lengths in the elementary intervals.

For the elementary intervals we introduce the following abbreviations:

$$I^p := \big((p-1)\alpha, p\alpha\big), \tag{2.20}$$

$$I^{-p} := \big(1-p\alpha, 1-(p-1)\alpha\big), \tag{2.21}$$

where $1 \leq p \leq n_{1/2}(\alpha)$, $p \in \mathbb{N}$. Recall that for the middle elementary interval we already introduced the symbol I_m. The sequence of hits in an elementary interval I will be denoted by S_α^I. For example, $S_\alpha^{I^1}$ is equal to the sequence (2.7). Because these sequences are scaled and translated versions of S_α (possibly with a different step size), we can also here introduce cycles. Every element of the kth cycle of S_α is, for some I and l also an element of the lth cycle of S_α^I, where l may be different from k. We are going to use these cycles later, but we do not need to specify them explicitly.

Induction Hypothesis: Assume that for all α and some $k \geq 1$ the theorem holds as long as we are in one of the first k cycles of S_α.

Let $2n_k(\alpha) \leq n < 2n_{k+1}(\alpha)$, implying that $S_\alpha(n+1)$ is an element of the $(k+1)$th cycle of S_α. Consider the partition of $\big(0,1\big)$ in subintervals by the first n terms of S_α. Denote the lengths of the subintervals by $l_1 > l_2 (> l_3)$. To prove that the theorem holds for the $(k+1)$th cycle, it suffices to show that the following three requirements are satisfied:

Requirement 1: If $S_\alpha(n+1)$ is the very first hit in an elementary interval, then it splits a subinterval of length l_1 into subintervals of length l_2 and $l_1 - l_2$.

Requirement 2: If $S_\alpha(n+1) \in I$, where I is one of the elementary intervals, then I contains a subinterval of length l_1 just before $S_\alpha(n+1)$ is added.

Requirement 3: If $S_\alpha(n+1)$ is not the very first hit in an elementary interval, denote the two largest lengths *in this elementary interval* by $\hat{l}_1 > \hat{l}_2$. Then $S_\alpha(n+1)$ splits a subinterval of length $\hat{l}_1$ into subintervals of length $\hat{l}_2$ and $\hat{l}_1 - \hat{l}_2$.

First we argue why it is sufficient that these three requirements hold and then we check each of them in the substeps below.

The idea of the proof is to use the self-similar structure by applying the induction hypothesis to the elementary intervals. The theorem only gives an assertion about the division in subintervals if we have already at least two lengths. Hence, our induction hypothesis makes no statement about the very first hit in an elementary interval. Therefore, in *Substep 2.1* we start by checking that in each of the elementary intervals the partition process starts in the right way, as indicated by Requirement 1. Suppose I is the elementary interval containing $S_\alpha(n+1)$. Denote the lengths occurring in I just before adding $S_\alpha(n+1)$ by $\hat{l}_1 > \hat{l}_2 (> \hat{l}_3)$. Then the maximal length in I should equal the maximal length in $[0,1]$: $\hat{l}_1 = l_1$, which is Requirement 2. Since all lengths in I are also lengths in $[0,1]$, either $\hat{l}_2 = l_2$ or $\hat{l}_2 = l_3$. In both cases, splitting an interval of length $\hat{l}_1$ in two subintervals of lengths $\hat{l}_2$ and $\hat{l}_1 - \hat{l}_2$ (as is demanded in Requirement 3) is the same as splitting an interval of length l_1 in two subintervals of lengths l_2 and $l_1 - l_2$, since $l_3 = l_1 - l_2$. We conclude that these three requirements are sufficient to complete the proof. The induction hypothesis is only needed to prove Requirement 3.

-*Substep 2.1*- All elementary intervals, except I_m if $L < \alpha$, get at least one hit in the first cycle of S_α. So here we have no problems, because we already checked that the theorem holds for the first cycle. Suppose that $L < \alpha$ and that $S_\alpha(n+1)$ is the first value we hit in I_m. Then $S_\alpha(n+1)$ can be written as $n_{1/2}(\alpha)\alpha + x$, where $x \in (0, L)$. In the same cycle the value x was already hit in $(0, L)$. The hit $n_{1/2}(\alpha)\alpha + x$ splits I_m in exactly the same way as x has divided $(0, L)$. That means, two subintervals are generated with lengths already occurring before the division. The two new subintervals have lengths l_2 and $l_3 = l_1 - l_2$.

-*Substep 2.2*- After each complete cycle of S_α, I_m is partitioned in a symmetric way. This implies that the longest subinterval occurring in I_m is a subinterval of $\big(n_{1/2}(\alpha)\alpha, (n_{1/2}(\alpha)+1)\alpha\big)$. From this observation, combined with (2.10) and (2.18) it follows that after each complete cycle all the intervals I^p,I^{-p}, where $1 \leq p \leq n_{1/2}(\alpha)$, contain a subinterval which has the maximal length. Now note that in each cycle the order in which the elementary intervals will get hits is as follows (writing $n_{1/2}(\alpha)$ as $n^\alpha_{1/2}$ for typographical reasons):

$$\underbrace{I^1, I^{-1}, I^2, I^{-2}, \ldots, I^{n^\alpha_{1/2}}, I^{-n^\alpha_{1/2}}}_{1st\ sequence}, \underbrace{I_m, \ldots, I_m}_{2nd\ sequence}, \underbrace{I^{-n^\alpha_{1/2}}, I^{n^\alpha_{1/2}}, \ldots, I^{-2}, I^2, I^{-1}, I^1}_{3rd\ sequence}, \tag{2.22}$$

where the second sequence contains 0, 2 or 4 elements. Observe that the equivalences

$$\begin{aligned} \big(0,\alpha\big) \simeq \ldots \simeq \big((n_{1/2}(\alpha)-1)\alpha, n_{1/2}(\alpha)\alpha\big) \overset{m}{\simeq} \\ \big(1-n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)-1)\alpha\big) \simeq \ldots \simeq \big(1-\alpha, 1\big) \end{aligned} \tag{2.23}$$

hold after the first sequence and after the second sequence. At the start of the cycle, all elementary intervals in the first sequence contain a subinterval of the maximal length. It follows that Requirement 2 is satisfied if $S_\alpha(n+1)$ belongs to the first sequence.

Now let us first check the third sequence. After the third sequence of hits a cycle is completed, so then again all the intervals I^p,I^{-p}, where $1 \leq p \leq n_{1/2}(\alpha)$, contain a subinterval which has the maximal length. Just before the third sequence the maximal subinterval in each of these elementary intervals was certainly not smaller. Since (2.23) holds after the second sequence, those maximal subintervals all had the same length, which shows that Requirement 2 is satisfied if $S_\alpha(n+1)$ belongs to the third sequence.

The hits corresponding to the second sequence in (2.22) can only violate Requirement 2 if the last of these hits does so. This last hit gives a value in $\big(n_{1/2}(\alpha)\alpha, (n_{1/2}(\alpha)+1)\alpha\big)$. After the third sequence the $(k+1)$th cycle is com-

plete and hence we have the equivalence

$$\big(0,\alpha\big)(n_{k+1}(\alpha)) \simeq \big(n_{1/2}(\alpha)\alpha, (n_{1/2}(\alpha)+1)\alpha\big)(n_{k+1}(\alpha)).$$

The third sequence gives only one hit in $\big(0,\alpha\big)$. The distance between this hit and the last hit of the second sequence is $n_{1/2}(\alpha)\alpha$. It follows that the last hit of the second sequence splits an interval in two subintervals in the exactly the same way as the third sequence does in $\big(0,\alpha\big)$. By equivalences and symmetry the same holds for the other elementary intervals. Hence, Requirement 2 is also satisfied if $S_\alpha(n+1)$ belongs to the second sequence.

-*Substep 2.3*- To check Requirement 3 we use our induction hypothesis. Suppose $S_\alpha(n+1)$ is a hit in I, where I is one of the elementary intervals. If $S_\alpha(n+1)$ is an element of one of the first k cycles of S_α^I, then by our induction hypothesis it follows that $S_\alpha(n+1)$ divides an interval of length $\hat{l}_1$ in a part of length $\hat{l}_2$ and a part of length $\hat{l}_1 - \hat{l}_2$, where $\hat{l}_1 > \hat{l}_2 (> \hat{l}_3)$ are the lengths of the subintervals in I, and we are ready.

Suppose that the elementary interval I containing $S_\alpha(n+1)$ has length not larger than α. Then each cycle of S_α gives at most 2 hits in I. After $k+1$ complete cycles of S_α, we have recorded at most $2(k+1)$ values in I. After k complete cycles of S_α^I, we have at least $4k$ hits in I. Since $4k \geq 2(k+1)$, $S_\alpha(n+1)$ belongs to one of the first k cycles of S_α^I and Requirement 3 is satisfied by the induction hypothesis.

Suppose that $S_\alpha(n+1) \in I_m$ and $L > \alpha$ (this is the only case where the argument from the previous paragraph fails). Note that from the definitions it follows that $L = 2\alpha - \beta$. In the first cycle of S_α we get 2 hits in I_m. Each next cycle of S_α gives either 2 or 4 hits in I_m. After $k+1$ complete cycles of S_α, we have recorded at most $4k+2$ values in I_m. After k complete cycles of $S_\alpha^{I_m}$, we have at least $4k$ hits in I_m. It follows that if the $(k+1)$th cycle of S_α gives two hits in I_m, then $S_\alpha(n+1)$ belongs to one of the first k cycles of $S_\alpha^{I_m}$ and again we use the induction hypothesis to conclude that Requirement 3 is satisfied.

If the $(k+1)$th cycle of S_α gives four hits in I_m, denote the last two hits by x_1 and x_2. We can only have a problem when $S_\alpha(n+1)$ is equal to x_1 or x_2, since otherwise $S_\alpha(n+1)$ belongs to one of the first k cycles of $S_\alpha^{I_m}$. So we check if x_1 and x_2 split an interval according to Requirement 3. Note that x_1 and x_2 are in $\big((n_{1/2}(\alpha)+1)\alpha, 1-n_{1/2}(\alpha)\alpha\big)$ and $\big(n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)+1)\alpha\big)$ respectively. These intervals have both length $L-\alpha=\alpha-\beta$. The distance between x_2 and the next hit x in $\big(0,\alpha-\beta\big)$ is a multiple of α and by (2.18) we know that at the moment that x is reached in S_α, we have $\big(n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)+1)\alpha\big) \simeq \big(0,\alpha-\beta\big)$. Hence x_2 splits $\big(n_{1/2}(\alpha)\alpha, 1-(n_{1/2}(\alpha)+1)\alpha\big)$ in exactly the same way as x splits $\big(0,\alpha-\beta\big)$. Since x belongs to one of the first k cycles of $S_\alpha^{I^1}$, we already know that x gives the right splitting. Therefore Requirement 3 is satisfied if $x_2 = S_\alpha(n+1)$. Using symmetry we see that Requirement 3 is also satisfied when $x_1 = S_\alpha(n+1)$, which completes the proof. □

2.3 A Four Gap Theorem

We are now in position to prove our main theorem, the 'Four Gap Theorem'.

Theorem 2.2. *(The Four Gap Theorem) Let $\alpha \in \mathbb{R}\setminus\mathbb{Q}$ and $n \in \mathbb{N}$. Let $||x||$ denote the distance from x to the nearest integer. The numbers*

$$||\alpha||, ||2\alpha||, ||3\alpha||, \ldots, ||n\alpha|| \tag{2.24}$$

divide the interval $[0, \frac{1}{2}]$ in subintervals of at least two and at most four different lengths. For these lengths the following assertions hold:

1. *The rightmost length, denoted by l_r, is unique.*
2. *There are two different lengths if and only if $n||\alpha|| < \frac{1}{2}$.*
3. *If we have three different lengths, denote the two lengths not equal to l_r by $l_1 > l_2$. Then exactly one of the following four equalities holds: $2l_r = l_1$, $2l_r = l_2$,*

$2l_r + l_2 = l_1$ *or* $l_1 + l_2 = 2l_r$.[1]

4. *If we have four different lengths, denote the three lengths not equal to* l_r *by* $l_1 > l_2 > l_3$. *Then* $l_1 = l_2 + l_3$ *and one of these lengths is equal to twice* l_r.

Proof. It is not possible to have only one length occurring, since α is irrational. Without loss of generality we assume that $\alpha \in [0, 1/2]$.

Observe that $\min\{\{x\}, \{-x\}\} \in [0, \frac{1}{2}]$. So if we look at the sequence

$$\min\Big\{\{\alpha\}, \{-\alpha\}\Big\}, \min\Big\{\{2\alpha\}, \{-2\alpha\}\Big\}, \min\Big\{\{3\alpha\}, \{-3\alpha\}\Big\}, \dots \tag{2.25}$$

we get a subsequence of the sequence S_α. A term of the sequence S_α is a term of the sequence (2.25) if and only if it is in $[0, \frac{1}{2}]$. Consequently, by Theorem 2.1, the first n terms of the sequence (2.25) divide the interval $[0, \frac{1}{2}]$ in subintervals of at least two and at most four different lengths. We possibly get a fourth length because the partition of $[0, 1]$ (which gave three lengths) is now truncated at $\frac{1}{2}$. Since

$$\min\Big\{\{n\alpha\}, \{-n\alpha\}\Big\} = ||n\alpha||, \tag{2.26}$$

the numbers in (2.24) divide $[0, \frac{1}{2}]$ in subintervals of at least two and at most four different lengths.

We now turn our attention to the four assertions about the lengths. If the rightmost length is not unique, then there exist integers $0 \leq k, l, m \leq n, l \neq m$ such that

$$\frac{1}{2} - ||k\alpha|| = ||l\alpha|| - ||m\alpha||, \tag{2.27}$$

which implies that $\frac{1}{2}$ is the sum of a multiple of α and an integer, contradicting the irrationality of α. Hence, the rightmost length l_r is unique.

If $n||\alpha|| < \frac{1}{2}$, then the only lengths are $||\alpha||$ and l_r, so we have only two different lengths. For the opposite implication, assume that we have only two different lengths. The leftmost interval has length $\min_{1 \leq k \leq n} ||k\alpha||$. It follows

[1] For all four possibilities we found an example.

that the numbers $||\alpha||, \ldots, ||n\alpha||$ are all multiples of $\min_{1\leq k\leq n} ||k\alpha||$. From the irrationality of α we conclude that $\min_{1\leq k\leq n} ||k\alpha|| = ||\alpha||$ and $||n\alpha|| = n||\alpha||$, which is only possible if $n||\alpha|| < \frac{1}{2}$.

Consider the partition of $[0,1]$ by the numbers

$$\{\alpha\}, \{-\alpha\}, \{2\alpha\}, \{-2\alpha\}, \ldots, \{n\alpha\}, \{-n\alpha\}. \tag{2.28}$$

This partition is symmetric with respect to $\frac{1}{2}$. The subintervals in which $[0,1]$ is divided by these numbers, have either two or three different lengths, according to Theorem 2.1. We check what happens in both cases. If we have two different lengths and after truncating the partition of $[0,1]$ at $\frac{1}{2}$ have three different lengths, then either $2l_r = l_1$ or $2l_r = l_2$. If we have three different lengths and after truncating the partition at $\frac{1}{2}$ again have three different lengths, then either $2l_r + l_2 = l_1$ or $l_1 + l_2 = 2l_r$.

The last assertion of the Four Gap Theorem follows immediately from Theorem 2.1 and the observations made before. □

Chapter 3

Polygons in billiard orbits

3.1 Introduction

Let a billiard ball be shot from a corner of a rectangular billiard. Consider the ball as a point, and truncate the orbit somewhere at the boundary. The truncated orbit of the ball generates a partition of the rectangular billiard into polygons, similar to Figure 3.1. Many of these triangles and quadrangles seem to have the same shape and size. In this chapter we will show that (for a fixed shooting angle and stopping point) the number of different areas is at most thirteen. This universal upper bound is the sharpest possible. We also consider rational shooting angles and irrational shooting angles for which the thirteen is never reached.

There is an extensive literature on billiard problems. Some introductory material can be found in [1, 5, 19], while [2] is a recent publication.

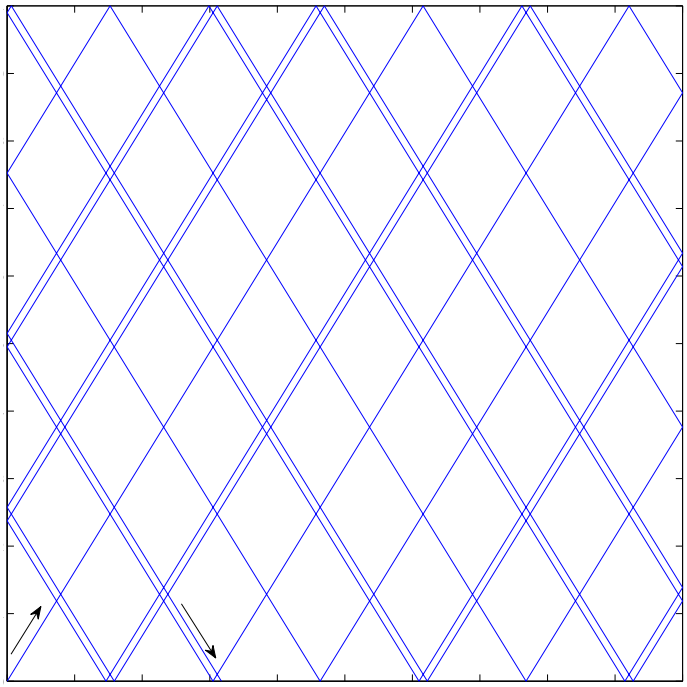

Figure 3.1: Truncated orbit of a billiard ball. The arrows indicate start and end of the orbit.

3.2 Rotations

The results in this chapter are closely related to the Three Gap Theorem (see e.g. [26], [24]) and the Four Gap Theorem (see [11]). The statements of these two theorems are best illustrated by a picture; see Figure 3.2.

The Three Gap Theorem is naturally associated to the concept of *rotations*. First we recall the theorem and then we discuss rotations on intervals. For $x \in \mathbb{R}$, let $\{x\} = x - \lfloor x \rfloor$ denote its fractional part.

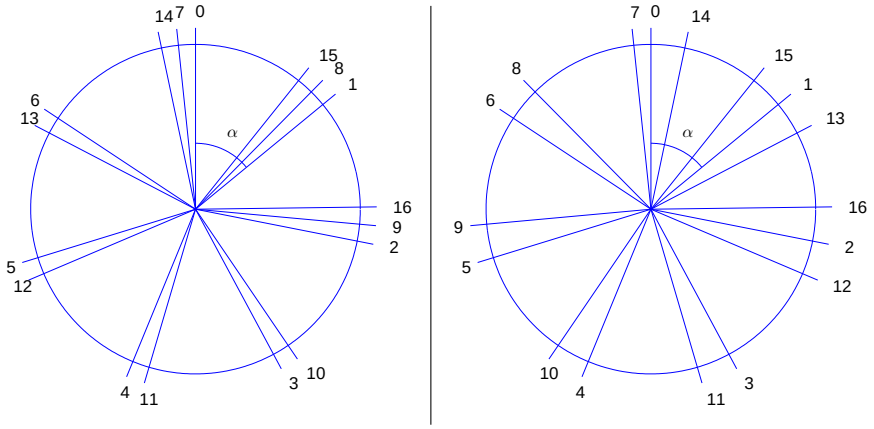

Figure 3.2: Left figure, the Three Gap Theorem: Cutting a pie n times where each next cut is obtained by shifting the previous one over a fixed angle α gives at most three different sizes of pieces of the pie. Right figure, the Four Gap Theorem: Now the first cut (at 0) works as a 'reflecting boundary'. As soon as it is reached, we continue in the opposite direction. In this case we have after n cuts at most four different sizes. For this picture we used $\alpha = 0.1405 * 2\pi$ and $n = 17$.

Theorem 3.1. *(The Three Gap Theorem) Let $n \in \mathbb{N}$ and $\alpha \in (0,1)$. The numbers*

$$0, \{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{n\alpha\} \tag{3.1}$$

induce a partition of the interval $[0,1]$ in subintervals which can have at most three different lengths. If there are three lengths, then the largest is the sum of the other two.

Letting $T_\alpha(x) = \{x + \alpha\}$ for $x \in [0,1]$, the numbers (3.1) transform into

$$0, T_\alpha(0), T_\alpha^2(0), \ldots, T_\alpha^n(0). \tag{3.2}$$

If we consider x as a point on the circle of unit circumference, then $T_\alpha(x)$ is obtained by rotating x over a distance α. This gives a more dynamical view of the partition of $[0,1]$: the partition is induced by a truncated orbit of the rotation map T_α. These observations lead to the following generalization of the Three

Gap Theorem:

Property 3.1. *Let $n_1, n_2 \in \mathbb{N}$, $\alpha \in (0,1)$ and $a, b \in \mathbb{R}$. The $n_1 + n_2 + 1$ numbers*

$$aT_\alpha^{-n_1}(0) + b, \ldots, aT_\alpha^{-1}(0) + b,\ b\ , aT_\alpha(0) + b, \ldots, aT_\alpha^{n_2}(0) + b \qquad (3.3)$$

induce a partition of $[b, b + a]$ in subintervals having at most three different lengths.

This can easily be obtained by taking $n = n_1 + n_2$ in (3.2), rotating over an appropriate angle and applying the affine map $a \cdot +b$ to the orbit. Actually, a special case of this property already appeared as a theorem in [11]. However, there a complicated proof was given to obtain this result. Vilmos Komornik came up with the idea to place the numbers on the circle, thus obtaining a much simplified and more natural argument [16]. In the sequel we will refer to (3.3) as a *rotation orbit on* $[b, b + a]$.

There is a slightly stronger property we will need in Remark 3.1 (see e.g. [11] and [24]):

Property 3.2. *Take a truncated orbit of a rotation on an interval. Suppose the orbit consists of n numbers. Create another orbit from this by removing the last number. The two partitions induced by these orbits give two sets of lengths. The union of these two sets contains at most three different lengths.*

3.3 Billiards and the Four Gap Theorem

The billiard in Figure 3.1 can be seen as a generalization to two dimensions of the pie-cutting process of the Four Gap Theorem, as illustrated in Figure 3.2. This statement deserves some explanation. Figure 3.3, a picture in some sense equivalent to the right panel of Figure 3.2, gives a description of the Four Gap Theorem in terms of a ball bouncing on the unit interval.

