



Masterthesis

# Incremental Aggregations

Tillmann Tristan Bosch

9. March 2021

Supervisor: PD. Dr. Steffen Winter

Faculty of Mathematics

Karlsruhe Institute of Technology



# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Preliminaries</b>	<b>5</b>
2.1	Symbols . . . . .	5
2.2	Expressions . . . . .	5
2.3	Graphs . . . . .	5
2.4	Random walks . . . . .	6
<b>3</b>	<b>Incremental aggregation</b>	<b>8</b>
3.1	Definition . . . . .	8
3.2	Notion of fractal dimension and growth rate . . . . .	9
3.3	A rigorous characterization for the fractal dimension . . . . .	13
<b>4</b>	<b>External diffusion limited aggregation</b>	<b>22</b>
4.1	Definition . . . . .	22
4.2	Fractal dimension and growth rate of external DLA in $\mathbb{Z}^2$ . . . . .	22
<b>5</b>	<b>Line hitting aggregation</b>	<b>24</b>
5.1	Motivation . . . . .	24
5.2	Excursion: Integral geometry . . . . .	27
5.2.1	General results . . . . .	28
5.2.2	Construction in the plane: Isotropic lines . . . . .	28
5.3	Definition . . . . .	35
5.4	Fractal dimension and growth rate of LHA in $\mathbb{Z}^2$ . . . . .	37
<b>6</b>	<b>Python simulations</b>	<b>38</b>
6.1	Code . . . . .	38
6.2	Empirical values of LHA . . . . .	42
6.3	Images . . . . .	47
<b>7</b>	<b>Review</b>	<b>49</b>
<b>8</b>	<b>Declaration</b>	<b>51</b>

# 1 Introduction

In mathematics and physics there are objects which may naturally awaken the interest of any human being, mathematician and non-mathematician, just by their natural visual appearance - and clusterings of particles or clusterlike occurrences like snowflakes or broken glasses are one example of them. Looking at Figure 1 we find a clusterlike formed occurrence in the radio screen of a car which might have arisen by the breaking of a crystal fluid in the screen. Other examples are cracking of solid glass (Figure 2 (b)), processes like bacterial colonies or electrodeposition and also more symmetrical examples as the formation of snowflake crystals (Figure 2 (a)).

One way of describing such processes is through clusters which get formed by random point by point aggregation. If for some natural number  $d \in \mathbb{N}$  we draw our attention to the  $d$ -dimensional lattice  $\mathbb{Z}^d$  and consider one particle in the origin  $(0, 0)$ , we can let a cluster grow by adding new particles one by one, always on the boundary of the current cluster. What we get is a random growth of a connected clusterlike set of which we then can think about its fractal dimension, how quick it grows and many other properties, and produce beautiful realizations.

In this paper we will look at two ways of letting grow such clusters and simulate them with python at the end to produce visuals and some empirical values. We will call such point by point growing clusters *incremental aggregations* and present the two models briefly here:

**External Diffusion Limited Aggregation** was initially introduced by [1, Witten and Sander, 1983]. In this model of incremental aggregation of particles, each new particle starts a random walk from infinity and when it hits the current cluster for the first time, it sticks at the place where it hit the cluster. Despite this simple definition there are only few rigorously proved results on this process. In  $d$  dimensions the fractal dimension of any incremental aggregation lies in between 1 and  $d$  and here we prove an almost sure lower bound of 1.5 in  $\mathbb{Z}^2$ . Empirical studies calculate a value around 1.7 to 1.8 in  $\mathbb{Z}^2$  (see for example [5]).

**Line Hitting Aggregation** is a ballistic model similar to the attempts of [4], but is introduced in its idea here by the author of this paper. In this model new particles do not approach the current cluster on a random walk but on straight lines which are chosen randomly before. The difficulty here is to find a fair measure on choosing random lines which hit the current cluster, where we base on already existing results from integral theory. So one way of understanding this process is by shooting new particles from randomly, fairly chosen directions to the cluster and let them stick where they hit the cluster the first time. It seems that such formed clusters are denser than external DLA and therefore have a higher fractal dimension which we empirically find to be true in the last chapter of this paper. In [4] they guess a fractal dimension of 2 for their ballistic model and here we will actually be able to prove this result for the ballistic model we

## *1 Introduction*

define in this paper.

We will first draw our attention to the general definition of incremental aggregations and a notion of their fractal dimension. We will then present a general relationship between the distribution family of an incremental aggregation and a lower bound for its fractal dimension. We will then turn our attention to the two models we discuss in this paper, first external DLA and the line hitting aggregation in the chapter after. With the help of the general theorem for incremental aggregations we will prove an lower bound for external DLA and even an exact result for line hitting aggregation. At the end we will present simulations for both aggregations of which only the one for the line hitting aggregation shall satisfy the claim of being implemented as close to its mathematical definition as possible. At the end we especially hope the reader to relate with the processes and enjoy the created images.