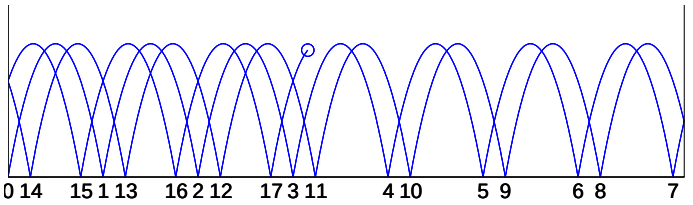

Figure 3.3: A ball bouncing on an interval between two walls. The Four Gap Theorem makes a statement about the subset of the interval consisting of the landing points of the ball. We used 0.1405 times the length of the interval as bouncing distance.

This figure shows the movement of a ball bouncing between two walls, where we assume that the ball is a point and that there is no loss of energy. The landing points of the ball build a sequence in the interval. The first n numbers in this sequence (0 included) define a splitting of the interval in n subintervals. The main statement of the Four Gap Theorem is that these subintervals can have at most four different lengths. In Figure 3.1 we now have a subset of a square, consisting of those points where the billiard ball appears. This observation gives already some reason to consider the billiard as a 2-dimensional generalization of the pie of the Four Gap Theorem. However, we can also argue this point of view in a more mathematical way.

Let $||x||$ denote the distance from x to the nearest integer. For $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, let

$$S_\alpha := (||k\alpha||)_{k=0}^{\infty}.$$

Obviously this is a sequence in $[0, \frac{1}{2}]$. Moreover, it is exactly the sequence of landing points of a ball bouncing between 0 and $\frac{1}{2}$ with horizontal bouncing distance α. The sequence S_α is obtained by 'folding' the sequence of integer multiples of α into the interval $[0, \frac{1}{2}]$. What we mean by this folding is illustrated in Figure 3.4, where we plot the function $f_1 : [0, \infty) \to [0, \frac{1}{2}]$

$$f_1(x) := ||x||,$$

and illustrate how $[0, \infty)$ is mapped to $[0, \frac{1}{2}]$ by f_1.

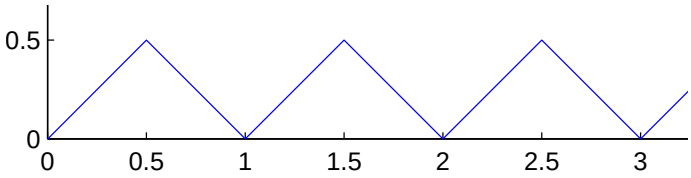

Figure 3.4: Plot of the folding map $f_1(x) = ||x||$.

Now we concentrate on the billiard: the orbit of the billiard ball is obtained by 'folding' a halfline into a rectangle. Since the shooting angle is arbitrary between 0 and $\pi/2$, we may without loss of generality assume that instead of a rectangle the billiard is a square and equal to $[0, \frac{1}{2}]^2$. The 'folding' map corresponding to this billiard is given by a two-variable function $f_2 : [0, \infty)^2 \to [0, \frac{1}{2}]^2$:

$$f_2(x, y) = (||x||, ||y||).$$

As we see, $f_2(x, y) = (f_1(x), f_1(y))$, which is why the billiard can be viewed as being a generalization of the setting of the Four Gap Theorem to two dimensions. The folding map f_2 applied to a line creates a billiard orbit. Let $\alpha > 0$, then

$$B^{\alpha}_{[0,M]} := \{(||x||, ||\alpha x||) : x \in [0, M]\}$$

describes a truncated billiard orbit that has initial slope α (the slope alternates between α and $-\alpha$). Let $\mathcal{A}^{\alpha}_{[0,M]}$ and $\mathcal{S}^{\alpha}_{[0,M]}$ denote the number of different areas respectively different shapes in the partition of $[0, \frac{1}{2}]^2$ induced by $B^{\alpha}_{[0,M]}$. Two shapes are different if one can not be obtained from the other by translating, rotating and reflecting. For orbits truncated in a boundary point, we will prove the following theorem:

Theorem 3.2. *Let $\alpha > 0$ and choose $M > 0$ such that $(||M||, ||\alpha M||) \in [0, \frac{1}{2}]^2 \setminus (0, \frac{1}{2})^2$. Then the billiard orbit $B^{\alpha}_{[0,M]}$ induces a partition of $[0, \frac{1}{2}]^2$ in polygons for*

which

$$\mathcal{A}^{\alpha}_{[0,M]} \leq 13 \quad \textit{and} \quad \mathcal{S}^{\alpha}_{[0,M]} \leq 16.$$

These upper bounds are the best possible.

In this theorem the billiard is square, but the result for rectangular billiards easily follows since the square can be scaled to any rectangle without changing the ratios between the shapes. From now on, we will assume that M satisfies the condition in the theorem.

3.4 Orbit construction

We already have an explicit expression for the billiard orbit $B^{\alpha}_{[0,M]}$, but we will need a more tractable description. Therefore, in this section we present a rough intuitive outline of the way one can think of the geometry and construction of the billiard. The corresponding lemmata and their proofs are given in Section 3.5. Consider the unit square and draw a line starting from the lower left corner with slope α. The boundaries are now considered to be connected as in a torus, so when we reach it, the line continues at the opposite boundary. Equivalently, if one of the coordinates is about to exceed 1, we subtract 1. But this is exactly taking fractional parts in both coordinates. Therefore, after we have traversed the unit square N times, we have a set which can be expressed as

$$\{(\{x\}, \{\alpha x\}) : 0 \leq x < M\},$$

for some $M \in \mathbb{R}$. A plot of such a set is shown in the left panel of Figure 3.5.

Now do the same starting from the other corners, traversing the square N times with a line either with slope α or $-\alpha$. Explicit expressions for these four sets (one for each corner) are given in Lemma 3.2. For an illustration, see the middle plot in Figure 3.5.

The key observation now is that intersection of all $4N$ lines with $[0, \frac{1}{2}]^2$ gives

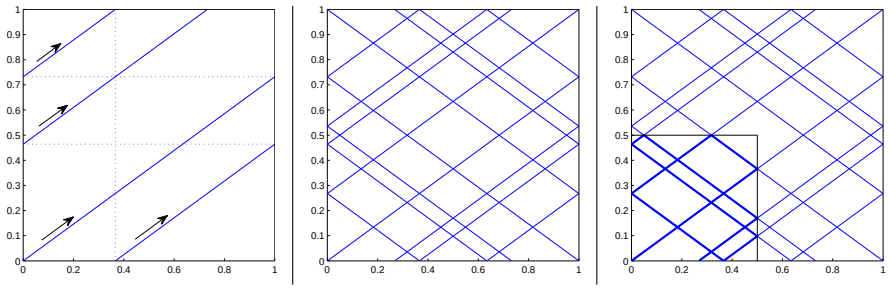

Figure 3.5: Construction of a billiard orbit in three steps. Here $N = 4$ and $\alpha = \sqrt{3} - 1$.

exactly a truncated billiard orbit with slope α, as is proved in Lemma 3.4. This fact is illustrated in the right plot in Figure 3.5. Obviously not all $4N$ lines actually contribute to the billiard orbit. However, there is a good reason to consider them all: the intercepts of the $2N$ lines with positive slope form a truncated orbit of a rotation on the interval $[-\alpha, 1]$, see Lemma 3.3. For the lines with negative slope a similar result holds. Having collected these insights, a simple counting argument suffices to obtain the upper bounds claimed in Theorem 3.2, see Section 3.6.

3.5 Lemmata and their proofs

Let $\alpha > 0$ be an irrational number and consider the halfline $l(x) = \alpha x, x \geq 0$. Let $S_1 = [0, 1)^2$ and define $S_2, S_3, S_4, \ldots$ to be the squares of the form $[k, k + 1) \times [m, m + 1)$, with k and m integers, that are consecutively traversed by the halfline, see Figure 3.6. Choosing an index N, there exists $M \in \mathbb{R}$ such that

$$\bigcup_{k=1}^{N} S_k \cap \{(x, \alpha x) : x \geq 0\} = \{(x, \alpha x) : 0 \leq x < M\}\,.$$

Taking fractional parts in both coordinates can be seen as mapping each of the

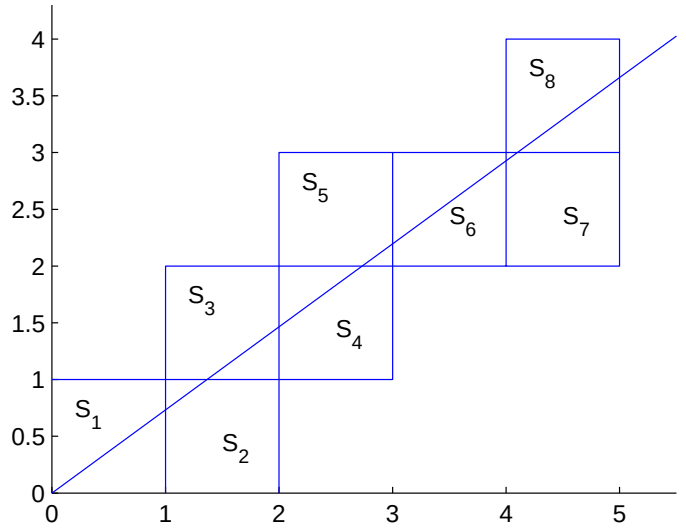

Figure 3.6: Construction of the squares S_k. Here $\alpha = \sqrt{3} - 1$, $N = 8$ and $M = 5$. The numbers y_k are approximately given by $y_0 = 0$, $y_1 \approx 0.732$, $y_2 \approx -0.268$, $y_3 \approx 0.464$,...Compare with Figure 3.5, left plot.

squares S_k to $[0,1)^2$. Therefore, doing this for the above set gives

$$\{(\{x\},\{\alpha x\}) : 0 \le x < M\} = [0,1)^2 \cap \bigcup_{k=1}^{N} \{(x, \alpha x + y_k) : x \in \mathbb{R}\} \tag{3.4}$$

for numbers y_k defined by the recursion

$$\begin{aligned} y_1 &= 0, \\ y_{k+1} &= \begin{cases} y_k + \alpha & \text{if} \quad y_k < 1 - \alpha, \\ y_k - 1 & \text{if} \quad y_k > 1 - \alpha. \end{cases} \end{aligned} \tag{3.5}$$

We will denote the set in (3.4) by A^{++}. The $++$ superscript reflects the fact that we started with a halfline in the first quadrant, so both coordinates are positive. Doing similar operations to halflines in the second, third and fourth quadrant,

we can define sets A^{-+}, A^{--} and A^{+-} respectively as follows:

$$\begin{aligned} A^{-+} &= \{(1-\{x\}, \{\alpha x\}) : 0 \le x < M\}, \\ A^{--} &= \{(1-\{x\}, 1-\{\alpha x\}) : 0 \le x < M\}, \\ A^{+-} &= \{(\{x\}, 1-\{\alpha x\}) : 0 \le x < M\}. \end{aligned}$$

Taking the union of these four sets and intersecting with $[0, \frac{1}{2}]$ gives us a billiard orbit, as is proved in the lemma below.

Lemma 3.1. *The billiard orbit $B^{\alpha}_{[0,M)}$ satisfies*

$$B^{\alpha}_{[0,M)} = \bigcup_{u,v\in\{+,-\}} A^{uv} \cap [0, \frac{1}{2}]^2.$$

Proof. Observe that

$$\begin{aligned} (||x||, ||\alpha x||) &= \Big(\min\Big\{\{x\}, 1-\{x\}\Big\}, \min\Big\{\{\alpha x\}, 1-\{\alpha x\}\Big\}\Big) \\ &= [0, \frac{1}{2}]^2 \cap \bigcup_{a\in\{\{x\},1-\{x\}\}} \bigcup_{b\in\{\{\alpha x\},1-\{\alpha x\}\}} (a, b), \end{aligned}$$

and now take the union over all $x \in [0, M)$. □

In the next lemma expressions similar to (3.4) are derived for A^{-+}, A^{--} and A^{+-}.

Lemma 3.2. *Let $y_{-k} = 1 - \alpha - y_k$ for $k = 1, 2, \ldots, N$. Then*

$$\begin{aligned} A^{-+} &= (0,1] \times [0,1) \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\}, \\ A^{--} &= (0,1]^2 \cap \bigcup_{k=1}^{N} \{(x, \alpha x + y_{-k}) : x \in \mathbb{R}\}, \\ A^{+-} &= [0,1) \times (0,1] \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_k) : x \in \mathbb{R}\}, \end{aligned}$$

Proof. Define the functions $f, g, h : \mathbb{R}^2 \to \mathbb{R}^2$ by $f((x,y)) = (1-x, y)$, $g((x,y)) = (1-x, 1-y)$ and $h((x,y)) = (x, 1-y)$. Applying these functions to the left hand side of (3.4), we get $f(A^{++}) = A^{-+}$, $g(A^{++}) = A^{--}$ and $h(A^{++}) = A^{+-}$. On the other hand,

$$\begin{aligned} f(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) &= \{(1 - x, \alpha x + y_k) : x \in \mathbb{R}\} \\ &= \{(x, \alpha(1-x) + y_k) : x \in \mathbb{R}\} \\ &= \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\}, \end{aligned}$$

whence application of f to the right hand side of (3.4) leads to

$$\begin{aligned} f\Big([0,1)^2 \cap \bigcup_{k=1}^{N} \{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) & \\ = f\Big([0,1)^2\Big) \cap f\Big(\bigcup_{k=1}^{N} \{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) & \\ = (0,1] \times [0,1) \cap \bigcup_{k=1}^{N} f\Big(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}\Big) & \\ = (0,1] \times [0,1) \cap \bigcup_{k=1}^{N} \{(x, -\alpha x + 1 - y_{-k}) : x \in \mathbb{R}\}, & \end{aligned}$$

so for A^{-+} we established the equality claimed in the lemma. The other two equalities for A^{--} and A^{+-} follow from a similar reasoning since

$$\begin{aligned} g(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) &= \{(1-x, 1-\alpha x - y_k) : x \in \mathbb{R}\} \\ &= \{(x, 1-\alpha(1-x) - y_k) : x \in \mathbb{R}\} \\ &= \{(x, \alpha x + y_{-k}) : x \in \mathbb{R}\}, \end{aligned}$$

and

$$h(\{(x, \alpha x + y_k) : x \in \mathbb{R}\}) = \{(x, 1-\alpha x - y_k) : x \in \mathbb{R}\}.$$

□

The numbers y_k and y_{-k} satisfy a nice relation, as is shown in the following lemma.

Lemma 3.3. *Let $y_0 = -\alpha$. Then the numbers $y_{-N}, \ldots, y_N$ form a rotation orbit on the interval $[-\alpha, 1]$. They are given by*

$$y_k = (1+\alpha)\left\{\frac{k\alpha}{1+\alpha}\right\} - \alpha \quad \textit{for} \quad -N \le k \le N. \tag{3.6}$$

Proof. The recursion (3.5) can be rewritten as

$$y_{k+1} = (y_k + 2\alpha \bmod (1+\alpha)) - \alpha,$$

and therefore

$$\frac{y_{k+1}+\alpha}{1+\alpha} = \frac{y_k + 2\alpha}{1+\alpha} \bmod 1 = \left\{\frac{y_k+2\alpha}{1+\alpha}\right\},$$

Letting $\tilde{y}_k = \frac{y_k+\alpha}{1+\alpha}, k = -N, \ldots, N$ and $\tilde{\alpha} = \frac{\alpha}{1+\alpha}$, this reduces to

$$\tilde{y}_{k+1} = \{\tilde{y}_k + \tilde{\alpha}\}.$$

Since $y_1 = 0$, we have $\tilde{y}_1 = \tilde{\alpha}$, which leads to

$$\tilde{y}_k = \{k\tilde{\alpha}\} \quad \text{for} \quad k \geq 1.$$

On the other hand, for $k \geq 1$,

$$\begin{aligned} \tilde{y}_{-k} &= \frac{y_{-k} + \alpha}{1+\alpha} = \frac{1 - \alpha - y_k + \alpha}{1+\alpha} = \frac{1+\alpha}{1+\alpha} - \frac{y_k + \alpha}{1+\alpha} \\ &= 1 - \tilde{y}_k = 1 - \{k\tilde{\alpha}\} = \{-k\tilde{\alpha}\}, \end{aligned}$$

since $\tilde{\alpha}$ is irrational. By definition we have $\tilde{y}_0 = 0$, and hence

$$\tilde{y}_k = \{k\tilde{\alpha}\} \quad \text{for} \quad -N \leq k \leq N.$$

Solving for y_k gives the result. □

In Lemma 3.1 we already derived an expression for $B^{\alpha}_{[0,M)}$, but this is not so easy to analyze directly. In the next lemma we describe $B^{\alpha}_{[0,M]}$ as the union of two collections of lines intersected with $[0, \frac{1}{2}]^2$. All lines in the first collection have slope α and all lines in the second collection have slope $-\alpha$.

Lemma 3.4. *Let $l^+_k(x) = \alpha x + y_k$ and $l^-_k(x) = -\alpha x + 1 - y_k$. Then*

$$B^{\alpha}_{[0,M]} = [0, \frac{1}{2}]^2 \cap \bigcup_{u \in \{+,-\}} \bigcup_{k=-N}^{N} \{(x, l^u_k(x) : x \in \mathbb{R})\}$$

Proof. This lemma will be proved by taking closures in the equation in Lemma 3.1.

$$\overline{B^{\alpha}_{[0,M)}} = B^{\alpha}_{[0,M]},$$

since $x \mapsto (||x||, ||\alpha x||)$ is a continuous function from $\mathbb{R}$ to $\mathbb{R}^2$. On the other hand,

$$\overline{[0, \frac{1}{2}]^2 \cap \bigcup_{u,v \in \{+,-\}} A^{uv}} = [0, \frac{1}{2}]^2 \cap \bigcup_{u,v \in \{+,-\}} \overline{A^{uv}},$$

and since A^{uv} is a finite collection of lines intersected by a 'half open' unit square its closure is the same collection of lines but now intersected by the closed square $[0,1]^2$. Therefore,

$$\overline{A^{++}} \cup \overline{A^{--}} = [0,1]^2 \cap \bigcup_{\substack{k=-N \\ k\neq 0}}^{N} \{(x, l_k^+(x)) : x \in \mathbb{R}\} \tag{3.7}$$

Now note that since $l_0^+(x) = \alpha x - \alpha$ we have

$$[0, \frac{1}{2}]^2 \cap \{(x, l_0^+(x)) : x \in \mathbb{R}\} = \emptyset.$$

Intersecting both sides of (3.7) with $[0, \frac{1}{2}]^2$ gives

$$[0, \frac{1}{2}]^2 \cap \left(\overline{A^{++}} \cup \overline{A^{--}}\right) = [0, \frac{1}{2}]^2 \cap \bigcup_{k=-N}^{N} \{(x, l_k^+(x)) : x \in \mathbb{R}\} \tag{3.8}$$

Analogously it follows that

$$[0, \frac{1}{2}]^2 \cap \left(\overline{A^{+-}} \cup \overline{A^{-+}}\right) = [0, \frac{1}{2}]^2 \cap \bigcup_{k=-N}^{N} \{(x, l_k^-(x)) : x \in \mathbb{R}\} \tag{3.9}$$

Combination of the last two equations gives the result. □

3.6 Proof of Theorem 3.2

Lemma 3.4 writes the billiard orbit as an intersection of the square $[0, \frac{1}{2}]^2$ with a set of lines. Let us concentrate on the lines with positive slope. By Lemma 3.3 the intercepts of these lines form a rotation orbit on the interval $[-\alpha, 1]$. So by Property 3.1 they induce a partition of this interval in subintervals of *at most*

three different lengths. Denote the set of these lengths by $D := \{d_1, \ldots, d_n\}$, where $n \leq 3$. For the lines with negative slope, the intercepts are the numbers $1 - y_k$, $-N \leq k \leq N$. They induce a partition of $[0, 1 + \alpha]$ in subintervals having lengths in the same set D. It now follows that vertical distances between adjacent parallel lines are in the set D.

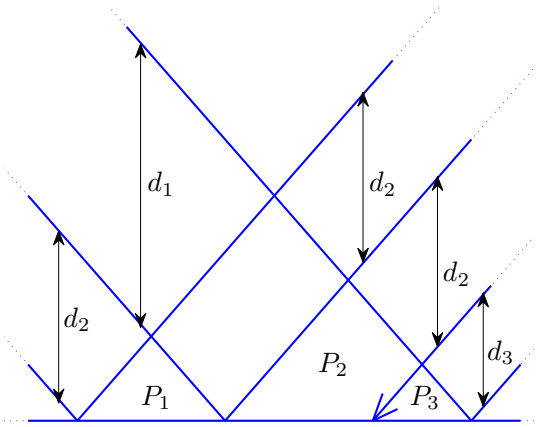

Figure 3.7: Local situation at the boundary where the orbit ends. Polygons of type 2 are triangular if the endpoint of the orbit is not one of the corners of the polygon (as is the case with P_1). There are only two shapes for which the endpoint of the orbit is one of the corners. One of them is still triangular (in this example P_3), the other is irregular (P_2).

We will distinguish between three types of polygons: those that have no side which is part of the boundary of $[0, \frac{1}{2}]^2$ (type 1), those that have exactly one such a side (type 2) and those that have two or more (type 3).

The polygons of type 1 must be parallelograms. The area of such a parallelogram is given by $d_i d_j / 2\alpha$ for some $d_i, d_j \in D$, and consequently they can have at most six different areas.

A polygon of type 2 that is triangular must be half of a rhombus of which the vertical diagonal has length $d \in D$, and therefore its area is $d^2/4a$. There is at most one non-triangular type 2 polygon, as is explained in Figure 3.7. So

polygons of type 2 can have at most four different areas.

Polygons of type 3 must be in one of the corners of $[0, \frac{1}{2}]^2$, but not in $(0, 0)$ since the orbit starts there. So this gives at most three more areas.

Putting everything together, it turns out that the number of different areas is bounded by thirteen.

For the number of shapes a similar counting argument holds. The number of parallelogram shapes is again six, since reflections do not count. The triangles that are half of a rhombus can have at most six different shapes, since there are three types of rhombi which can be cut either horizontally or vertically. The rest of the argument doesn't change, so there are at most three more different shapes than different areas, which establishes the upper bound of at most sixteen different shapes.

The sharpness of these bounds follows from Example 3.1 in section 3.7. □

Remark 3.1. As the careful reader may have noted, the construction of the billiard orbit always gives a truncation on the left boundary or on the lower boundary of the square. So strictly speaking, Theorem 3.2 is not proved in full generality yet. Suppose we have an orbit truncated at the upper or right boundary. By removing the last linear part or adding the next linear part, we can transform this orbit into an orbit truncated at the left or lower boundary. This means that in the proof above, the rotation orbit on the interval $[-\alpha, 1]$ contains one element more or one less than the rotation orbit on $[0, 1 + \alpha]$. Now Property 3.2 tells us that vertical distances between adjacent parallel lines can still have at most three different values, completing the proof.

3.7 Sharpness of the bounds

In this section we present an example in which the upper bounds of Theorem 3.2 are reached. This proves sharpness of the bounds.

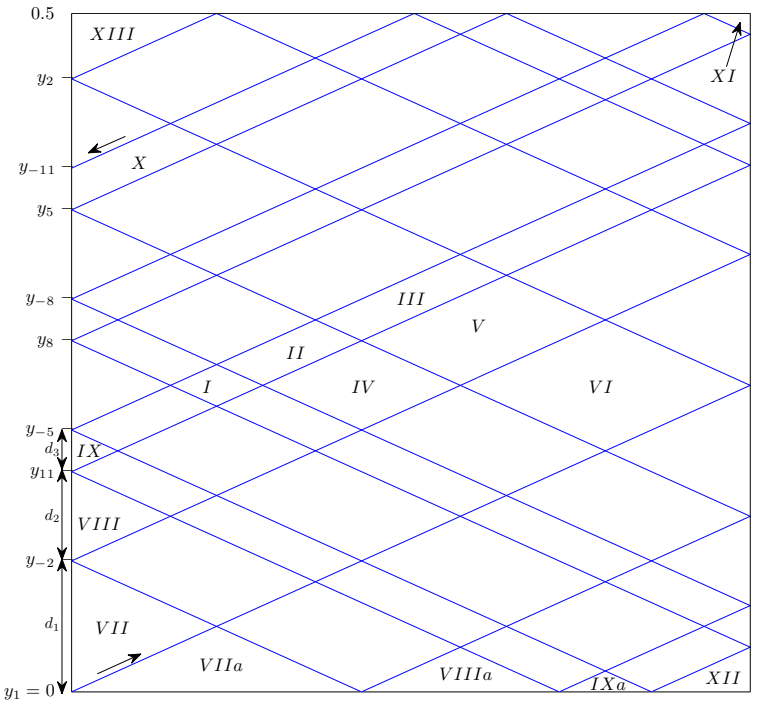

Figure 3.8: Thirteen different areas, sixteen different shapes.

Example 3.1. Let $\alpha = \frac{\sqrt{10}}{7}$ and choose $N = 11$. The corresponding orbit is shown in Figure 3.8. Use Lemma 3.3 to find the numbers y_k and let

$$d_1 = y_{-2} - y_1 \approx 0.0965, \quad d_2 = y_{11} - y_{-2} \approx 0.0658, \quad d_3 = y_{-5} - y_{11} \approx 0.0307$$

denote the three different vertical distances between adjacent parallel lines. The

areas of the shapes are of the following form:

$$\begin{array}{ll} \text{Shapes } I,\ II,\ III,\ IV,\ V,\ VI: & d_i d_j/2\alpha, \quad i \leq j \in \{1,2,3\} \\ \text{Shapes } VII,\ VIII,\ IX: & d_i^2/4\alpha, \quad i \in \{1,2,3\} \\ \text{Shape } X: & d_3 d_1/2\alpha - d_3^2/4\alpha \\ \text{Shapes } XI,\ XII,\ XIII: & d_i^2/8\alpha, \quad i \in \{1,2,3\} \end{array} \tag{3.10}$$

Calculating these thirteen areas indeed gives thirteen different values, where a precision of two decimals suffices. The flakes VII, $VIII$ and IX have the same areas as $VIIa$, $VIIIa$ and IXa respectively, so the maximal number of sixteen different shapes is also reached. We checked the calculations by using the outcomes to determine the area of $[0, \frac{1}{2}]^2$.

■

3.8 Rational angles and a golden exception

Theorem 3.2 gives an upper bound for the number of different areas of shapes on the billiard table. Some natural questions remain. For example, what happens if α is rational? Can we prove sharper upper bounds under suitable conditions? In this section we explore these properties.

Obviously, taking α rational gives a special case. The first thing to note is that the orbit will be periodic: if $\alpha = p/q$, then for $x \in \mathbb{R}$

$$(||x+q||, ||\alpha(x+q)||) = (||x||, ||\alpha x||).$$

A bit less trivial is the following result.

Proposition 3.1. *The best upper bound for $\mathcal{A}^{\alpha}_{[0,M]}$ with $\alpha \in \mathbb{Q}$ is 13, but for all $\alpha \in \mathbb{Q}$ there is an M_0 such that $1 \leq \mathcal{A}^{\alpha}_{[0,M]} \leq 3$ for $M \geq M_0$. These bounds are sharp.*

Proof. Note that the areas of the polygons continuously depend on α. So if we

have an $\tilde{\alpha}$ and M such that $\mathcal{A}^{\tilde{\alpha}}_{[0,M]} = 13$, then we can find $\varepsilon > 0$ such that the upper bound of thirteen is reached for all $\alpha \in (\tilde{\alpha} - \varepsilon, \tilde{\alpha} + \varepsilon)$. Since this interval contains rationals, we see that rationality is not sufficient for a sharper upper bound.

Since the orbit is periodic, the partition doesn't change anymore if M is large enough. Taking $\alpha = 1$ shows that 1 is a sharp lower bound for the limiting number of shapes. For the upper bound, suppose that $\alpha = p/q$. By Lemma 3.3 the intercepts satisfy

$$y_{p+q} = \left(1 + \frac{p}{q}\right)\left\{\frac{(p+q)p/q}{1+p/q}\right\} - \frac{p}{q} = -\frac{p}{q} = y_0.$$

It follows that the numbers y_k form a periodic rotation orbit on $[-\alpha, 1]$ and therefore the set D as defined in the proof of Theorem 3.2 contains only one length if M is large enough. If p and q are relative prime, then this length is $1/q$. Now a type 1 polygon is a rhombus with area $1/2pq$. Since there is no endpoint of the orbit anymore, a type 2 polygon is half of such a rhombus. Polygons in the corners are also triangular, because the orbit touches all sides of the square before becoming periodic. These triangles are quarters of the rhombus, thus having area $1/8pq$. This makes at most three different areas in total. To see that this upper bound is sharp, see Figure 3.9. □

Surprisingly, there exist irrational α for which the upper bound of thirteen different areas is never reached:

Proposition 3.2. *Let $\phi = (\sqrt{5} - 1)/2$ denote the small golden mean. If $\alpha = \frac{1}{n+\phi}$ for some $n \in \mathbb{N}$, then $\mathcal{A}^{\alpha}_{[0,M]} \leq 12$.*

Proof. Consider the numbers y_k that form a rotation orbit on $[-\alpha, 1]$. The partition of $[-\alpha, 1]$ induced by this orbit gives subintervals with lengths in a set D. This set D changes if we extend the orbit (i.e. we increase M): some lengths will disappear and new lengths will be created. In [11] and [24] it was shown that the largest length is always the first to disappear. A new length only pops

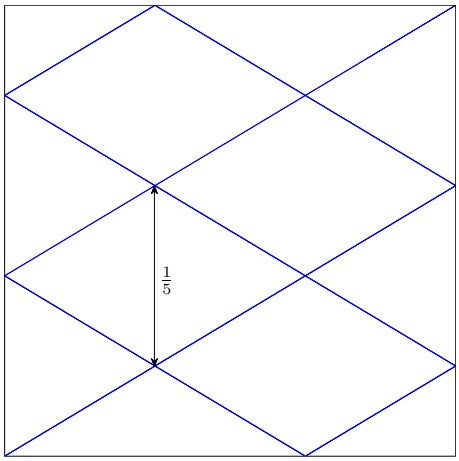

Figure 3.9: The periodic orbit for $\alpha = 3/5$. There are three different areas: the rhombi have area $1/(2 \cdot 3 \cdot 5) = 1/30$. The triangles have area $1/60$ or $1/120$.

up if there are only two lengths in D, and the new length is the difference of these two existing lengths. Together with the fact that $1 - \phi = \phi^2$, this is the basis of our argument.

Let $\alpha = 1/(n+\phi)$. From the way points are added to the rotation orbit it is clear that we can choose M such that $[-\alpha, 1]$ will be partitioned in $n+1$ intervals of length α and an interval of length $1+\alpha-(n+1)\alpha = \phi\alpha$. This gives $D = \{\alpha, \phi\alpha\}$. Extending the orbit with one more point transforms D into $\{\alpha, \phi\alpha, \phi^2\alpha\}$ and this is the first time that D contains three lengths. Increasing M further, D will change into $\{\phi\alpha, \phi^2\alpha\}$ and then into $\{\phi\alpha, \phi^2\alpha, \phi^3\alpha\}$. An inductive argument suffices to show that the ratios between the lengths in D are preserved.

Recall that the areas of the parallelograms are determined by a product of two lengths in D. By the above reasoning, if $D = \{d_1, d_2, d_3\}$, then $d_1 d_3 = d_2^2$, which implies that the parallelograms can have at most five different areas. Consequently $\mathcal{A}^{\alpha}_{[0,M]} \leq 12$. □

Chapter 4

Correlated fractal percolation and the Palis conjecture

4.1 Introduction

In this chapter we consider a natural class (called correlated fractal percolation) of random Cantor sets with dependence, as opposed to the independent case, which is know as fractal percolation or Mandelbrot percolation. Two- and three-dimensional versions of both types of sets have occurred before in the literature, especially as a modeling tool, see e.g., [27], where the dependent case is called the 'homogeneous algorithm', and the independent case the 'heterogeneous algorithm' (See Figure 4.1 Left, respectively Right for an illustration of these two processes by two realizations). In [21] they are called 'constrained curdling', respectively 'canonical curdling'. All this work has its roots in the seminal paper [20].

Our main goal is to answer the question whether or not an interval occurs

in the algebraic difference of two independent random Cantor sets from the correlated fractal percolation class. A complete answer is given in Theorem 4.3 in Section 4.5.

We also call correlated fractal percolation m out of M percolation (cf. Subsection 4.2.2), where m is an integer with $1 \leq m \leq M$. It will appear that the transition from no interval to interval lies at values of $m \approx \sqrt{M}$. The combinatorial Lemma 4.6 lies at the basis for a solution of all cases, except the case $m = \sqrt{M+1}$, which is a tough nut to crack (Lemma 4.7).

The key idea to obtain these results is that we introduce a new condition on the survival distributions, which improves on the condition given in [8]. As a bonus, this gives a more general and more simple proof of the basic theorem (Theorem 4.2). It is more simple since we do not need the combinatorial 'color lemma' of [10] and [8], nor the irreducibility condition of [8].

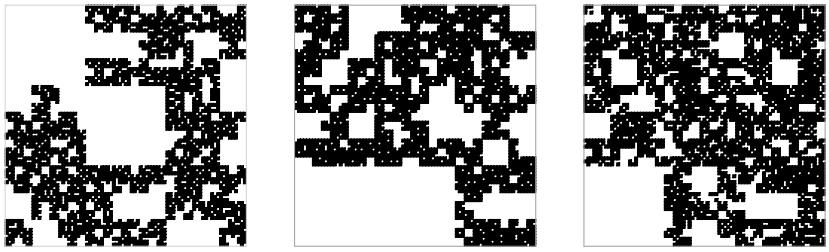

Figure 4.1: Left: Two-dimensional 7 out of 9 correlated fractal percolation with $\mu(\emptyset) = 0$. Middle: Two-dimensional 8 out of 9 correlated fractal percolation with $\mu(\emptyset) = \frac{1}{8}$. Right: Ordinary two-dimensional fractal percolation with $p = 7/9$.

4.2 Differences of random Cantor sets

Here we will introduce M-adic random Cantor sets and their differences, and the main result (Theorem 4.1) from [8] regarding the Palis conjecture, with a rough sketch of the proof. Finally we describe higher order Cantor sets which are particularly useful to obtain a more complete characterization from Theorem 4.1.

4.2.1 $M-$adic random Cantor sets

An $M-$adic random Cantor set F is constructed using the following mechanism: take the unit interval and divide it into M subintervals of equal length. Each of those subintervals corresponds to a letter in the alphabet

$$\mathbb{A} = \{0, \dots, M-1\}.$$

It will be convenient to consider $\mathbb{A}$ as an Abelian group with addition. So for instance if $M = 6$ we have $5 + 3 = 2$. Now define a *joint survival measure* μ on $2^{2^{\mathbb{A}}}$. It is determined by its values $(\mu(A))$ on the singletons $A \subset \mathbb{A}$. According to this distribution we choose which subintervals are kept and which are discarded. Then in each next construction step, each of the surviving subintervals is again divided in M subintervals of equal length, of which a subset survives according to the distribution μ.

More formally, we consider the space of $\{0,1\}$-labeled M-adic trees $\{0,1\}^{\mathcal{T}}$, where we label each node $i_1 \dots i_n \in \mathcal{T}$ with $X_{i_1 \dots i_n} \in \{0,1\}$. The root of $\mathcal{T}$ is denoted by $\emptyset$.

The probability measure $\mathbb{P}_\mu$ on this space is defined by letting $\mathbb{P}_\mu(X_\emptyset = 1) = 1$, and requiring that for all $i_1 \dots i_n \in \mathcal{T}$ the random sets

$$\left\{ i_{n+1} \in \mathbb{A} : X_{i_1 \dots i_n i_{n+1}} = 1 \right\}$$

are independent and identically distributed according to μ. We let $\mathcal{T}_n$ denote the set of nodes at level n, and for any $\underline{i}_n = i_1 \dots i_n$ from $\mathcal{T}_n$ we define the associated M-adic interval by

$$I_{i_1 \dots i_n} := \left[\frac{i_1}{M} + \dots + \frac{i_{n-1}}{M^{n-1}} + \frac{i_n}{M^n}, \frac{i_1}{M} + \dots + \frac{i_{n-1}}{M^{n-1}} + \frac{i_n + 1}{M^n} \right].$$

The n-th level approximation F^n of the random Cantor set is a union of such n-th level M-adic intervals selected by the sets S_n defined by

$$S_n = \{ i_1 \dots i_n : X_{i_1} = X_{i_1 i_2} = \dots = X_{i_1 \dots i_n} = 1 \}.$$

The random Cantor set F is

$$F = \bigcap_{n=1}^{\infty} F^n = \bigcap_{n=1}^{\infty} \bigcup_{i_1 \dots i_n \in S_n} I_{i_1 \dots i_n}.$$

The *marginal probabilities* p_i of μ are defined for $i \in \mathbb{A}$ by

$$p_i := \sum_{X \subseteq \mathbb{A} : i \in X} \mu(X). \tag{4.1}$$

We start with the definition of the class of random Cantor sets which we will take into consideration.

4.2.2 Correlated fractal percolation

From now on we will consider one-dimensional fractal percolation.

Definition 4.1. *Suppose μ assigns the* same *positive probability to all subsets of $\mathbb{A}$ with m elements for some fixed integer $1 \leq m \leq M$, and that μ assigns probability zero to all other non-empty subsets of $\mathbb{A}$. If $p := (1 - \mu(\emptyset)) \frac{m}{M}$ then we call this (m, M, p)-percolation.*

We can compute the marginal probabilities of (m, M, p)-percolation as follows. Let X be a subset of $\mathbb{A}$, chosen according to the joint survival distribution μ. The probability that X is non-empty is $1 - \mu(\emptyset)$. Given that X is non-empty, the probability that a fixed $k \in \mathbb{A}$ belongs to X equals m/M. It follows that for $k \in \mathbb{A}$ the marginal probability p_k is given by

$$p_k = (1 - \mu(\emptyset))\frac{m}{M} = p,$$

which is exactly the reason why we defined (m, M, p)-percolation by requiring that $p = (1 - \mu(\emptyset))m/M$. Because $0 \leq \mu(\emptyset) \leq 1$, (m, M, p)-percolation is only defined for $0 \leq p \leq \frac{m}{M}$. From now on we will assume that $p > 0$ and $m > 0$, since giving the empty set probability one does not yield the most exciting situation.

4.2.3 Algebraic differences of sets

The *algebraic difference* $F_1 - F_2$ of the sets F_1 and F_2 is defined by

$$F_1 - F_2 = \{x - y : \ x \in F_1, \ y \in F_2\}.$$

The well known Palis conjecture ([23]) states that 'generically' $\dim_{\mathrm{H}}(F_1) + \dim_{\mathrm{H}}(F_2) > 1$ should imply that the algebraic difference $F_1 - F_2$ will contain an interval.

This question is considered in [10] and [8] for two M-adic random Cantor sets F_1 and F_2 with the same M but not necessarily the same joint survival distribution.
One can distinguish between joint survival distributions selecting intervals independently and joint survival distributions not having this property. In the independent case, the problem is somewhat less complicated, but still far from trivial. Intervals are selected and discarded independently if and only if the

joint survival distribution satisfies for all $X \subseteq \mathbb{A}$ the equality

$$\mu(X) = \prod_{i \in X} p_i \prod_{i \notin X} (1 - p_i). \tag{4.2}$$

An important role in the answer to the main question is played by the *cyclic cross-correlation coefficients* (mostly simply called correlation coefficients)

$$\gamma_k := \sum_{i=0}^{M-1} q_i p_{i+k}, \quad \text{for } k \in \mathbb{A},$$

where (p_i) and (q_i) are the vectors of marginal probabilities of the joint survival distributions μ, respectively λ.

The result of [8] needs the following condition (which is satisfied in the independent case of Equation (4.2)).

Condition 4.1. *A joint survival distribution $(\mu(A))_{A \subseteq \mathbb{A}}$ satisfies the* joint survival condition (JSC) *if it assigns positive probability to the marginal support* $\text{Supp}_m(\mu)$ *of μ, which is defined by*

$$\text{Supp}_m(\mu) := \bigcup \{X \subseteq \mathbb{A} : \mu(X) > 0\} = \{i \in \mathbb{A} : p_i > 0\}.$$

The following result of [8] generalizes the main theorem of [10].

Theorem 4.1. *Consider two independent random Cantor sets F_1 and F_2 whose joint survival distributions μ and λ both satisfy Condition 4.1, the* (JSC).

1. *If $\gamma_k > 1$ for all $k \in \mathbb{A}$, then $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.*
2. *If $\gamma_k < 1, \gamma_{k+1} < 1$ for some $k \in \mathbb{A}$, then $F_1 - F_2$ contains no interval a.s.*

Obviously for (m, M, p)-percolation the JSC is not satisfied, unless we are in the case $m = M$, giving positive probability only to the full alphabet and the empty

set (actually, this is ordinary fractal percolation, where intervals are discarded independently and the marginal probabilities p_k are all equal to p).

4.2.4 The geometry of the algebraic difference

We will give in this subsection the tools and the notation introduced in [10] and [8].

Let $\phi : [0,1]^2 \to [-1,1]$ be given by $\phi(x,y) = x - y$, then $F_1 - F_2 = \phi(F_1 \times F_2)$. Thus $F_1 - F_2$ is defined on the product space of the probability spaces of F_1 and F_2. We will use $\mathbb{P} := \mathbb{P}_\mu \times \mathbb{P}_\lambda$ to denote the corresponding product measure and $\mathbb{E}$ to denote expectations with respect to this probability.

Let F_1 and F_2 be two independent M-adic random Cantor sets with joint survival distributions μ and λ, respectively. Denote by F_1^n and F_2^n their n^{th} level approximations ($n \geq 0$) and define the following subsets of the unit square $[0,1]^2$:

$$\Lambda^n := F_1^n \times F_2^n, \quad n \geq 0, \qquad\qquad \Lambda := F_1 \times F_2 = \bigcap_{n=0}^{\infty} \Lambda^n.$$

Note that as $F_1^n \downarrow F_1$ and $F_2^n \downarrow F_2$, also $\Lambda^n \downarrow \Lambda$.

The Λ^n are unions of M-adic squares

$$Q_{i_1 \ldots i_n, j_1 \ldots j_n} := I_{i_1 \ldots i_n} \times I_{j_1 \ldots j_n},$$

with $i_1 \ldots i_n, j_1 \ldots j_n \in \mathcal{T}_n$ and $n \geq 0$.

Note that ϕ acts as a 45° projection on the x-axis. Similarly to [10] and [8] we scale and rotate the unit square over 45° counterclockwise, to rather see it as a 90° projection on $[-1,1]$. See Figure 4.2 for a graphical representation of some of the squares Q and their ϕ-images. Here we denote the M-adic intervals $I_{i_1 \ldots i_n}$ in $[0,1]$ by $I^R_{i_1 \ldots i_n}$ (they are projections of squares in the right side of the tilted

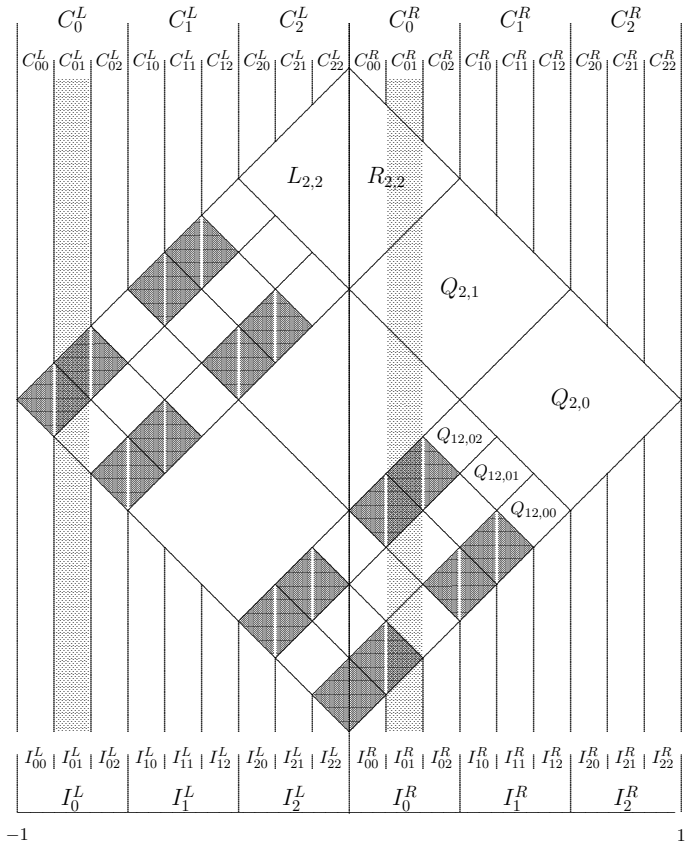

Figure 4.2: An illustration for $M = 3$ of the unit square $[0,1]^2$, scaled and rotated by 45°. The shaded squares form a realization of Λ^2 for 2 out of 3 fractal percolation. The vertical projection gives the ϕ-image $[-1,\ 5/9]$ of Λ^2.

square), and define

$$I^L_{i_1\dots i_n} = I^R_{i_1\dots i_n} - 1,$$

for the M-adic intervals $I_{i_1\dots i_n}$ in $[-1, 0]$ (they come from the left side). The

columns $C^U_{k_1\dots k_n}$, where $U = L$ or $U = R$ are defined for each $k_1 \dots k_n \in \mathcal{T}$ by

$$C^U_{k_1\dots k_n} := \phi^{-1}\left(I^U_{k_1\dots k_n}\right).$$

Note that any n^{th} level M-adic square $Q_{i_1\dots i_n, j_1\dots j_n}$ is split into a 'left' and a 'right' triangle by the M-adic columns. These triangles are called L-triangles and R-triangles, and will be denoted by $L_{i_1\dots i_n, j_1\dots j_n}$ and $R_{i_1\dots i_n, j_1\dots j_n}$ respectively, for any $i_1 \dots i_n, j_1 \dots j_n \in \mathcal{T}$.

For all $U, V \in \{L, R\}$ and $\underline{k}_n \in \mathcal{T}$ we let

$$Z^{UV}(\underline{k}_n) := \#\Big\{(\underline{i}_n, \underline{j}_n) : Q_{\underline{i}_n, \underline{j}_n} \subseteq \Lambda^n, V_{\underline{i}_n, \underline{j}_n} \subseteq C^U_{\underline{k}_n}\Big\}$$

denote the number of level n V-triangles in $\Lambda^n \cap C^U_{\underline{k}_n}$. We also denote the total number of V-triangles in columns $C^L_{\underline{k}_n}$ and $C^R_{\underline{k}_n}$ together by

$$Z^V(\underline{k}_n) := Z^{LV}(\underline{k}_n) + Z^{RV}(\underline{k}_n),$$

for all $\underline{k}_n \in \mathcal{T}$. For example, in Figure 4.2 we have $Z^R(01) = 1 + 2 = 3$.

An important observation is that an M-adic interval $I^U_{\underline{k}_n}$ is absent in $\phi(\Lambda^n)$ exactly when there are no triangles in the corresponding column $C^U_{\underline{k}_n}$ in Λ^n:

$$I^U_{\underline{k}_n} \not\subseteq \phi(\Lambda^n) \iff Z^{UL}(\underline{k}_n) = Z^{UR}(\underline{k}_n) = 0.$$

The triangle counts $Z^{UV}(\underline{k}_n)$, with $k_1, k_2, \dots$ a fixed path, constitute a two type branching process in a varying environment with interaction: the interaction comes from the dependency between triangles that are *aligned*, i.e., triangles contained in respective squares $Q_{i_1\dots i_n, j_1\dots j_n}$ and $Q_{i'_1\dots i'_n, j'_1\dots j'_n}$ with $i_1 \dots i_n = i'_1 \dots i'_n$ or $j_1 \dots j_n = j'_1 \dots j'_n$. Squares that are not aligned will be called *unaligned*.

The *expectation matrices* of the two type branching process are for $\underline{k}_n \in \mathcal{T}$ given

by:

$$\mathcal{M}\left(\underline{k}_n\right) := \begin{bmatrix} \mathbb{E}Z^{LL}(\underline{k}_n) & \mathbb{E}Z^{LR}(\underline{k}_n) \\ \mathbb{E}Z^{RL}(\underline{k}_n) & \mathbb{E}Z^{RR}(\underline{k}_n) \end{bmatrix}. \tag{4.3}$$

These matrices satisfy the basic relation

$$\mathcal{M}\left(k_1 \dots k_n\right) = \mathcal{M}\left(k_1\right) \cdots \mathcal{M}\left(k_n\right), \tag{4.4}$$

for all $k_1 \dots k_n \in \mathcal{T}$.

Lemma 4.1 shows the importance of the correlation coefficients.

Lemma 4.1. ([10]) *For all $k \in \mathbb{A}$ we have*

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \mathcal{M}\left(k\right) = \begin{bmatrix} \mathbb{E}Z^{L}(k) & \mathbb{E}Z^{R}(k) \end{bmatrix} = \begin{bmatrix} \gamma_{k+1} & \gamma_k \end{bmatrix}. \tag{4.5}$$

Proof. As in [10] this follows from some careful bookkeeping and

$$\mathbb{P}(Q_{i,j} \subseteq \Lambda^1) = \mathbb{P}(I_i \subseteq F_1^1, I_j \subseteq F_2^1) = \mathbb{P}_\mu\left(I_i \subseteq F_1^1\right)\mathbb{P}_\lambda\left(I_j \subseteq F_2^1\right) = p_i q_j.$$

□

4.2.5 Rough sketch of the proof of Theorem 4.1

The idea of the proof is to pair unaligned left and right triangles that survive in the *same* column into what are called Δ-*pairs*.

Suppose we have a Δ-pair in one of the columns with positive probability. If we can prove that there is a strictly positive probability that the number of L-triangles and R-triangles in *all subcolumns* of this column grows exponentially, then it can be shown that with positive probability the M-adic interval corresponding to this column is in the projection $\phi(\Lambda)$. The determining quantity for

exponential growth is the smallest correlation coefficient

$$\gamma := \min_{k \in \mathbb{A}} \gamma_k. \tag{4.6}$$

Now we make use of the fact that conditioned on $\Lambda \neq \emptyset$ the Hausdorff dimension of Λ is almost surely larger than 1, which is implied by $\gamma > 1$.

It can be shown (see [10]) that from this it follows that the number of unaligned squares grows to infinity. By self-similarity of the process each of the unaligned squares has positive probability to generate an interval in the projection, and hence with probability one there will be an interval in the projection.

To show that a Δ-pair occurs somewhere with positive probability it suffices that $\gamma > 1$. So the joint survival condition is only needed to ensure positive probability of exponential growth in all subcolumns of a Δ-pair. For any level l Δ-pair (L^l, R^l) that is contained in a level l column C, the distribution of the number of level $l+n$ V-triangles surviving in Λ^{l+n} in the $\underline{k}_n$-th subcolumn of (L^l, R^l), conditional on the survival of (L^l, R^l) in Λ^l, is independent of l, the particular choice of the column C and the Δ-pair in this column. Therefore, we can unambigiously denote a random variable having this distribution by

$$\tilde{Z}^V(\underline{k}_n) \tag{4.7}$$

for all $V \in \{L, R\}$ and $\underline{k}_n \in \mathcal{T}$. In general $\tilde{Z}^V(\underline{k}_n)$ does not have the distribution of $Z^V(\underline{k}_n)$ because there is possible dependence between the offspring generation of two level 0 triangles, whereas there is no dependence between the offspring generation of the L-triangle and the R-triangle of a Δ-pair, because they are unaligned by definition of a Δ-pair. However, both do have the same expected value.

In [8] the following lemma on exponential growth of triangles is proved:

Lemma 4.2. *If $\gamma > 1$, and the joint survival distributions satisfy the joint survival*

condition, then for all $n \geq 0$

$$\mathbb{P}(\tilde{Z}^L(\underline{k}_l) \geq \gamma^l, \tilde{Z}^R(\underline{k}_l) \geq \gamma^l \text{ for all } \underline{k}_l \in \mathcal{T}_l \text{ for all } 0 \leq l \leq n) > 0.$$

In Lemma 4.4 in Section 4.4 we obtain this lemma (with a different growth factor) under weaker conditions than the joint survival condition.

4.2.6 Higher order Cantor sets

The idea of higher order Cantor sets is to collapse n construction steps into one step. Since $\Lambda^n \downarrow \Lambda$ we can for all $n \geq 1$ write

$$\Lambda = \bigcap_{m=1}^{\infty} \Lambda^m = \bigcap_{m=1}^{\infty} \Lambda^{nm}.$$

The sets $(\Lambda^{nm})_{m=1}^{\infty}$ are constructed by joint survival distributions which will be denoted by $\mu^{(n)}$ and $\lambda^{(n)}$. If Theorem 4.2 fails to answer the interval or not question for the pair (μ, λ), one can hope to get an answer by considering Λ as generated by $(\mu^{(n)}, \lambda^{(n)})$.

The success of this idea is illustrated by Theorem 6.1 in [8], and by Theorem 4.4. We will also use it for the proof of Lemma 4.7.

All entities of the nth order random Cantor set will be denoted with a superscript (n). The alphabet now is $\mathbb{A}^{(n)} = \{0, \ldots, M^n - 1\}$ and $\mu^{(n)}$ and $\lambda^{(n)}$ are probability measures on the subsets of $\mathbb{A}^{(n)}$ which are completely determined by μ and λ.

Example 4.1. Let $M = 2$ and define μ by

$$\mu(\{0, 1\}) = \mu(\{1\}) = 1/2.$$

For the corresponding second order Cantor set we have $\mathbb{A}^{(2)} = \{0, 1, 2, 3\}$ and

$$\mu^{(2)}(\{0,1,2,3\}) = \mu^{(2)}(\{1,2,3\}) = \mu^{(2)}(\{0,1,3\}) = \mu^{(2)}(\{1,3\}) = \tfrac{1}{8},$$
$$\mu^{(2)}(\{2,3\}) = \mu^{(2)}(\{3\}) = \tfrac{1}{4}.$$

■

4.3 The critical case

What happens in the critical case when $\gamma = 1$? This was left open in [10] and [8]. Here we will give a simple argument, independent of the other results in this chapter, that under some conditions shows that there is almost surely no interval in the difference set. In particular this result permits us to give a complete classification in Theorem 4.3. We also can tell what happens for critical classical fractal percolation: if $p = 1/\sqrt{M}$, then there is almost surely no interval in the difference set.

Definition 4.2. *The joint survival distributions μ and λ are called* entangled *if for sets $X, Y \subseteq \mathbb{A}$ the inequality $\mu(X)\lambda(Y) > 0$ implies that $X \cap Y \neq \emptyset$.*

Proposition 4.1. *Consider two independent random Cantor sets F_1 and F_2 with joint survival distributions μ and λ having marginal probabilities (p_i) and (q_j), such that $\gamma_0 \leq 1$. Then $F_1 - F_2$ contains no interval a.s., provided that μ and λ are not entangled.*

Proof. Let Z_n be the number of 'central' squares in Λ^n, i.e.,

$$Z_n = \#\{i_1 \dots i_n \in \mathcal{T} : Q_{i_1 \dots i_n, i_1 \dots i_n} \in \Lambda^n\}.$$

Then $Z_0 = 1$, and since these central squares are unaligned, (Z_n) is an ordinary branching process with mean offspring

$$\mathbb{E}[Z_1] = p_0 q_0 + p_1 q_1 + \dots + p_{M-1} q_{M-1} = \gamma_0 \leq 1.$$

Now if $\gamma_0 = 1$, then the offspring distribution is deterministic ($Z_1 \equiv 1$) if and only if μ and λ are entangled ($\mathbb{P}(Z_1 = 0) \geq \mu(X)\lambda(Y) > 0$ if X and Y are the sets with $X \cap Y = \emptyset$ and $\mu(X)\lambda(Y) > 0$). Hence, (Z_n) will die out a.s., say at time N. In the sequel we will write the string $i_1 \dots i_n = (k, k, \dots, k)$ for $k \in \mathbb{A}$ as k^n.

Then, because there are no central squares left, $C^R_{0^{N+n}}$ only contains left triangles for all $n \geq 0$. Moreover, the number of left triangles in $(C^R_{0^{N+n}})$ is an ordinary branching process (Y^R_n) with random initial distribution Y^R_0, and mean offspring

$$\mathbb{E}[Y^R_1] = p_0 q_{M-1} \leq 1.$$

Similarly, $C^L_{(M-1)^{N+n}}$ only contains right triangles for all $n \geq 0$. Moreover, the number of right triangles in $(C^L_{(M-1)^{N+n}})$ is a branching process (Y^L_n) with Y^L_0, and mean offspring

$$\mathbb{E}[Y^L_1] = p_{M-1} q_0 \leq 1.$$

If both $\mathbb{E}[Y^R_1]$ and $\mathbb{E}[Y^L_1]$ would equal 1, then $p_0 q_{M-1} = p_{M-1} q_0 = 1$ and consequently $p_0 q_0 = p_{M-1} q_{M-1} = 1$ implying that $\gamma_0 \geq 2$. Hence either $\mathbb{E}[Y^R_1] < 1$ or $\mathbb{E}[Y^L_1] < 1$, such that at least one of the two branching processes (Y^R_n) and (Y^L_n) will die out almost surely, implying that $F_1 - F_2$ has a 'gap' directly left or right of 0. It then follows from selfsimilarity and the denseness of the points $k_1 M^{-1} + \dots + k_n M^{-n}$ that $F_1 - F_2$ contains no interval a.s. (cf. [10]) □

That we need at least some restriction on the joint survival distributions in addition to the requirement $\gamma_0 \leq 1$ is shown in the following example: Let $M = 2$ and define the joint survival distributions μ and λ by setting $\mu(\{0\}) = \mu(\{1\}) = 1/2$ and $\lambda(\{0, 1\}) = 1$. Then $\gamma = \gamma_0 = 1$, and there is an interval of length 1 in the difference set at a random position. Proposition 4.1 does not hold since μ and λ are entangled.

4.4 The distributed growth condition

In this section we introduce a condition for exponential growth of triangles which is based on the following idea: if we can find a column C where we have a sufficient number of Δ-pairs, then under some conditions each of these Δ-pairs can be used to guarantee exponential growth of triangles in a proper subset of the set of subcolumns of C. In some sense we 'spread the burden of proof', and this gives the condition a flexible nature. This is illustrated by the fact that with help of this condition, we can completely classify correlated fractal percolation.

For $X, Y \subseteq \mathbb{A}$ and $e \in \mathbb{A}$ we define $\gamma_e(X, Y)$ to be the e^{th} correlation coefficient corresponding to the joint survival distributions $\mu^\star$ and $\lambda^\star$ assigning probability one to X and Y respectively, i.e.,

$$\gamma_e(X, Y) = \sum_{i \in \mathbb{A}} \mathbf{1}_Y(i)\mathbf{1}_X(i + e). \tag{4.8}$$

Condition 4.2. *The pair of joint survival distributions (μ, λ) satisfies the* distributed growth condition (DGC) *if for all $k \in \mathbb{A}$ we can find sets $X_k, Y_k \subseteq \mathbb{A}$ such that*

(DG0) $\mu(X_k) > 0$ *and* $\lambda(Y_k) > 0$,

(DG1) $\min_{e \in \mathbb{A}} \gamma_e(X_k, Y_k) \geq 1$,

(DG2) $\gamma_k(X_k, Y_k) \geq 2,\ \gamma_{k+1}(X_k, Y_k) \geq 2$.

Lemma 4.3. *Let E denote the event that there exists $l \geq 1$, $\underline{k}_l \in \mathcal{T}_l$ and $U \in \{L, R\}$ such that $C^U_{\underline{k}_l}$ contains at least M left and M right triangles which are all pairwise unaligned. If the pair of joint survival distributions (μ, λ) satisfies the DGC, then*

$$\mathbb{P}(E) > 0.$$

Proof. Choose $X_0, Y_0 \subseteq \mathbb{A}$ according to the DGC. Define the joint survival dis-

tributions $\mu^\star$ and $\lambda^\star$ by $\mu^\star(X_0) = \lambda^\star(Y_0) = 1$. Then by (DG2) both column sums of the expectation matrix $\mathcal{M}^\star(0)$ are at least 2, implying that

$$[1 \quad 1]\,\mathcal{M}^\star(0^n) \geq [2^n \quad 2^n],$$

elementwise. The first row of $\mathcal{M}^\star(0^n)$ corresponds to $C^L_{0^n}$, which can contain at most one left triangle and no right triangles. Therefore, both numbers in the second row of $\mathcal{M}^\star(0^n)$ are bounded below by $2^n - 1$. It follows that the numbers of left and right triangles in $C^R_{0^n}$ grow arbitrary large if n is sufficiently large. Since μ and λ assign positive probability to X_0 and Y_0 respectively, the statement of the lemma follows. □

We can now formulate our exponential growth lemma.

Lemma 4.4. *If the pair of joint survival distributions (μ, λ) satisfies the distributed growth condition, then there exist $l \geq 1$, $\underline{k}_l \in \mathcal{T}_l$ and $\eta > 1$ such that for all $n \geq 0$*

$$\mathbb{P}(Z^L(\underline{k}_l\underline{k}_p) \geq \eta^p, Z^R(\underline{k}_l\underline{k}_p) \geq \eta^p \text{ for all } \underline{k}_p \in \mathcal{T}_p \text{ for all } 0 \leq p \leq n) > 0.$$

Proof. Choose $n \geq 0$ arbitrary. For all $k \in \mathbb{A}$ choose $X_k \subseteq \mathbb{A}$ and $Y_k \subseteq \mathbb{A}$ such that these sets satisfy the DGC. Define the joint survival distributions $\mu^\star_k$ and $\lambda^\star_k$ by requiring that $\mu^\star_k(X_k) = \lambda^\star_k(Y_k) = 1$.

Let $k \in \mathbb{A}$ be fixed and consider the expectation matrices corresponding to the triangle growth process defined by $(\mu^\star_k,\lambda^\star_k)$. By (4.5), their column sums are given by the correlation coefficients corresponding to the pair of joint survival distributions $(\mu^\star_k, \lambda^\star_k)$. So, for all $e \in \mathbb{A}$, both column sums of $\mathcal{M}^\star_k(e)$ are at least 1 and both column sums of $\mathcal{M}^\star_k(k)$ are at least 2. Let p be an integer with $0 \leq p \leq n$. Since for $\underline{k}_p = k_1 \dots k_p \in \mathcal{T}_p$ we have

$$\mathcal{M}^\star_k(\underline{k}_p) = \mathcal{M}^\star_k(k_1) \dots \mathcal{M}^\star_k(k_p),$$

it follows that a lower bound for the column sums of $\mathcal{M}^\star_k(\underline{k}_p)$ is determined by

the number of k's in the string $\underline{k}_p$. We obtain (omitting the dependence on k, and writing k_j for the jth element in the string $\underline{k}_p$.)

$$\gamma^\star_{\underline{k}_p} \geq 2^{\#\{0\leq j\leq p : k_j = k\}}, \quad \gamma^\star_{\underline{k}_p+1} \geq 2^{\#\{0\leq j\leq p : k_j = k\}}.$$

From the deterministic nature of $\mu^\star_k$ and $\lambda^\star_k$, it follows that the expectation of the number of triangles in some column is simply the number that will occur. This means that for all $0 \leq p \leq n$

$$\begin{aligned} Z^{L;\star}_k(\underline{k}_p) &= \mathbb{E}[Z^{L;\star}_k(\underline{k}_p)] = \gamma^\star_{\underline{k}_p+1} \geq 2^{\#\{0\leq j\leq p : k_j = k\}}, \\ Z^{R;\star}_k(\underline{k}_p) &= \mathbb{E}[Z^{R;\star}_k(\underline{k}_p)] = \gamma^\star_{\underline{k}_p} \geq 2^{\#\{0\leq j\leq p : k_j = k\}}. \end{aligned}$$

Since (μ, λ) satisfies the DGC, we can by Lemma 4.3 find an l-adic column $C^U_{\underline{k}_l}$ containing with strictly positive probability at least M left- and M right triangles being all pairwise unaligned. Let this event be denoted by E and abbreviate the notation of this column by C and its subcolumns $C^U_{\underline{k}_l \underline{k}_p}$ by $C_{\underline{k}_p}$.

Now suppose we have a Δ-pair (L, R) in C, in which the growth process behaves according to the pair of joint survival distributions $(\mu^\star_k, \lambda^\star_k)$. Then, for all p and all subcolumns $C_{\underline{k}_p}$ of C, both the number of left and the number of right triangles in $C_{\underline{k}_p} \cap (L \cup R)$ is at least $2^{\#\{0\leq j\leq p : k_j = k\}}$.

Conditional on the event E, we have M left and right triangles in C. We can label them by the elements of $\mathbb{A}$ such that we have M Δ-pairs. These $2M$ triangles are all pairwise unaligned (also if they belong to different Δ-pairs) and hence there is completely no dependence between these triangles. It follows that it is possible that in each of the Δ-pairs the growth process takes place as prescribed by $\mu^\star_k$ and $\lambda^\star_k$, where k is the label of the Δ-pair. Denoting the event that this happens in the first n construction steps after occurrence of E by E_n, we can find a strictly positive lower bound for $\mathbb{P}(E_n|E)$:

$$\mathbb{P}(E_n|E) \geq \prod_{k\in\mathbb{A}} \mu(X_k)^{\sum_{j=1}^n 2(\#X_k)^{j-1}} \lambda(Y_k)^{\sum_{j=1}^n 2(\#Y_k)^{j-1}} > 0.$$

Let $0 \leq p \leq n$ and let $C_{\underline{k}_p}$ be an arbitrary M^p-adic subcolumn of C. There must exist a $k = k(\underline{k}_p) \in \mathbb{A}$ such that $\#\{0 \leq j \leq p : k_j = k\} \geq \lceil \frac{p}{M} \rceil$. Hence, given the event E_n, for the numbers of left and right triangles in $C_{\underline{k}_p}$ we have

$$Z^L(\underline{k}_l \underline{k}_p) \geq 2^{\lceil \frac{p}{M} \rceil}, \quad Z^R(\underline{k}_l \underline{k}_p) \geq 2^{\lceil \frac{p}{M} \rceil}.$$

Taking $\eta = \sqrt[M]{2}$, we obtain

$$\begin{aligned} &\mathbb{P}(Z^L(\underline{k}_l \underline{k}_p) \geq \eta^p, Z^R(\underline{k}_l \underline{k}_p) \geq \eta^p \text{ for all } \underline{k}_p \in \mathcal{T}_p \text{ for all } 0 \leq p \leq n) \\ &\geq \mathbb{P}(E)\mathbb{P}(E_n|E) > 0. \end{aligned}$$

□

Collecting the results established so far, we can replace the joint survival condition (Condition 4.1) and Lemma 4.2 by the distributed growth condition and Lemma 4.4 to obtain the following useful variation on Theorem 4.1:

Theorem 4.2. *Consider two independent random Cantor sets F_1 and F_2 whose joint survival distributions satisfy Condition 4.2, the* DGC.

1. *If $\gamma_k > 1$ for all $k \in \mathbb{A}$, then $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.*
2. *If $\gamma_k < 1, \gamma_{k+1} < 1$ for some $k \in \mathbb{A}$, then $F_1 - F_2$ contains no interval a.s.*

This result is useful since it can be successfully applied to the class of correlated fractal percolation, whilst the JSC is never satisfied for the members of this class. Actually our new condition can always supersede the JSC.

Lemma 4.5. *Suppose that the joint survival distributions μ and λ satisfy the* JSC. *If $\gamma_k > 1$ for all $k \in \mathbb{A}$, then the pair (μ, λ) satisfies the* DGC.

Proof. We take for the sets X_k and Y_k in (4.8) the marginal supports of μ and λ. Then the JSC implies that (DG0) holds. Since $q_i = 0$ if $i \notin \mathrm{Supp}_m(\lambda)$, and

similarly for p_i, we have for all $e \in \mathbb{A}$

$$\begin{aligned}\gamma_e(\mathrm{Supp}_m(\mu), \mathrm{Supp}_m(\lambda)) &= \sum_{i\in\mathbb{A}} \mathbf{1}_{\mathrm{Supp}_m(\lambda)}(i)\mathbf{1}_{\mathrm{Supp}_m(\mu)}(i+e) \\ &\geq \sum_{i\in\mathbb{A}} q_i p_{i+e} = \gamma_e \geq 2,\end{aligned}$$

since the number on the left hand side is an integer larger than 1. Therefore X_k and Y_k certainly satisfy (DG1) and (DG2) for all $k \in \mathbb{A}$. Thus (μ, λ) satisfies the DGC. □

4.5 Classifying correlated fractal percolation

With the distributed growth condition at our disposal we can make an attempt to solve the Palis problem for correlated fractal percolation. To facilitate our search for sets satisfying the DGC, we introduce an alternative notation for subsets of the alphabet. A subset S of the alphabet $\mathbb{A}$ can be represented as a string of length M with at the ith position a zero or a one, indicating whether or not i is contained in S. For (m, M, p)-percolation, all subsets of $\mathbb{A}$ to which is assigned positive probability correspond to a string consisting of m ones and $M - m$ zeros, where any order of the symbols is allowed. Next we need the notion of the *cyclic shift operator* σ. For any string $X = x_0x_1 \ldots x_{M-2}x_{M-1}$ we define

$$\sigma(X) = x_1x_2 \ldots x_{M-1}x_0. \tag{4.9}$$

For the kth iterate of σ we use the notation σ^k and for its inverse σ^{-k}. Computing $\gamma_k(X, Y)$ can be done by writing down the two binary strings corresponding to $\sigma^k(X)$ and Y, and then counting in how many positions both strings have a one (this will be called a *coincidence*). This procedure is illustrated in (4.10) for $M = 9$, $k = 4$ and the sets $X = \{3, 5, 7, 8\}$ and $Y = \{0, 1, 6, 7\}$, where we abuse notation by *also* writing X for the indicator string of X, and similarly

for Y (this will never cause confusion).

$$\begin{array}{lcccccccccc} X & : & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ \sigma^4(X) & : & 0 & \mathbf{1} & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ Y & : & 1 & \mathbf{1} & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array} \tag{4.10}$$

As we see, there is one coincidence, so $\gamma_4(X, Y) = 1$. Checking the DGC boils down to finding binary strings with the right properties as given in (DG0), (DG1) and (DG2).

Let X and Y be two subsets of the M-adic alphabet $\mathbb{A}$ containing m elements in order to satisfy (DG0). Our strategy is to choose X such that we get a binary string with all ones at the beginning and Y such that the ones are distributed evenly over the string in such a way that at most $m-1$ consecutive zeros occur. This pattern will lead to fulfillment of requirement (DG1). If we have sufficient freedom to choose Y within this framework, then we will also succeed in letting (DG2) be satisfied. The details of this strategy are filled in in the proof of the lemma below.

Lemma 4.6. *For (m, M, p)-percolation the following two assertions hold:*

1. *If $m < \sqrt{M}$ or $p < \dfrac{1}{\sqrt{M}}$, then $F_1 - F_2$ contains no interval a.s.*[1]
2. *If $m \geq \sqrt{M+2}$ and $p > \dfrac{1}{\sqrt{M}}$, then $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.*

Proof. Suppose that $p < \frac{1}{\sqrt{M}}$, then for all $k \in \mathbb{A}$ we have

$$\gamma_k = Mp^2 < M\left(\frac{1}{\sqrt{M}}\right)^2 = 1,$$

[1] Actually, $m < \sqrt{M}$ implies that $p < 1/\sqrt{M}$. Hence the statement *"If $p < 1/\sqrt{M}$, then $F_1 - F_2$ contains no interval a.s."* is equivalent to the first assertion of Lemma 4.6. We formulated the lemma in this way to emphasize what the bounds on m are.

and consequently $F_1 - F_2$ contains no interval a.s. by Theorem 4.2. If $m < \sqrt{M}$, then $p = (1-\mu(\emptyset))\frac{m}{M} < \frac{1}{\sqrt{M}}$ and consequently the same argument is applicable, completing the proof of the first part of Lemma 4.6.

For the proof of the second assertion, assume that $m \geq \sqrt{M+2}$ and define $X, Y' \subseteq \mathbb{A}$ by their strings

$$\begin{aligned} X &= 1^m\, 0^{M-m} \\ Y' &= R\,[1\,0^{m-1}]^q, \end{aligned}$$

where $q = \lfloor M/m \rfloor$,R is a left substring of $1\,0^{m-2}$ (R is empty when m divides M), and $[1\,0^{m-1}]^q$ denotes the string $1\,0^{m-1}$, q times repeated. Ignoring the trivial case $M = m = 2$ we obtain from $m \geq \sqrt{M+2}$ that we may assume $m \geq 3$.

Since Y' does not contain m consecutive zeros (also cyclically), whereas X begins with m consecutive 1's, we must have

$$\gamma_e(X, Y') \geq 1 \qquad \text{for } e = 0, 1, \ldots, M-1.$$

So X and Y' satisfy (DG1). The set X contains m elements, which means that $\mu(X) > 0$.

Note that $q = \lfloor M/m \rfloor$ can not exceed $m-1$, since that would imply $m \leq \sqrt{M}$.

Case 1: $q \leq m-2$ or R is empty.
Then Y' contains at most $m-1$ ones. In order to obtain (DG2), we construct Y'' from Y' by putting a one in the second position (if there is a zero)—note that X and Y'' will then certainly still satisfy (DG1). Moreover, we now have

$$\gamma_0(X, Y'') \geq 2, \qquad \gamma_1(X, Y'') \geq 2,$$

since $m \geq 3$. Finally Y is obtained by adding 1's to Y'' (if necessary) till Y contains m ones—and thus $\mu(Y) > 0$. As an illustration for $M = 7$ and $m = 4$,

X is given by 1111000 and (writing $\gamma_k(\cdot)$ for $\gamma_k(X,\cdot)$):

$\cdot$	String	$\mu(\cdot)>0$	$\gamma_e(\cdot)\geq 1 \quad \forall e\in\mathbb{A}$	$\gamma_0(\cdot),\gamma_1(\cdot)\geq 2$
Y' :	1001000	no	yes	no
Y'' :	1101000	no	yes	yes
Y :	1111000	yes	yes	yes

Now we have found $X_0 := X$ and $Y_0 := Y$ satisfying (DG0), (DG1) and (DG2) for $k = 0$. By observing that

$$\gamma_k(X,\sigma^k Y) = \gamma_0(X,Y); \quad \gamma_{k+1}(X,\sigma^k Y) = \gamma_1(X,Y), \tag{4.11}$$

it follows that the DGC holds for any $k \in \mathbb{A}$ if we take $X_k = X$ en $Y_k = \sigma^k Y$.

Case 2: $q = m-1$ and $R \neq \emptyset$.
Since $m \geq \sqrt{M+2}$, we have (with r the length of R)

$$M = m^2 - m + r \geq M + 2 - m + r,$$

so $r \leq m-2$, implying that R does not contain more than $m-3$ zero's. This gives that $\gamma_0(X,Y) \geq 2$ and $\gamma_1(X,Y) \geq 2$. Now again we can take $X_k = X$ en $Y_k = \sigma^k Y$. Summarizing, for all cases of correlated fractal percolation in part (2) we have shown that (DG0), (DG1) and (DG2) hold. We conclude that the DGC is satisfied.

Moreover, for all $k \in \mathbb{A}$ we find

$$\gamma_k = \sum_{j=0}^{M-1} p_j p_{j+k} = Mp^2 > M\left(\frac{1}{\sqrt{M}}\right)^2 = 1,$$

and therefore, by Theorem 4.2, $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$. □

Lemma 4.6 still gives no conclusive answer for some combinations of m and M

when $p > 1/\sqrt{M}$, namely, those where $m = \sqrt{M+1}$. By having a look at the 2^{nd} order sets for (m, M, p)-percolation this can be resolved.

Lemma 4.7. *Consider (m, M, p)-percolation. If $p > \dfrac{1}{\sqrt{M}}$ and*

$$m = \sqrt{M+1}, \tag{4.12}$$

then $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.

Proof. First we have a look at the shape of the binary strings corresponding to 2^{nd} order sets to which is assigned positive probability by the 2^{nd} order joint survival distribution $\mu^{(2)}$ of correlated fractal (m, M, p)-percolation. Such a string has length M^2. It should be regarded as consisting of M blocks of length M. Each of these blocks contains either exclusively zeros, or it contains $M - m$ zeros and m ones. Blocks of the latter kind occur exactly m times. Positions in the binary string can be identified with numbers in $\mathbb{A}^{(2)}$: an M^2-adic number represented by $\underline{k}_2 = k_1 k_2$ corresponds to the (k_2+1)th position in the (k_1+1)th block.

Note that (4.12) implies that $M - m(m-1) = m - 1$ and $\lfloor M/m \rfloor = m - 1$. This means that the two strings X and Y' defined in the proof of Lemma 4.6 are now equal to (we omit from now on the prime on Y)

$$\begin{aligned} X &= 1^m\, 0^{M-m}, \\ Y &= [1\, 0^{m-2}]\, [1\, 0^{m-1}]^{m-1}. \end{aligned}$$

The basic idea of the proof is to replace the 0's in these two strings by blocks 0^M, and the 1's by blocks similar to X or Y to obtain for all $\underline{k}_2 \in \mathbb{A}^{(2)}$ the order 2 strings $X^{(2)}_{\underline{k}_2}$ and $Y^{(2)}_{\underline{k}_2}$ which will satisfy (DG1) and (DG2)—note that by construction (DG0) is then obviously satisfied.

Actually we will replace all the m 1's in X by the string Y. Replacing additionally the $M - m$ 0's by blocks 0^M we obtain $X^{(2)}_{\underline{k}_2}$ independent of $\underline{k}_2$, and hence

we will denote it by $X^{(2)}$.

The definition of $Y^{(2)}_{\underline{k}_2}$ is slightly more involved. We first restrict ourselves to the case $k_1 = 0$ and define:

$$Y^{(2)}_{0k_2} := \sigma^{Ms}\left([\sigma^{k_2}(X)\,0^{(m-2)M}]\,[\sigma^{k_2}(X)\,0^{(m-1)M}]^{m-1}\right),$$

where s is given by

$$s := \begin{cases} 0 & \text{if } 0 \le k_2 \le m-2, \\ 1 & \text{if } m-1 \le k_2 \le M-1. \end{cases}$$

So the m 1's in Y are replaced by shifted versions of X and 0's by blocks 0^M and finally an additional shift over M positions is applied on the complete string if k_2 is at least $m-1$.

Example 4.2. Let $M = 8$ and $m = 3$. Then

$$X = 11100000 \quad \text{and} \quad Y = 10100100.$$

Writing $O = 0^8$ and $s = \mathbf{1}_{\{n:n\ge 2\}}(k_2)$, we have for $0k_2 \in \mathbb{A}^{(2)}$

$$\begin{array}{rcccccccccc} X^{(2)} & = & & Y & Y & Y & O & O & O & O & O \\ Y^{(2)}_{0k_2} & = & \sigma^{8s}\big(\sigma^{k_2}(X) & O & \sigma^{k_2}(X) & O & O & \sigma^{k_2}(X) & O & O\big). \end{array}$$

■

Suppose that $X^{(2)}$ and $Y^{(2)}_{0k_2}$ satisfy the DGC. Then it is easy to construct sets $X^{(2)}$ and $Y^{(2)}_{k_1k_2}$ satisfying requirements (DG1) and (DG2) for other values of k_1. First observe that all shifted versions of $X^{(2)}$ and $Y^{(2)}_{0k_2}$ still satisfy (DG1).

Furthermore we use the fact that

$$\begin{aligned}\gamma^{(2)}_{\underline{k}_2}\big(X^{(2)},\sigma^{k_1M}(Y^{(2)}_{0k_2})\big) &= \gamma^{(2)}_{0k_2}(X^{(2)},Y^{(2)}_{0k_2})\geq 2,\\ \gamma^{(2)}_{\underline{k}_2+1}\big(X^{(2)},\sigma^{k_1M}(Y^{(2)}_{0k_2})\big) &= \gamma^{(2)}_{(0k_2)+1}(X^{(2)},Y^{(2)}_{0k_2})\geq 2.\end{aligned}$$

Now it follows that we can choose $Y^{(2)}_{\underline{k}_2}=\sigma^{k_1M}(Y^{(2)}_{0k_2})$.

To complete the proof, it suffices to check that the sets $X^{(2)}$ and $Y^{(2)}_{0k_2}$ satisfy requirements (DG1) and (DG2) of the DGC. Therefore, we consider the correlation coefficients $\gamma^{(2)}_{\underline{e}_2}(X^{(2)},Y^{(2)}_{0k_2})$ where $\underline{e}_2=e_1e_2\in\mathbb{A}^{(2)}$. We will focus first on the 'coarse' structure, i.e. on those correlation coefficients for which $e_2=0$. Here we will always have a string $\sigma^{k_2}(X)$ in $Y^{(2)}_{0k_2}$ coinciding with a string Y in $X^{(2)}$ for the same reason that we always have a coincidence at level 1. This implies that we also always have a string $\sigma^{k_2}(X)$ in $Y^{(2)}_{0k_2}$ coinciding with a zero string of length M in $X^{(2)}$ which is followed (cyclically) by a string Y.

It follows that if we will shift on the 'fine' level by varying e_2, then in all cases we are in the same situation of one $\sigma^{k_2}(X)$ block 'entering' an Y block, and one $\sigma^{k_2}(X)$ 'leaving' an Y block. Thus we get the same coincidences as in the case where $\sigma^{k_2}(X)$ and Y are compared cyclically, and therefore the second order correlation coefficients can be related to the first order correlation coefficients $\gamma_e(\sigma^{k_2}(X),Y)$:

$$\gamma^{(2)}_{\underline{e}_2}(X^{(2)},Y^{(2)}_{0k_2})\geq\gamma_{e_2}(Y,\sigma^{k_2}(X))\geq 1 \tag{4.13}$$

for all $\underline{e}_2=e_1e_2\in\mathbb{A}^{(2)}$. As we see, (DG1) holds for all $\underline{e}_2\in\mathbb{A}^{(2)}$.

Now we turn to (DG2). If $e_2=k_2$, then in (4.13) we even have by equation (4.11) that

$$\gamma^{(2)}_{\underline{e}_2}(X^{(2)},Y^{(2)}_{0k_2})\geq\gamma_{e_2}(Y,\sigma^{k_2}(X))=\gamma_0(Y,X)=2,$$

which means that

$$\gamma^{(2)}_{0k_2}(X^{(2)},Y^{(2)}_{0k_2})\geq 2. \tag{4.14}$$

We still have to check that also $\gamma^{(2)}_{(0k_2)+1}(X^{(2)},Y^{(2)}_{0k_2})\geq 2$. First we concentrate

on the case where both the first and the last Y-block in $X^{(2)}$ coincide with a $\sigma^{k_2}(X)$ block in $Y^{(2)}_{0k_2}$. To illustrate this in the terms of Example 4.2, we have:

$$\begin{array}{rcccccccccc} X^{(2)} & = & Y & & Y & & Y & O & O & O & O & O \\ Y^{(2)}_{0k_2} & = & \sigma^{k_2}(X) & O & \sigma^{k_2}(X) & O & O & \sigma^{k_2}(X) & O & O. \end{array}$$

Keeping k_1 fixed to zero and varying k_2, the structure of coincidences we obtain will look like:

$YY:$																										$k_2\downarrow$	$s\downarrow$
$\sigma^{k_2}(X)\sigma^{k_2}(X):$							|1	1	1	1	0	0	0	0	0|	1	1	1	0	0	0	0	0|			0	0
								|1	1	1	0	0	0	0	0	1|	1	1	0	0	0	0	0	0|		1	0
|0	0	0	0	0	0	0	1	1	|1	0	0	0	0	0	1	1|										2	1
	|0	0	0	0	0	1	1	1	|0	0	0	0	0	1	1	1|										3	1
		|0	0	0	0	1	1	1	0	|0	0	0	0	1	1	1	0|									4	1
			|0	0	0	1	1	1	0	0	|0	0	0	1	1	1	0	0|								5	1
				|0	0	1	1	1	0	0	0	|0	0	1	1	1	0	0	0|							6	1
					|0	1	1	1	0	0	0	0	|0	1	1	1	0	0	0	0|						7	1

Each line in the table corresponds to a value of k_2 and displays the string $\sigma^{k_2}(X)\sigma^{k_2}(X)$. This string is moved over k_2+1 positions to the right, since we are interested in $\gamma^{(2)}_{(0k_2)+1}(X^{(2)}, Y^{(2)}_{0k_2})$. Then, for each value of k_2 the corresponding value of s (being either 0 or 1) is computed. If $s=1$, then the string is moved over $M=8$ positions back to the left. By construction, the number of coincidences of YY with the k_2-line in the table is a lower bound for $\gamma^{(2)}_{(0k_2)+1}(X^{(2)}, Y^{(2)}_{0k_2})$. In each of the lines of the table, we have coincidences with both bold ones in YY. Therefore,

$$\gamma^{(2)}_{(0k_2)+1}(X^{(2)}, Y^{(2)}_{0k_2}) \geq 2.$$

Combining this with (4.14), we see that (DG2) holds. Adapting this argument for other values of M and m is straightforward.

As we have seen in the proof of the previous lemma, it is possible to find sufficient independent left and right triangles. Therefore, we have completed our

proof that the distributed growth condition is satisfied. We also already saw $p > 1/\sqrt{M}$ implies that $\gamma > 1$, and hence we can use Theorem 4.2 to finish the proof of Lemma 4.7. □

Theorem 4.3. *For correlated fractal (m, M, p)-percolation we have*

1. *If $\gamma > 1$ then $F_1 - F_2$ contains an interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.*
2. *If $\gamma \leq 1$, then $F_1 - F_2$ contains no interval a.s.*

Proof. This result is the combination of Lemma 4.6, Lemma 4.7 and Proposition 4.1. In the latter case we use that $\mu(\emptyset) > 0$ implies that μ is not entangled with itself, and that otherwise $m = \sqrt{M}$ is for $M \geq 4$ smaller or equal to $M/2$, and thus is also not entangled with itself. □

We remark here that since these results will also hold if we merely require that all sets with m elements have positive probability to occur, the theorem will also be true in this more general case.

4.6 The lower spectral radius in the symmetric case

In this section we show that the distributed growth condition propagates to higher order Cantor sets. As a consequence, the spectral radius characterization obtained in [8] can be extended to joint survival distributions satisfying the DGC.

Lemma 4.8. *(Propagation of the distributed growth condition to higher orders) Suppose the pair of joint survival distributions (μ, λ) satisfies the* DGC. *Then for all $n \geq 1$, the pair of nth order joint survival distributions $(\mu^{(n)}, \lambda^{(n)})$ satisfies the* DGC.

Proof. Choose a string $\underline{k}_n \in \mathbb{A}^{(n)}$, which we write as $\underline{k}_n = kk_2 \dots k_n$, with $k \in \mathbb{A}$ and $k_2 \dots k_n \in \mathbb{A}^{(n-1)}$. We check that we can find nth order sets satisfying the DGC for this $\underline{k}_n$. Since the pair (μ, λ) satisfies the DGC, there exist first order

sets $X_k, Y_k \subseteq \mathbb{A}$ satisfying (DG0), (DG1) and (DG2). Define

$$\begin{aligned} X_k^{(n)} &:= \left\{ \underline{l}_n = l_1 \dots l_n \in \mathbb{A}^{(n)} : l_j \in X_k \text{ for all } j = 1, \dots, n \right\}, \\ Y_k^{(n)} &:= \left\{ \underline{l}_n = l_1 \dots l_n \in \mathbb{A}^{(n)} : l_j \in Y_k \text{ for all } j = 1, \dots, n \right\}. \end{aligned}$$

Obviously, $\mu^{(n)}(X_k^{(n)}) > 0$ and $\lambda^{(n)}(Y_k^{(n)}) > 0$. Define a new pair of nth order joint survival distributions by $\mu_k^{(n)}(X_k^{(n)}) = \lambda_k^{(n)}(Y_k^{(n)}) = 1$. Also define a first order deterministic pair of joint survival distributions by $\mu_k(X_k) = \lambda_k(Y_k) = 1$. The expectation matrices belonging to these nth order survival distributions are related to those belonging to the first order survival distributions by

$$\mathcal{M}_k^{(n)}(\underline{k}_n) = \mathcal{M}_k(\underline{k}_n) = \mathcal{M}_k(k)\mathcal{M}_k(k_2) \dots \mathcal{M}_k(k_n).$$

Using that X_k and Y_k satisfy (DG1) and (DG2), and that the columns sums of the expectation matrices are equal to the correlation coefficients, we obtain that

$$[1 \;\; 1]\mathcal{M}_k^{(n)}(\underline{k}_n) = [1 \;\; 1]\mathcal{M}_k(k) \prod_{j=2}^{n} \mathcal{M}_k(k_j) \geq [2 \;\; 2] \prod_{j=2}^{n} \mathcal{M}_k(k_j) \geq [2 \;\; 2]$$

elementwise, which means that $Z_k^{(n);L}(\underline{k}_n) \geq 2$ and $Z_k^{(n);R}(\underline{k}_n) \geq 2$, or equivalently

$$\gamma_{\underline{k}_n}(X_k^{(n)}, Y_k^{(n)}) \geq 2; \qquad \gamma_{\underline{k}_n+1}(X_k^{(n)}, Y_k^{(n)}) \geq 2.$$

Similarly $\gamma_{\underline{l}_n}(X_k^{(n)}, Y_k^{(n)}) \geq 1$ for all $\underline{l}_n \in \mathbb{A}^{(n)}$. It follows that the pair $(\mu^{(n)}, \lambda^{(n)})$ satisfies the DGC. □

This propagation property leads to the theorem below. The lower spectral radius $\underline{\rho}(\Sigma)$ of a set Σ of square matrices is defined by

$$\underline{\rho}(\Sigma) := \liminf_{n\to\infty} \min_{A_1,\dots,A_n \in \Sigma} ||A_1 \dots A_n||^{1/n},$$

for some matrix norm $||\cdot||$. For two M-adic random Cantor sets, let Σ_M be the

corresponding collection of expectation matrices

$$\Sigma_M := \{\mathcal{M}(0), \ldots, \mathcal{M}(M-1)\}\,. \tag{4.15}$$

Then we obtain the following result:

Theorem 4.4. *Consider the algebraic difference $F_1 - F_2$ between two M-adic independent random Cantor sets F_1 and F_2 with the same joint survival distribution satisfying the distributed growth condition.*

1. *If $\underline{\rho}(\Sigma_M) > 1$, then $F_1 - F_2$ contains no interval a.s. on $\{F_1 - F_2 \neq \emptyset\}$.*
2. *If $\underline{\rho}(\Sigma_M) < 1$, then $F_1 - F_2$ contains no interval a.s.*

Proof. The proof is basically the same as the proof of Theorem 6.1 in [8]. There is a difference in the fact that here we do not require irreducibility explicitly. From the symmetry $\mu = \lambda$ it follows that $m_e = m_{-e}$ for all $e \in \mathbb{A} \cup -\mathbb{A}$. Now, since the DGC holds, we get the irreducibility for free.

After derivation of the same statements concerning the nth order correlation coefficients as in [8], we apply our Theorem 4.2. This is justified by the fact that the DGC propagates to higher orders, as was shown in Lemma 4.8. □

Chapter 5

New methods to determine the critical probability in fractal percolation

5.1 Introduction

Fractal percolation has been introduced by Mandelbrot in 1974 as a model for turbulence and is discussed in his book *The Fractal Geometry of Nature* [21]. Several equivalent formal definitions of this process can be found in the literature (see e.g. [4, 7, 14]). Here we only give an informal definition of the two-dimensional case. Let K_0 be the unit square and choose an integer $M \geq 2$ and a parameter $p \in [0, 1]$. To obtain K_1, divide K_0 into M^2 equal subsquares, each of which survives with probability p and is discarded with probability $1 - p$, independently of all other subsquares. Now do the same procedure in all surviving squares, in order to obtain K_2. Iterating this process gives a decreasing

sequence of sets $(K_n)_{n\in\mathbb{N}}$, see Figure 5.1. Let $K = \bigcap_{n\in\mathbb{N}} K_n$ be the limit set.

Figure 5.1: Realizations of $K_n, n = 1, 2, 3, 5, 7$ for $M = 3$ and $p = 0.85$.

It was shown in 1988 by Chayes, Chayes and Durrett [4] that there exists a non-trivial critical value $p_c(M)$ such that a.s. the largest connected component in K is a point for $p < p_c(M)$ and with positive probability there is a connected component intersecting opposite sides of the unit square for $p \geq p_c(M)$.

For all $M \geq 2$, the value of $p_c(M)$ is unknown. Several attempts have been made to find bounds for $p_c(M)$. It is easy to see that K is empty a.s. if $p \leq 1/M^2$, which implies $p_c(M) > 1/M^2$. The argument in [4] is already a bit smarter: any left-right crossing has to cross the line $\{1/M\} \times [0, 1]$ somewhere. A crossing of this line in K_n means that there is a pair of adjacent squares on opposite sides of this line. Such pairs form a branching process with mean offspring p^2M and consequently $p_c(M) > 1/\sqrt{M}$. For the case $M = 2$ this was sharpened by White in 2001 to $p_c(2) \geq 0.810$, who used a set that dominates K and has a simpler structure to study.

Sharp upper bounds are harder to obtain. The first idea to get rigorous upper bounds for $M \geq 2$ was given by Chayes, Chayes and Durrett [4], but (in their own words) these bounds are ridiculously close to 1. For $M = 3$, they show that $p_c(3) < 0.9999$ (although in fact one can prove that $p_c(3) < 0.993$ with their method), which was improved by Dekking and Meester [9] to $p_c(3) < 0.991$. Chayes et al only treat $M = 3$, but they point out that the same idea works for any $M \geq 3$. The case $M = 2$ can be treated by comparing with $M = 4$. As is noted by van der Wal [30], a coupling argument gives $p_c(2) \leq 1-(1-\sqrt{p_c(4)}\,)^4$. Following this approach gives $p_c(4) < 0.998$ and $p_c(2) < 1 - 10^{-12}$.

In this chapter we present ideas to find significantly sharper lower and upper bounds. To find lower bounds, we compare fractal percolation with site percolation. A fundamentally new result is that for all M we construct an increasing computable sequence that converges to $p_c(M)$. The terms in the sequence can in principle be calculated algorithmically. To actually obtain numerical results we bound them from below. These ideas lead to (computer aided) proofs that $p_c(2) > 0.881$ and $p_c(3) > 0.784$.

For the upper bounds, we map all possible realizations of K_n to a finite alphabet $\mathscr{A}$. The choice of the alphabet $\mathscr{A}$ is inspired by van der Wal's work [30]. The fractal percolation iteration process now induces an iterative random process on $\mathscr{A}$, which is easier to analyze than the original process. This theoretical framework is the basis of computer aided proofs for the following upper bounds: $p_c(2) < 0.993$, $p_c(3) < 0.940$ and $p_c(4) < 0.972$.

5.2 Lower bounds for $p_c(M)$

In this section we develop methods to calculate lower bounds for the critical value of two-dimensional fractal percolation. First we briefly introduce site percolation and then we prove a coupling with fractal percolation that allows us to find lower bounds for $p_c(M)$. In particular, we construct an increasing sequence of lower bounds and we prove that this sequence converges to $p_c(M)$. At the end of this section we show how to use these insights to obtain numerical results.

5.2.1 Site percolation

Consider the infinite two-dimensional square lattice in which each vertex is open with probability p and closed otherwise. In this model the percolation probability $\zeta(p)$ is defined as the probability that the origin belongs to an infi-

nite open cluster. The critical probability is given by

$$p_c^{site} := \inf \{p : \zeta(p) > 0\}.$$

It has been shown by van den Berg and Ermakov [3] that $p_c^{site} > 0.556$. Now consider the probability that a box of $n \times n$ vertices is crossed by an open cluster. It is a classical result that in the subcritical regime, this probability converges to zero as n increases. This property will be used to couple site percolation to fractal percolation.

Property 5.1. *Take a box of $n \times n$ vertices. Suppose $p < p_c^{site}$. Then the probability that there is an open cluster intersecting opposite sides of the box converges to 0 as $n \to \infty$.*

5.2.2 Coupling site percolation and fractal percolation

In fractal percolation, a set in the unit square is said to percolate if it contains a connected component intersecting opposite sides of the square. Let

$$\theta_n(p, M) = \mathbb{P}(K_n(p, M) \text{ percolates}), \qquad \theta(p, M) = \mathbb{P}(K(p, M) \text{ percolates}).$$

Then $p_c(M) := \inf \{p : \theta(p, M) > 0\}$. We will often suppress some of the dependence on M and p. It is well known (see [22]) that

$$\lim_{n\to\infty} \theta_n(p) = \theta(p) = \mathbb{P}(\bigcap_{n=0}^{\infty} \{K_n(p) \text{ percolates}\}).$$

To obtain a proper coupling, we have to deal with diagonal connections. For example, the set $[0, 1/2]^2 \cup [1/2, 1]^2$ percolates. We would like to ignore such diagonal connections, since in site percolation diagonal connections do not exist. Therefore we redefine percolation as follows. We say K_n percolates if there is a left-right crossing of the square that does not use diagonal connections.

Since diagonal connections break down almost surely, this does not change the limiting percolation probability. Also from now on connections, crossings and connected components in K_n are similarly redefined.

Theorem 5.1. *Let $\pi_n(p, M) = \mathbb{P}$(at least two sides are connected in $K_n(p, M)$). If $\pi_n(p) < p_c^{site}$ for some n, then $p < p_c(M)$.*

Proof. First define *delayed* fractal percolation: $F_{m,n}$ is constructed in the same way as K_{m+n}, the only difference being that we do not discard any squares in the first m construction steps. So we first divide the unit square into $M^m \times M^m$ subsquares and only then we start the fractal percolation process in each of these squares. Delayed fractal percolation stochastically dominates fractal percolation.

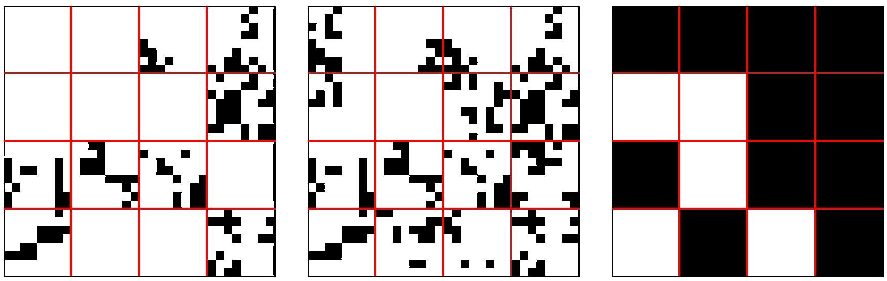

Figure 5.2: Three coupled realizations to illustrate (5.1). Left: K_5 for $M = 2$ and $p = 2/3$. Middle: $F_{2,3}$ for $M = 2$ and $p = 2/3$. Right: K_1 for $M = 4$ and $p = \pi_3(2/3, 2)$.

Then we have the following inequalities (illustrated for $M = m = 2$ and $n = 3$ in Figure 5.2):

$$\theta_{m+n}(p, M) \leq \mathbb{P}(F_{m,n}(p, M) \text{ percolates}) \leq \theta_1(\pi_n(p), M^m). \tag{5.1}$$

The first inequality follows from the fact that $F_{m,n}$ is stochastically larger than K_{m+n}. The second inequality can be explained as follows. Suppose we have two types of squares: realizations of $K_n(p, M)$ in which at least two sides are

connected (type 1) and realizations in which no side is connected to any other side (type 2). Suppose we tile a larger square with M^{2m} independent realizations of $K_n(p, M)$. So the probability on a type 1 square is $\pi_n(p)$. This larger square is a (scaled) realization of $F_{m,n}(p, M)$. Now replace all type 1 squares by a full square and discard all type 2 squares. This gives a realization of $K_1(\pi(p), M^m)$. Moreover, this replacement procedure can not destroy percolation. To see this, suppose we have a left-right crossing in $F_{m,n}(p, M)$. All parts of this crossing that are in type 1 squares are preserved. If the crossing uses a type 2 square S, then it enters and leaves S at the same side of S. At this side the neighboring square T must be of type 1. So if we remove S and replace T by a full square, there still is a left-right crossing.

A first level fractal percolation set can be seen as site percolation in a finite box. Suppose $\pi_n(p) < p_c^{site}$ for some n and let the box size M^m tend to ∞. By Property 5.1 we arrive at

$$\lim_{m\to\infty} \theta_1(\pi_n(p), M^m) = 0. \tag{5.2}$$

Therefore, if $\pi_n(p) < p_c^{site}$, by (5.1) we find that

$$\theta(p) = \lim_{m\to\infty} \theta_{m+n}(p) = 0,$$

which in turn implies $p < p_c(M)$. □

5.2.3 A convergent sequence of lower bounds for $p_c(M)$

In this section we define a sequence of lower bounds for $p_c(M)$. We prove that this sequence converges to the $p_c(M)$. Let

$$p_c^n(M) = \sup\left\{p : \pi_n(p, M) < p_c^{site}\right\}. \tag{5.3}$$

Note that $\pi_n(p_c^n) \leq p_c^{site}$ for all n, since $\pi_n(p)$ is continuous in p. Theorem 5.1 requires a strict inequality, but this does not give any problem since $\pi_n(p)$ is strictly decreasing in n, so $\pi_{n+1}(p_c^n) < p_c^{site}$ for all n and hence indeed $p_c^n(M) < p_c(M)$ for all n. The strict monotonicity in n also implies that $(p_c^n(M))_{n=0}^{\infty}$ is increasing. The obvious question now is whether $(p_c^n(M))_{n=0}^{\infty}$ converges to $p_c(M)$. We will show that this is indeed the case. First we need that $\pi_n(p)$ goes to zero if the fractal percolation is subcritical:

Lemma 5.1. *If $p < p_c(M)$, then $\lim_{n\to\infty} \pi_n(p, M) = 0$.*

Proof. Suppose $p < p_c(M)$, so $\theta(p) = 0$. Note that a.s. there is an n such that in K_n the two squares in the top left and bottom left corner are discarded already. Conditioned on this event, a connection in the limiting set from the left side to any other side can only occur if it horizontally crosses the vertical strip S consisting of the squares $[0, M^{-n}] \times [jM^{-n}, (j+1)M^{-1}]$ for $j = 1, \ldots, M^n - 2$. From selfsimilarity and subcriticality it follows that in the limit each of these squares has zero probability to contain a component connecting opposite sides (horizontally and vertically). So in K a horizontal crossing of S can only occur if it crosses a block of two vertically adjacent squares. But as Dekking and Meester showed (Lemma 5.1 in [9]), $\theta(p) = 0$ implies that such a block crossing has zero probability as well. It follows that connections from the left side to any other side have zero probability to occur. If this holds for the left side, then it holds for all sides, and so $\lim_{n\to\infty} \pi_n(p) = 0$. □

The previous lemma makes it easy to prove convergence:

Theorem 5.2. *The sequence $(p_c^n(M))_{n=0}^{\infty}$ converges to $p_c(M)$ if $n \to \infty$.*

Proof. Let $\varepsilon > 0$ and suppose $p = p_c(M) - \varepsilon$. Then by Lemma 5.1, there exists an N such that $\pi_n(p) < p_c^{site}$ for all $n \geq N$. Therefore, $p_c^n(M) > p$ for all $n \geq N$. □

The theory developed so far gives us an algorithmic tool to calculate an in-

creasing and converging sequence of lower bounds for $p_c(M)$. Still, it is not so easy to actually obtain sharp bounds. For example, a little thought gives

$$\begin{aligned}\pi_1(p,2) &= 1-(1-p)^4,\\ \pi_1(p,3) &= 1+(1-p)^4(p^5+4p^4(1-p)+6p^3(1-p)^2-1),\end{aligned}$$

so using the bound of van den Berg and Ermakov for p_c^{site}, we find $p_c^1(2) > 0.183$ and $p_c^1(3) > 0.178$. Of course, we want to find a bit sharper bounds, so we should take larger values of n. However, for large n, the functions $\pi_n(p)$ are very complicated polynomials, and it is not clear how to find them in reasonable time. In the next section we will discuss a way to avoid this problem.

5.3 Classifying realizations

In this section we introduce the idea to map realizations of K_n to a finite alphabet $\mathscr{A}$ that does not depend on n. Such a map will be called a *classification*. In this way, we can simplify the substitution process without losing too much essential information on the connectivity structure in K_n. The construction of the random sets $(K_n)_{n=0}^{\infty}$ now induces a sequence of probability measures on $\mathscr{A}$. These ideas can be used to obtain both lower and upper bounds for $p_c(M)$. For the lower bounds we choose the classification in such a way that we can calculate upper estimates for $\pi_n(p)$ and then we can approximate terms of $(p_c^n)_{n=0}^{\infty}$ using the van den Berg and Ermakov bound for p_c^{site}. For the upper bounds we will construct a classification that permits us to bound $\theta(p)$ away from 0 for p sufficiently large.

Let $\mathscr{K}_n$ be the set of all possible realizations of K_n. We will define a partially ordered alphabet $\mathscr{A}$ and the classification will actually be a sequence of maps $\mathscr{C} = (\mathscr{C}_n)_{n=0}^{\infty}$, where $\mathscr{C}_n$ is a map from $\mathscr{K}_n$ to $\mathscr{A}$. The alphabet will have a

unique minimum and maximum, with the property that

$$\mathscr{C}_n(\emptyset) = \min(\mathscr{A}) \quad \text{and} \quad \mathscr{C}_n([0,1]^2) = \max(\mathscr{A}). \tag{5.4}$$

Realizations of K_n can be mapped in a natural way to words in $\mathscr{A}^{M\times M}$. A realization of K_n is obtained by first generating a realization of K_1 and then replacing all squares that survived by independent scaled realizations of K_{n-1}. Given $\mathscr{C}_{n-1}$, this induces a map $\mathscr{W}_n$ from $\mathscr{K}_n$ to $\mathscr{A}^{M\times M}$. If $\mathscr{C}_n(K_n)$ only depends on the word $\mathscr{W}_n(K_n)$ and not on n, we say the classification is *regular*:

Definition 5.1. *Let ϕ be a map from $\mathscr{A}^{M\times M}$ to $\mathscr{A}$. Define $\mathscr{C} = (\mathscr{C}_n)_{n=0}^{\infty}$ by letting*

$$\mathscr{C}_n = \phi \circ \mathscr{W}_n, \qquad n \geq 1. \tag{5.5}$$

Then we say $\mathscr{C}$ is a regular classification *and ϕ is called the* word code *of $\mathscr{C}$.*

Example 5.1. Let $M = 2$ and take the alphabet $\mathscr{A} = \{\square, \boxplus\}$, where $\min \mathscr{A} = \square$ and $\max \mathscr{A} = \boxplus$. For a 2×2 word w over $\mathscr{A}$ let $\phi(w) = \boxplus$ if and only if at least one of the letters in w equals $\boxplus$. Let $\mathscr{C} = (\mathscr{C}_n)_{n=0}^{\infty}$ be regular with word code ϕ. Then $\mathscr{C}_0$ is determined by (5.4):

$$\mathscr{C}_0(K_0(p)) = \mathscr{C}_0([0,1]^2) = \boxplus.$$

For $n \geq 1, \mathscr{C}_n = \phi \circ \mathscr{W}_n$. For instance

$$\mathscr{C}_1([\frac{1}{2},1]^2) = \phi(\mathscr{W}_1([\frac{1}{2},1]^2)) = \phi\begin{pmatrix} \mathscr{C}_0(\emptyset) & \mathscr{C}_0([0,1]^2) \\ \mathscr{C}_0(\emptyset) & \mathscr{C}_0(\emptyset) \end{pmatrix} = \phi\begin{pmatrix} \square & \boxplus \\ \square & \square \end{pmatrix} = \boxplus.$$

■

We now want to analyze the probabilities $\mathbb{P}(\mathscr{C}(K_n(p)) = a)$, where $a \in \mathscr{A}$. Suppose $\mathscr{C}$ is a regular classification with word code ϕ. Let $x \in [0,1]^{|\mathscr{A}|}$ be a probability vector on $\mathscr{A}$, and suppose we construct an $M \times M$ word w in which

all letters are chosen independently according to x. Define $F_{\mathscr{C}}(x) \in [0,1]^{|\mathscr{A}|}$ by

$$(F_{\mathscr{C}}(x))_a = \mathbb{P}_x(\phi(w) = a), \qquad a \in \mathscr{A}.$$

The function $F_{\mathscr{C}} : [0,1]^{|\mathscr{A}|} \to [0,1]^{|\mathscr{A}|}$ will be the key to calculate the probabilities $\mathbb{P}(\mathscr{C}(K_n(p)) = a)$ in an iterative way, as is shown in the next lemma. Let $\tau^n(p)$ be a vector indexed by the elements of $\mathscr{A}$ such that

$$\tau^n_a(p) = \mathbb{P}(\mathscr{C}(K_n(p)) = a).$$

Let $\tau^{\square}$ and $\tau^{\boxplus}$ be the vectors that assign full probability to $\min(\mathscr{A})$ and $\max(\mathscr{A})$ respectively.

Lemma 5.2. *If the classification $\mathscr{C}$ is regular, then*

$$\tau^{n+1}(p) = F_{\mathscr{C}}(p\tau^n(p) + (1-p)\tau^{\square}) \quad \textit{with initial condition} \quad \tau^0(p) = \tau^{\boxplus}.$$

Proof. The letters in $\mathscr{W}_{n+1}(K_{n+1}(p))$ are independent. With probability p a letter corresponds to a scaled realization of $K_n(p)$, with probability $1-p$ it corresponds to an empty square. So each letter occurs according to the probability vector $p\tau^n(p) + (1-p)\tau^{\square}$. This gives the recursion. The initial condition follows from (5.4). □

This recursion formula is essentially a generalization of the recursion given in [9].

5.3.1 Strategy for lower bounds

The strategy to find lower bounds for p_c now is as follows. Define an alphabet $\mathscr{A}$ with subset $\mathscr{A}_\pi$ and define a classification $\mathscr{C}$ (by choosing ϕ) in such a way that $\mathbb{P}(\mathscr{C}(K_n(p)) \in \mathscr{A}_\pi) \geq \pi_n(p)$ for all n. Now search for the largest p for which this probability is smaller than 0.556. We will give an example to illustrate this

procedure. In this example (which only gives a very moderate bound) $\mathscr{A}$ only has two elements, and therefore exactly the same recursion as in [9] pops up.

Example 5.2. (A lower bound for $M = 2$) Let $\mathscr{C}$ be the classification of Example 5.1. By induction it follows that if two sides are connected in K_n, then $\mathscr{C}_n(K_n) = \boxplus$. Consequently $\tau^n_{\boxplus}(p) \geq \pi_n(p)$. From the definition of ϕ it follows that in this case

$$\tau^{n+1}_{\boxplus}(p) = 1 - (1 - p\tau^n_{\boxplus}(p))^4 = \tau^1_{\boxplus}(p\tau^n_{\boxplus}(p)).$$

The function $G_p(x) := \tau^1_{\boxplus}(px)$ is increasing and $\tau^n_{\boxplus}(p)$ decreases to the largest fixed point of $G_p(\cdot)$. Choosing $p = 0.33$, we find $\pi_{50}(p) \leq \tau^{50}_{\boxplus}(p) \approx 0.554 < 0.556 \leq p_c^{site}$, and consequently $p_c^{50}(2) > 0.33$ and therefore also $p_c(2) > 0.33$. ■

5.3.2 Strategy for upper bounds

Our recipe to find upper bounds for $p_c(M)$ is a bit more involved. We already have a partial ordering on $\mathscr{A}$, but to find upper bounds it will also be convenient to partially order the probability vectors on $\mathscr{A}$. A set $S \subseteq \mathscr{A}$ will be called *increasing* if $a \in S$ implies $b \in S$ for all $b \succeq a$. For probability vectors x and y, we now write $x \succeq y$ if x assigns larger probabilities to all increasing subsets of $\mathscr{A}$. We say the function $F_{\mathscr{C}}$ is increasing if $F_{\mathscr{C}}(x) \succeq F_{\mathscr{C}}(y)$ for $x \succeq y$.

Lemma 5.3. *Let $\mathscr{C}$ be a regular classification for which $F_{\mathscr{C}}$ is increasing. Then the sequence $(\tau^n(p))_{n=0}^{\infty}$ is decreasing and $\tau^{\infty}(p) := \lim_{n\to\infty} \tau^n(p)$ exists. If $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x$ for some probability vector $x \in [0,1]^{|\mathscr{A}|}$ and $p \in (0,1]$, then $\tau^{\infty}(p) \succeq x$.*

Proof. Since $\tau^0(p) = \tau^{\boxplus}$, we have $\tau^0(p) \succeq \tau^1(p)$. Suppose $\tau^n(p) \succeq \tau^{n+1}(p)$. Then $\tau^{n+1}(p) = F_{\mathscr{C}}(p\tau^n(p) + (1-p)\tau^{\square}) \succeq F_{\mathscr{C}}(p\tau^{n+1}(p) + (1-p)\tau^{\square}) = \tau^{n+2}(p)$ since $F_{\mathscr{C}}$ is increasing. So $(\tau^n(p))_{n=0}^{\infty}$ is decreasing. If $S \subseteq \mathscr{A}$ is an increasing

set, then

$$\tau_S^n(p) := \mathbb{P}(\mathscr{C}(K_n) \in S)$$

decreases in n and is bounded from below by 0, so $\lim_{n\to\infty} \tau_S^n(p)$ exists. There are only finitely many increasing subsets of $\mathscr{A}$ and their limiting probabilities uniquely determine $\tau^\infty(p)$.

For the second statement, $\tau^0(p) = \tau^{\boxplus} \succeq x$. Now suppose $\tau^n(p) \succeq x$ for some n. Then

$$\tau^{n+1}(p) = F_{\mathscr{C}}(p\tau^n(p) + (1-p)\tau^{\square}) \succeq F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x.$$

and therefore $\tau^n(p) \succeq x$ for all $n \in \mathbb{N}$. Hence $\tau^\infty(p) \succeq x$. □

Corollary 5.1. *Let $\mathscr{C}$ be a regular classification for which $F_{\mathscr{C}}$ is increasing. Suppose $x \in [0,1]^{|\mathscr{A}|}$ is a probability vector for which $x_{\max(\mathscr{A})} > 0$. If $\theta_n(p) \geq \tau^n_{\max(\mathscr{A})}(p)$ for all n, and $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x$ for some $p \in (0,1]$, then*

$$\theta(p) = \lim_{n\to\infty} \theta_n(p) \geq \lim_{n\to\infty} \tau^n_{\max(\mathscr{A})}(p) \geq x_{\max(\mathscr{A})} > 0,$$

and consequently $p > p_c$.

Before we give an example of the procedure to find upper bounds for $p_c(M)$, we will first construct suitable alphabets in the next section.

5.4 Construction of the alphabet and word codes

Our alphabets will be defined by means of non-crossing equivalence relations. Figure 5.3 displays all equivalence relations on an ordered set of 5 elements, of which 42 are non-crossing.

Let $E = \{e_1, \ldots, e_n\}$ be a collection of line segments that together form the boundary of the unit square, numbered clockwise, starting from $(0, 0)$. We define the alphabet $\mathscr{A}_E$ as the set of all non-crossing equivalence relations on E. The size of $\mathscr{A}_E$ is given by the Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

On the alphabet $\mathscr{A}_E$, we can naturally define a partial ordering. For $a, b \in \mathscr{A}_E$ we say $a \succeq b$ if all equivalences that hold in b also hold in a. The equivalences (denoted by $\sim$) represent connections between the elements. The letters can be represented as pictures in a square. For example, if E is the set containing the four sides of the square (so $n = 4$), then $\mathscr{A}_E$ can be represented as

$$\mathscr{A}_E = \{\square, \square, \square, \square, \square, \square, \square, \square, \square, \square, \square, \square, \square\}.$$

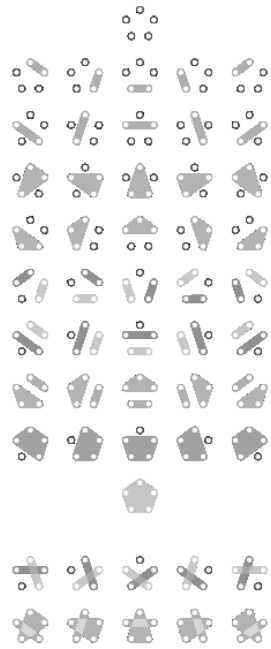

Figure 5.3: Equivalence relations on 5 elements

The alphabet has a unique maximum and minimum, $\max(\mathscr{A}_E) = \boxplus$ and $\min(\mathscr{A}_E) = \square$. From now on we will denote the alphabet that corresponds to dividing each side in M^n elements of equal size by $\mathscr{A}_{M,n}$.

Now that we have letters, we can start making words. We only allow square words, obtained by tiling $[0, M]^2$ by M^2 letters. An example of a 3×3 word over $\mathscr{A}_{3,0}$ is given in the right panel of Figure 5.4. A word w also has a boundary element set E_w. We enumerate clockwise, starting from $(0, 0)$, as follows: $E_w = \left\{e_1^w, \ldots, e_{M|E|}^w\right\}$.

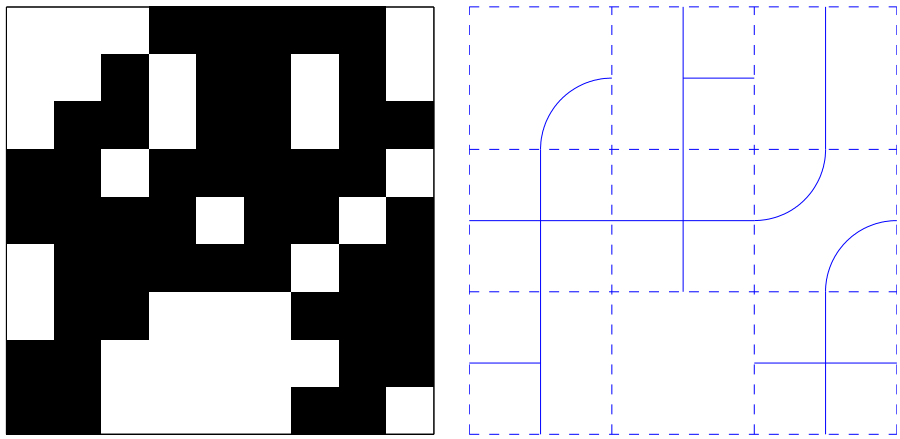

Figure 5.4: Left: A realization of K_2 for $M = 3$. Right: A 3×3 word w over $\mathscr{A}_{3,0}$. If we use the word code $\Psi_{3,0}$, then this word corresponds to the realization on the left.

5.4.1 Weak and strong connectivity

Suppose we partition E_w into subsets $E_w^1, \ldots, E_w^k$ for some integer $k \geq 2$. If $e \sim f$ for some $e \in E_w^i$ and $f \in E_w^j$, we say E_w^i and E_w^j are *connected*. If $E_w^{i_1}, \ldots, E_w^{i_m}$ form a chain of pairwise connected sets, we say $E_w^{i_1}$ and $E_w^{i_m}$ are *weakly connected*. If strictly more than half of the elements in E_w^i and strictly more than half of E_w^j are in the same equivalence class, we say E_w^i and E_w^j are *strongly connected*. Observe that weak and strong connectivity are equivalence relations on $\left\{E_w^1, \ldots, E_w^k\right\}$.

Choose the alphabet $\mathscr{A}_{M,n}$, then $|E| = 4M^n$, so words of size $M \times M$ have $4M^{n+1}$ boundary elements. Partition E_w in the subsets

$$E_w^i = \left\{e_{M(i-1)+j}^w : 1 \leq j \leq M\right\}, \quad 1 \leq i \leq 4M^n.$$

This gives a tool to define word codes: a word defines an equivalence relation on $\left\{E_w^1, \ldots, E_w^{4M^n}\right\}$ that obviously can be mapped to an equivalence relation

on $E = \{e_1, \ldots, e_{4M^n}\}$, which is just a letter in $\mathscr{A}_E$. In this way, define word codes

$$\Phi_{M,n} : \mathscr{A}_{M,n}^{M\times M} \to \mathscr{A}_{M,n} \quad \text{and} \quad \Psi_{M,n} : \mathscr{A}_{M,n}^{M\times M} \to \mathscr{A}_{M,n}$$

based on weak and strong connectivity respectively. The corresponding classifications are increasing.

Example 5.3. Let w be the word in Figure 5.4. This word has 12 boundary elements that will be grouped into four partition sets. For instance, E_w^3 contains the three elements at the right side. Then E_w^1, E_w^2 and E_w^4 are all pairwise connected and E_w^3 is connected to E_w^4. Consequently E_w^i is weakly connected to E_w^j for all i and j. Therefore $\Phi_{3,0}(w) = \boxplus$. The only partition sets that are strongly connected are E_w^1 and E_w^2. Therefore $\Psi_{3,0}(w) = \boxdot$. ■

The idea behind the definitions of $\Phi_{M,n}$ and $\Psi_{M,n}$ is that they guarantee the following key properties:

Property 5.2. *Using the alphabet $\mathscr{A}_{M,k}$ and word code $\Phi_{M,k}$, the following implication holds: if $e_i, e_j \in E$ are connected in K_n, then $e_i \sim e_j$ in $\mathscr{C}(K_n)$.*

Property 5.3. *Using the alphabet $\mathscr{A}_{M,k}$ and word code $\Psi_{M,k}$, the reversed implication holds: if $e_i \sim e_j$ in $\mathscr{C}(K_n)$, then e_i and e_j are connected in K_n.*

These properties are not hard to prove. Therefore we do not give a fully detailed proof. For Property 5.2 a straightforward inductive argument suffices.

For Property 5.3, one should note that we actually have a stronger implication, namely that e_i and e_j are connected in a special way that we will explain using the realization in Figure 5.4. In this realization the left side and the top side are connected. At both these sides there are two (strictly more than $M/2$) first level squares each of which contains two second level squares that survived. These second level squares are all in the same connected component. This property is easily generalized to higher level realizations and other values of M. If two neighbouring squares both have such connection from their joint side to an

other side, then these two connections are in the same connected component.

The following lemma shows that these alphabets and word codes are suitable for our purposes, see the discussion in Section 5.3.

Lemma 5.4. *Using the word codes $\Phi_{M,k}$ and $\Psi_{M,k}$ leads to two inequalities:*

1. *Take the alphabet $\mathscr{A}_{M,k}$ and define a classification $\mathscr{C}$ by the word code $\Phi_{M,k}$. Let $\mathscr{A}_\pi \subseteq \mathscr{A}_{M,k}$ be the set of all letters in which at least two of the sides are connected. Then $\tau_\pi^n(p) := \mathbb{P}(\mathscr{C}(K_n(p)) \in \mathscr{A}_\pi) \geq \pi_n(p)$ for all n.*
2. *Take the alphabet $\mathscr{A}_{M,k}$ and define a classification $\mathscr{C}$ by the word code $\Psi_{M,k}$. Then $\tau^n_{\max(\mathscr{A})}(p) \leq \theta_n(p)$ for all n.*

Proof. These statements follow from Property 5.2 and 5.3 respectively. □

Now we are ready to give an example illustrating how to find upper bounds. We will keep the example as simple as possible, so that it can be checked by hand. Therefore our alphabet will contain only two letters and we will use a word code that is even simpler than $\Psi_{3,0}$. Nevertheless, it leads to a bound that already improves upon the best bound known so far.

Example 5.4. (An upper bound for $M = 3$) Let $\mathscr{A} = \{\max(\mathscr{A}_{3,0}), \min(\mathscr{A}_{3,0})\} = \{\square, \boxplus\}$. A 3×3 word w has twelve boundary elements that we partition into four sets as before. Define a classification $\mathscr{C}$ by letting $\phi(w) = \boxplus$ if and only if all sets $E_w^1, \ldots, E_w^4$ are strongly connected. This classification is increasing and satisfies Property 5.3, so $\tau^n_{\max(\mathscr{A})}(p) \leq \theta_n(p)$ for all n. The recursion of Lemma 5.2 reduces to $\tau_\boxplus^{n+1}(p) = \tau_\boxplus^1(p\tau_\boxplus^n(p))$, where

$$\tau_\boxplus^1(p) = p^9 + 9p^8(1-p) + 20p^7(1-p)^2.$$

This function is the first component of $F_\mathscr{C}(p\tau^\boxplus + (1-p)\tau^\square) = F_\mathscr{C}((p, 1-p))$. Now choose $p = 0.984$ and $x = 0.9720$. Then $\tau_\boxplus^1(px) \approx 0.9721 > x$. Therefore the probability vector $(x, 1-x)$ satisfies the requirement of Corollary 5.1. Hence

$p_c(3) < 0.984$. ■

5.4.2 Monotonicity and convergence

So far we developed some tools to find bounds for $p_c(M)$. One would expect that taking larger alphabets results in sharper bounds, since we can approximate the connectivity structure in K_n more accurately. In this section we show that this is indeed the case and that the lower bounds even convergence to $p_c(M)$. Unfortunately, we do not know if the upper bounds also converge to $p_c(M)$.

For the word code $\Phi_{M,k}$ over $\mathscr{A}_{M,k}$, define the corresponding classification and let $\tau_\pi^n(p)$ be defined as before. Then define a critical value as follows:

$$p_c(\Phi_{M,k}) := \sup\left\{p : \tau_\pi^\infty(p) < p_c^{site}\right\}.$$

Let $\mathscr{A} = \mathscr{A}_{M,k}$ and define $\mathscr{C}$ by the word code $\Psi_{M,k}$. Also here we define a critical value:

$$p_c(\Psi_{M,k}) := \inf\left\{p : \tau_{\max \mathscr{A}}^\infty(p) > 0\right\}.$$

Now we have the following proposition:

Proposition 5.1. *The sequence $(p_c(\Phi_{M,k}))_{k=0}^\infty$ is increasing and $(p_c(\Psi_{M,k}))_{k=0}^\infty$ is decreasing. Moreover,*

$$\lim_{k\to\infty} p_c(\Phi_{M,k}) = p_c(M).$$

Proof. The alphabet $\mathscr{A}_{M,k}$ contains equivalence relations on an element set $E_{M,k}$. The element set $E_{M,k+1}$ of $\mathscr{A}_{M,k+1}$ can be obtained by dividing each element $e \in E_{M,k}$ into M equal pieces. These pieces will be called children of e, and e will be called the parent of its children. Define classifications $\mathscr{C}^k$ and $\mathscr{C}^{k+1}$ by the word codes $\Phi_{M,k}$ and $\Phi_{M,k+1}$. Denote the corresponding probability vectors by ${}^k\tau$ and ${}^{k+1}\tau$. By induction it follows that if two elements are

connected in $\mathscr{C}^{k+1}(K_n)$, then their parents are connected in $\mathscr{C}^k(K_n)$. Consequently, ${}^k\tau_\pi^n(p) \geq {}^{k+1}\tau_\pi^n(p)$ and so

$$p_c(\Phi(M,k+1)) \geq p_c(\Phi(M,k)).$$

Now define $\mathscr{C}^k$ and $\mathscr{C}^{k+1}$ by the word codes $\Psi_{M,k}$ and $\Psi_{M,k+1}$. If two elements are connected in $\mathscr{C}^k(K_n)$, then two of their children are connected in $\mathscr{C}^{k+1}(K_n)$. Henceforth, ${}^k\tau^n_{\max\mathscr{A}}(p) \leq {}^{k+1}\tau^n_{\max\mathscr{A}}(p)$, so

$$p_c(\Psi(M,k+1)) \leq p_c(\Psi(M,k)).$$

Take the alphabet $\mathscr{A}_{M,k}$ and define $\mathscr{C}^k$ by the word code $\Phi_{M,k}$. A realization of K_n consists of $M^n \times M^n$ squares, and letters in $\mathscr{A}_{M,k}$ have M^k boundary elements at each side. This means that for $n \leq k$ the classification describes the connectivity structure exactly: two elements in $\mathscr{C}^k(K_n)$ are connected if and only if they are connected in K_n. So $\pi_n(p) = {}^k\tau_\pi^n(p)$ if $n \leq k$. This implies that for $n \leq k$ we can rewrite (5.3):

$$\begin{aligned} p_c^n(M) &= \sup\left\{p : {}^k\tau_\pi^n(p) < p_c^{site}\right\} \\ &\leq \sup\left\{p : {}^k\tau_\pi^\infty(p) < p_c^{site}\right\} = p_c(\Phi_{M,k}) \leq p_c(M), \end{aligned}$$

where we used that ${}^k\tau_\pi^n(p)$ decreases in n by Lemma 5.3. Theorem 5.2 states that $p_c^n(M)$ converges to $p_c(M)$, so we conclude that $\lim_{k\to\infty} p_c(\Phi_{M,k}) = p_c(M)$. □

5.5 Numerical results

In this section we present our numerical results. Our principal goal was to compute bounds using the alphabets $\mathscr{A}_{M,k}$ which were constructed in the previous section. In these calculations we encountered the problem of accumulating rounding errors, disturbing the convergence. Therefore we had to normalize

the probability vector in each step. We did the rounding and normalization in such a way that our conclusions are not violated. For example to show that $\tau_\pi^n(p) < p_c^{site}$, we made sure that we computed an upper estimate for $\tau_\pi^n(p)$. Our implementation in Matlab (everything available from the author on request) gives the following results:

Proposition 5.2. *Take the alphabet $\mathscr{A}_{M,k}$ and define a classification $\mathscr{C}$ by the word code $\Phi_{M,k}$. Let $n = 1000$ and define $\tau_\pi^n(p)$ as before. Then*

- *For $M = 2$ and $k = 0$, we have $\tau_\pi^n(0.785) < p_c^{site}$.*
- *For $M = 2$ and $k = 1$, we have $\tau_\pi^n(0.859) < p_c^{site}$.*
- *For $M = 3$ and $k = 0$, we have $\tau_\pi^n(0.715) < p_c^{site}$.*

Corollary 5.2. $p_c(2) > 0.859$ *and* $p_c(3) > 0.715$.

Proof. This follows from Lemma 5.4 and Theorem 5.1. □

Figure 5.5 illustrates for the case $M = 2$ and $k = 0$ how $\tau_\pi^n(p)$ behaves as a function of n for some values of p. The values of $\tau_\pi^n(p)$ were calculated by iterating the recursion of Lemma 5.2.

For larger values of k the computations were too complicated to perform in a reasonable computation time. For example, the alphabet $\mathscr{A}_{2,2}$ already contains 35357670 letters. Nevertheless we will show that it is possible to improve the bounds of Corollary 5.2 by taking other alphabets or word codes.

For $M = 2$, define the element set E by dividing the left and right side of the unit square into 4 equal pieces and the bottom and top side into 2 equal pieces. This leads to an alphabet $\mathscr{A}_E$ that is in some sense in between $\mathscr{A}_{2,1}$ and $\mathscr{A}_{2,2}$. Analogous to our previous approach, we define a classification by choosing a word code based on weak connectivity. For this classification we find $\tau_\pi^{50}(0.876) < p_c^{site}$, which implies $p_c(2) > 0.876$.

One can improve this even a bit more by taking the alphabet $\mathscr{A}_{2,2}$ and defining

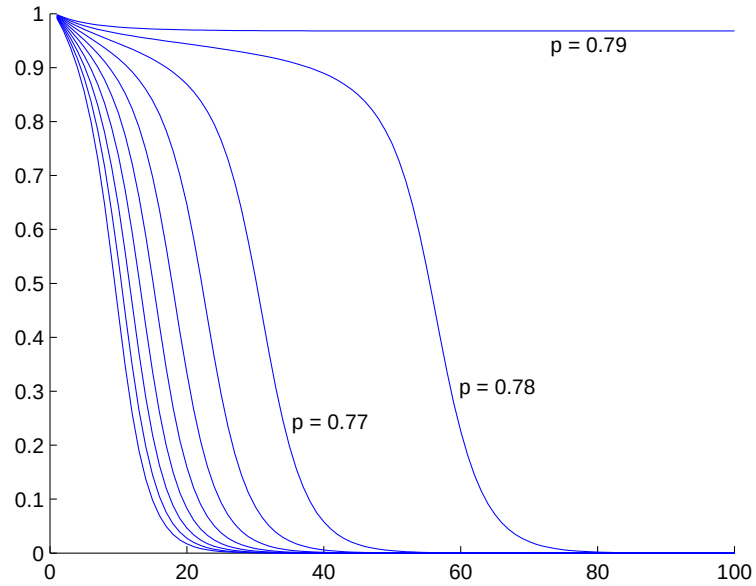

Figure 5.5: Plot of $\tau_\pi^n(p)$ for $p = 0.7 + 0.01k$ where $k = 0, \ldots, 9$ as functions of n. Note that there is a kind of phase transition between $p = 0.78$ and $p = 0.79$.

a word code $\tilde{\Phi}_{2,2}$ that is a bit simpler than $\Phi_{2,2}$ as follows. If at least one of the letters in a 2×2 word w equals $\min(\mathscr{A}_{2,2})$, then $\tilde{\Phi}_{2,2}(w) = \Phi_{2,2}(w)$. Otherwise, define $\tilde{\Phi}_{2,2}(w)$ by first mapping each of the four letters to $\mathscr{A}_E$ and then mapping the new word to $\mathscr{A}_{2,2}$, in both steps using weak connectivity. This simplifies the required calculations a lot, and leads to $p_c(2) > 0.881$, in 200 iterations.

For $M = 3$ we improved the lower bound by using an alphabet that divides the left and right side of the unit square into 3 equal pieces. The bottom and top side still consist of one element. This leads to $p_c(3) > 0.784$, using 100 iterations.

These calculations have been checked by Arthur Bik, a mathematics student at Delft University of Technology. He independently implemented the algorithms

and reproduced all results, except the bound $p_c(2) > 0.881$. This was due to the fact that his program was not fast enough to perform the calculations in a reasonable time. Concluding, the best lower bounds we found are

Theorem 5.3. $p_c(2) > 0.881$ *and* $p_c(3) > 0.784$.

Now let us turn to the upper bounds. Here we want to make use of Corollary 5.1. We already know $\tau^n_{\max(\mathscr{A})}(p) \leq \theta_n(p)$ for all n (Lemma 5.4) and that the classification is increasing. That means, we only have to find a probability vector $x \in [0,1]^{|\mathscr{A}|}$ for which $x_{\max(\mathscr{A})} > 0$ and $p \in (0,1]$ such that $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x$. The trick we use here is to take a value of p, and to search for the fixed point τ^∞ by iterating the recursion of Lemma 5.2. Assume that the numerical results suggest that $\tau^\infty_{\max(\mathscr{A})} > 0$. Letting $x = \tau^n(p)$ for n large, we have a combination of x and p for which $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \approx x$. Taking p slightly larger gives a combination for which the desired inequality $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x$ holds. This strategy leads to the following results:

Proposition 5.3. *Take the alphabet $\mathscr{A}_{M,k}$ and define a classification $\mathscr{C}$ by the word code $\Psi_{M,k}$. Let $n = 1000$. The conditions $F_{\mathscr{C}}(px + (1-p)\tau^{\square}) \succeq x$ and $x_{\max(\mathscr{A})} > 0$ hold if x and p are chosen as follows:*

- *For $M = 3$ and $k = 0$, choose $p = 0.958$ and $x = \tau^n(0.9579)$.*
- *For $M = 4$ and $k = 0$, choose $p = 0.972$ and $x = \tau^n(0.9719)$.*

Corollary 5.3. $p_c(3) < 0.958$ *and* $p_c(4) < 0.972$.

Proof. This follows from Lemma 5.4 and Corollary 5.1. □

For $M = 3$, the result can be sharpened by using the same alphabet that was used to improve the lower bound. The classification is again defined by a word code based on strong connectivity. In that case the choice $p = 0.940$ and $x = \tau^{1000}(0.9399)$ satisfies all conditions, so $p_c(3) < 0.940$.

The algorithm for $M = 4$ can be slightly adapted to find a bound for $M =$

2. Each realization of K_n for $M = 4$ can be seen as a realization of K_{2n} for $M = 2$. So the only thing that changes is the iteration function $F_{\mathscr{C}}$. Given a probability vector on $\mathscr{A}$, the probability on each 4×4 word can be calculated. The word code is still the same function $\Psi_{4,0}$. These ingredients are sufficient to determine $F_{\mathscr{C}}$. Performing the calculations we find that the conditions are satisfied for $p = 0.993$ and $x = \tau^{1000}(0.9929)$, henceforth $p_c(2) < 0.993$.

Also for the upper bounds Arthur Bik checked our results. He independently reproduced our bounds, except for the bound $p_c(3) < 0.940$ (for similar reasons as before). Summarizing, our best upper bounds are

Theorem 5.4. $p_c(2) < 0.993$, $p_c(3) < 0.940$ *and* $p_c(4) < 0.972$.

Summary

The Three Gap Theorem of Steinhaus states that given an irrational number α and a natural number n, the fractional parts

$$\{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{n\alpha\}$$

divide $[0, 1]$ into subintervals of at least two and at most three different lengths. In Chapter 2 we prove a variation stating that the partition of $[0, 1]$ defined by

$$\{\pm\alpha\}, \{\pm 2\alpha\}, \{\pm 3\alpha\}, \ldots, \{\pm n\alpha\}$$

also yields at most three different interval lengths. If there are three, then the longest one is the sum of the other two. From this theorem we then deduce an analogous "Four Gap Theorem" for the distance to the nearest integer, denoted by $||\cdot||$. Namely, this theorem states that the distances

$$||\alpha||, ||2\alpha||, ||3\alpha||, \ldots, ||n\alpha||$$

divide the interval $[0, 1/2]$ into subintervals of at least two and at most four different lengths. Four is the sharpest possible bound. In addition, other properties of the lengths are determined. For example, the rightmost length is unique, there are two different lengths if and only if $||\alpha|| < 1/2n$, and if there are three or four lengths, then all relations among them are found.

In Chapter 3, we consider the orbit of a point representing an idealized billiard ball which starts from a corner of a rectangular table. If one truncates the orbit when the ball is at some boundary point, then a partition of the rectangle into a finite number of polygons is obtained. The main result of this chapter is that, for a fixed initial shooting angle and boundary stopping point, the number of different shapes these polygons can have (up to translation, rotation or reflection) is at most 16, and the number of different areas is at most 13. An example is constructed which shows that these bounds are sharp.

In addition, we consider two special cases. If the shooting angle has a rational slope, then the orbit is periodic and eventually the partition into polygons remains fixed. This fixed partition can have at most 3 different areas, although the upper bound of 13 is still sharp before reaching this limiting partition. On the other hand, if the shooting slope α has the form $1/(n+\phi)$ where $\phi = (\sqrt{5}-1)/2$ is the small golden mean, then the total number of possible areas is at most 12.

Chapter 4 deals with algebraic differences of random Cantor sets. Let F_1 and F_2 be independent copies of one-dimensional correlated fractal percolation, with almost sure Hausdorff dimensions $\dim_{\mathrm{H}}(F_1)$ and $\dim_{\mathrm{H}}(F_2)$. We study the following question: does $\dim_{\mathrm{H}}(F_1) + \dim_{\mathrm{H}}(F_2) > 1$ imply that their algebraic difference $F_1 - F_2$ will contain an interval? The well known Palis conjecture states that 'generically' this should be true. Recent work by Dekking and Kuijvenhoven [8] on random Cantor sets can not answer this question as their condition on the joint survival distributions of the generating process is not satisfied by correlated fractal percolation. We develop a new condition which permits us to solve the problem, and we prove that the condition of [8] implies our condition. Independently of this we give a solution to the critical case, yielding that a strong version of the Palis conjecture holds for fractal percolation and correlated fractal percolation: the algebraic difference contains an interval almost surely if and only if the sum of the Hausdorff dimensions of the random Cantor sets exceeds one.

In Chapter 5, we investigate the critical probability $p_c(M)$ in two-dimensional

M-adic fractal percolation. To find lower bounds, we compare fractal percolation with site percolation. Fundamentally new is the construction of an increasing principally computable sequence that converges to $p_c(M)$. Unfortunately, these computations are too complicated to perform in reasonable time, so to obtain sharp numerical results we need some additional ideas.

We introduce the idea to classify realizations: they are mapped to a finite alphabet $\mathscr{A}$. This induces a sequence of probability measures on $\mathscr{A}$. We show that these probability measures converge to a fixed point. Choosing the alphabet and the classification rule in a proper way allows us to prove the following lower bounds for the critical probability: $p_c(2) > 0.881$ and $p_c(3) > 0.784$. For the upper bounds a similar approach gives $p_c(2) < 0.993$, $p_c(3) < 0.940$ and $p_c(4) < 0.972$.

Samenvatting

De Drie Gaten Stelling van Steinhaus zegt dat voor een gegeven irrationaal getal α en een natuurlijk getal n, de breukdelen

$$\{\alpha\}, \{2\alpha\}, \{3\alpha\}, \ldots, \{n\alpha\}$$

het interval $[0, 1]$ in deelintervallen met ten minste twee en ten hoogste drie verschillende lengtes verdelen. In Hoofdstuk 2 wordt een variant hierop bewezen, namelijk dat de partitie van $[0, 1]$ die gedefinieerd wordt door

$$\{\pm\alpha\}, \{\pm 2\alpha\}, \{\pm 3\alpha\}, \ldots, \{\pm n\alpha\}$$

eveneens ten hoogste drie verschillende intervallengtes geeft. Als er drie zijn, dan is de langste de som van de andere twee. Uit dit resultaat wordt een analoge "Vier Gaten Stelling" afgeleid voor de afstanden tot het dichtstbijzijnde gehele getal, genoteerd als $||\cdot||$. Deze stelling zegt dat de afstanden

$$||\alpha||, ||2\alpha||, ||3\alpha||, \ldots, ||n\alpha||$$

het interval $[0, 1/2]$ verdelen in subintervallen van ten minste twee en ten hoogste vier verschillende lengtes. Vier is de scherpst mogelijke bovengrens. Bovendien worden andere eigenschappen van de lengtes bewezen. De lengte van het meest rechtse interval is bijvoorbeeld uniek, en er zijn precies twee verschil-

lende lengtes dan en slechts dan als $||\alpha|| < 1/2n$. Voor de gevallen dat er drie of vier lengtes zijn worden alle mogelijk relaties tussen deze lengtes opgesomd.

In Hoofdstuk 3 bestuderen we de baan van een geïdealiseerde biljartbal die vertrekt vanuit een van de hoeken van een rechthoekige tafel. Door de baan te stoppen als de bal in een randpunt is, kan een verdeling van de rechthoek in een eindig aantal veelhoeken verkregen worden. Het belangrijkste resultaat in dit hoofdstuk is dat (voor een vaste lanceringshoek en een gegeven eindpunt) het aantal verschillende vormen van deze veelhoeken maximaal zestien is. Hierbij tellen translaties, rotaties en spiegelingen maar een keer mee. Het aantal mogelijke oppervlaktes van de veelhoeken is ten hoogste dertien. We construeren bovendien een voorbeeld dat aantoont dat deze bovengrenzen scherp zijn.

Verder worden twee bijzondere gevallen bekeken. Als de schiethoek rationaal is, dan is de baan periodiek en uiteindelijk krijgen we dan een partitie die niet meer verandert. Deze vaste partitie geeft maximaal drie verschillende oppervlaktes, hoewel de bovengrens van dertien nog steeds bereikt kan worden voordat het periodieke gedrag zich laat zien. Anderzijds zijn er ook irrationale getallen waarvoor de bovengrens nooit bereikt wordt. Als de schiethoek α van de vorm $1/(n+\phi)$ is, waarbij $\phi = (\sqrt{5}-1)/2$ de kleine gulden snede is, dan is het aantal verschillende oppervlaktes maximaal twaalf.

Hoofstuk 4 behandelt algebraïsche verschillen van stochastische Cantor verzamelingen. Stel dat F_1 en F_2 onafhankelijke realisaties van eendimensionale gecorreleerde fractale percolatie zijn, waarvan de bijna zekere Hausdorff dimensies gegeven worden door $\dim_{\mathrm{H}}(F_1)$ en $\dim_{\mathrm{H}}(F_2)$. Onze centrale vraag is dan: volgt uit $\dim_{\mathrm{H}}(F_1) + \dim_{\mathrm{H}}(F_2) > 1$ dat het algebraïsch verschil $F_1 - F_2$ een interval bevat? Het welbekende vermoeden van Palis zegt dat dit in het algemeen waar zou moeten zijn. Recent werk van Dekking en Kuijvenhoven [8] blijft het antwoord op deze vraag schuldig, aangezien aan hun voorwaarde op de gezamenlijke kansverdelingen van het genererende proces niet voldaan is in het geval van gecorreleerde fractale percolatie. In dit proefschrift wordt

een nieuwe voorwaarde ontwikkeld, die hier wel uitkomst biedt en waarvan we aantonen dat hij zwakker is dan de conditie in [8].

Onafhankelijk hiervan wordt het kritieke geval opgelost, waarmee we laten zien dat een sterke versie van het vermoeden van Palis geldt voor fractale percolatie en gecorreleerde fractale percolatie: het algebraïsch verschil bevat met kans 1 een interval dan en slechts dan als de som van de Hausdorff dimensies van de Cantor verzamelingen strikt groter is dan 1.

In Hoofdstuk 5 richten we onze aandacht op de kritieke kans $p_c(M)$ in tweedimensionale M-adische fractale percolatie. Om ondergrenzen te vinden, vergelijken we fractale percolatie met percolatie op de roosterpunten in het Euclidische vlak. We construeren voor alle M een stijgende rij die convergeert naar $p_c(M)$. De termen in deze rij zijn in principe algoritmisch te berekenen, maar aangezien dat zeer tijdrovend is leidt dit nog niet tot goede numerieke resultaten.

Om dit probleem te omzeilen gaan we realisaties classificeren: alle mogelijke realisaties worden afgebeeld naar een letter uit een eindig alfabet $\mathscr{A}$. De kansverdeling op mogelijke realisaties wordt hierdoor vereenvoudigd tot een kansverdeling op $\mathscr{A}$. We bewijzen dat deze kansverdelingen convergeren naar een dekpunt. Door nu het alfabet en de classificatieregel goed te kiezen kunnen we ondergrenzen voor de kritieke kans afleiden: $p_c(2) > 0.881$ and $p_c(3) > 0.784$. Een vergelijkbare aanpak voor de bovengrenzen geeft $p_c(2) < 0.993$, $p_c(3) < 0.940$ and $p_c(4) < 0.972$.

Bibliography

[1] Acquistapace, P.; Broglia, F. – *Some calculations about billiards*, (Italian) Archimede 36 (1984), no. 1-2, 72-88.

[2] Bálint, P.; Halász, M.; Hernández-Tahuilán, J. A.; Sanders, D. P. – *Chaos and stability in a two-parameter family of convex billiard tables*, Nonlinearity 24 (2011), no. 5, 1499-1521.

[3] Berg, J. van den and Ermakov, A. – *A New Lower Bound for the Critical Probability of Site Percolation on the Square Lattice*, Random Structures and Algorithms, Vol. 8, No. 3 (1996).

[4] Chayes, J.T., Chayes, L. and Durrett, R. – *Connectivity Properties of Mandelbrot's Percolation Process*, Probability Theory and Related Fields, 77 (1988), 307–324.

[5] Chernov, N.; Markarian, R. – *Chaotic billiards*, Mathematical Surveys and Monographs, 127. American Mathematical Society, Providence, RI, 2006. ISBN: 0-8218-4096-7.

[6] Dekking, F.M. and Don, H. – *Correlated fractal percolation and the Palis conjecture*, Journal of Statistical Physics. 139 (2010), no. 2, 307-325.

[7] Dekking, F.M. and Grimmett, G.R. – *Superbranching processes and projections of random Cantor sets*, Probability Theory and Related Fields, 78 (1988), 335–

355.

[8] Dekking, F.M. and Kuijvenhoven, B. – *Differences of random Cantor sets and lower spectral radii*, Journal of the European Mathematical Society (JEMS) 13 (2011), no. 3, 733-760.

[9] Dekking, F.M. and Meester, R.W.J. – *On the Structure of Mandelbrot's Percolation and Other Random Cantor Sets*, Journal of Statistical Physics, Vol. 58, Nos. 5/6 (1990), 1109–1126.

[10] Dekking, F.M. and Simon, K. – *On the size of the algebraic difference of two random Cantor sets*, Random Structures Algorithms **32** (2008), no. 2, 205–222.

[11] Don, H. – *On the distribution of the distances of multiples of an irrational number to the nearest integer*, Acta Arith. 139 (2009), 253–264.

[12] Don, H. – *Polygons in billiard orbits*, Journal of Number Theory 132 (2012), no. 6, 1151-1163.

[13] Don, H. – *New methods to determine the critical probability in fractal percolation*, submitted to Random Structures and Algorithms. (arXiv:1210.4150)

[14] Falconer, K. – *Random fractals*, Math. Proc. Cambr. Phil. Soc. 100 (1986), 559–582.

[15] Grimmett G.R. – *Percolation*, Second edition, Springer-Verlag, Berlin (1999).

[16] Komornik, V. – Personal communication.

[17] Larsson, P. – *L'ensemble différence de deux ensembles de Cantor aléatoires*, C. R. Acad. Sci. Paris Sr. I Math. 310 (1990), no. 10, 735-738.

[18] Larsson, P. – *The difference set of two Cantor sets* PhD Thesis, Uppsala Universitet, 1991.

[19] Lennes, N. J. – *On the Motion of a Ball on a Billiard Table*, Amer. Math. Monthly 12 (1905), no. 8-9, 152-155.

[20] Mandelbrot, B.B. – *Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier*, Journal of Fluid Mechanics **62** (1974), part2, 331–358.

[21] Mandelbrot, B.B. – *The fractal geometry of nature*, W. H. Freeman and Co.,San Francisco, Calif., (1982),v+460, 0-7167-1186-9, MR665254.

[22] Orzechowski, M. – *Geometrical and topological properties of fractal percolation*, PhD Thesis, University of St. Andrews, 1997.

[23] Palis, J. and Takens, F. – *Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations*, volume 35 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1993. Fractal dimensions and infinitely many attractors.

[24] van Ravenstein, T. – *The three gap theorem (Steinhaus conjecture)*, J. Austral. Math. Soc. Ser. A 45 (1988), no. 3, 360–370.

[25] Slater, N.B. – *Gaps and steps for the sequence $n\theta$ mod 1*, Proc. Cambridge Philos. Soc. 73 (1967), 1115–1122.

[26] Sós, V.T. – *On the distribution mod 1 of the sequence $n\alpha$*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 1 (1958), 127–134.

[27] Sukop, M.C.; Perfect, E. and Bird, N.R.A. – *Water retention of prefractal porous media generated with the homogeneous and heterogeneous algorithms*, Water Resources Research **32** (2001), no. 10, 2631–2636.

[28] Surányi, J. – *Über die Anordnung der Vielfachen einer reelen Zahl mod 1*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 1 (1958), 107–111.

[29] Świerckowski, S. – *On successive settings of an arc on the circumference of a circle*, Fund. Math. 46 (1958), 187–189.

[30] Wal, P. van der – *Random substitutions and fractal percolation*, PhD thesis, Delft University of Technology, 2002.

[31] White, D.G. – *On the Value of the Critical Point in Fractal Percolation*, Random

Structures and Algorithms, 18 (2001), 332–345.

Dankwoord

Verscheidene personen zijn mij bij de vervaardiging van dit proefschrift zeer behulpzaam geweest. In de eerste plaats gaat mijn dank uit naar mijn promotor, Michel Dekking, die mij de mogelijkheid geboden heeft dit onderzoek onder zijn supervisie te verrichten. Hij heeft mij veel fraaie wiskunde laten zien, en tegelijkertijd liet hij mij de vrijheid zelf op ontdekking te gaan. Zijn begeleiding bij het onderzoek was van grote waarde. Een woord van dank verdient ook de copromotor, Cor Kraaikamp, die bij het meer getaltheoretische deel van het onderzoek over mijn schouder meekeek. In de afgelopen jaren hebben we vele leerzame en onderhoudende discussies gevoerd over verscheidene onderwerpen.

Voor het oplossen van ondoorgrondelijke computerkuren was de helpende hand van Carl altijd afdoende. Alle andere problemen van de meest uiteenlopende aard werden door de hulpvaardigheid van Cindy snel de wereld uit geholpen. Alle andere werknemers van de stochastiekgroep wil ik eveneens graag danken voor hun collegialiteit.

Ook alle mede-promovendi betuig ik mijn welgemeende dank. Om de drukkosten te drukken noem ik slechts mijn drie huidige kamergenoten bij naam: Christos, Moritz en Richard. De vele gedachtewisselingen en gezamenlijke wandelingen naar de koffiemachine gaven kleur aan mijn kantooruren als promovendus.

Graag maak ik gewag van mijn dankbaarheid voor de niet aflatende blijken van belangstelling van familie en vrienden voor mijn vaak raadselachtige en onbevattelijke bezigheden. In het bijzonder dank ik mijn ouders die onophoudelijk mijn arbeidzaamheid aanwakkerden. Ook Catharina ben ik veel dank verschuldigd voor de steun en bemoediging bij mijn onderzoek en voor het geduldig dragen van de ongerieflijkheden die horen bij het leven met een wiskundige onder hetzelfde dak. Bert en Ernst, jullie ongelooflijke vrolijkheid en energie was me heel veel waard.

Henk Don
Renswoude, januari 2013

Curriculum Vitae

Hendrik Don werd geboren op 8 april 1984 te Rhenen. In 2002 behaalde hij zijn gymnasiumdiploma aan het Van Lodenstein College te Amersfoort. Zijn loopbaan werd vervolgd met een studie wiskunde aan de Technische Universiteit Delft. In januari 2009 studeerde hij cum laude af onder begeleiding van Prof. dr. F.M. Dekking.

Op 1 maart 2009 ging hij aan de slag als promovendus in de kansrekening, onder supervisie van Prof. dr. F.M. Dekking en Dr. C. Kraaikamp. Het onderzoek dat hij in deze periode verrichtte heeft geleid tot dit proefschrift.