## 1 Introduction

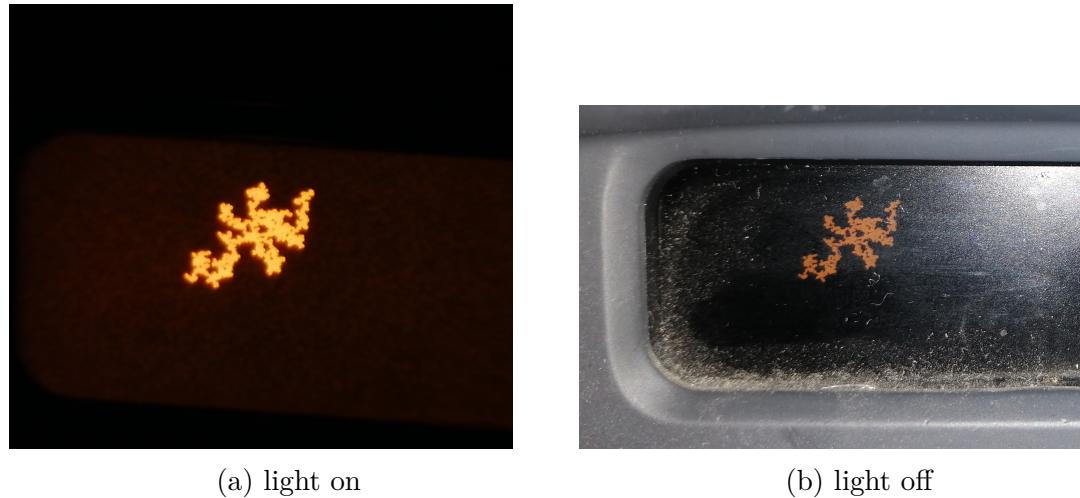


Figure 1: Cluster appearance in the radio screen of a car

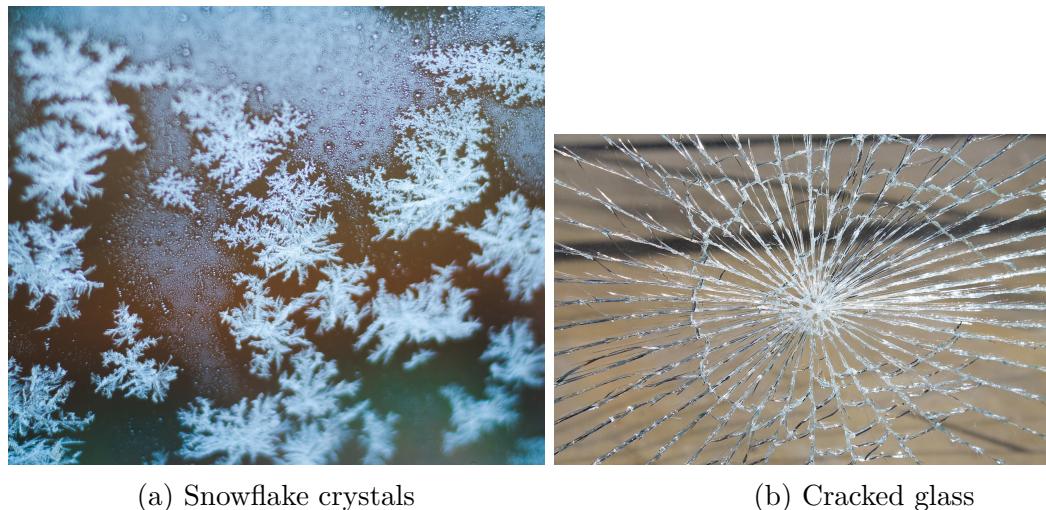


Figure 2: More examples of clusterlike formations

## 2 Preliminaries

### 2.1 Symbols

Let  $d \in \mathbb{N}$  and  $q \in \{0, \dots, d\}$ .

$\mathbb{N} = \{1, 2, 3, \dots\}$ , the set of natural numbers (without 0)

$\mathbb{N}_0 = \mathbb{N} \cup \{0\}$

$\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$

$\mathcal{B}^d$ ,  $d$ -dimensional Borel- $\sigma$ -algebra of  $\mathbb{R}^d$

$\mathcal{K}^d$ , the set of convex and compact sets in  $\mathbb{R}^d$

$B_d(x, r) = \{y \in \mathbb{R}^d \mid |x - y| \leq r\}$ , the  $d$ -dimensional closed ball of radius  $r$  around  $x$

$B_r := B_2(0, r)$

$S_{d-1}(x, r) = \partial B_d(x, r)$ , the  $(d-1)$ -dimensional surface of the  $d$ -dimensional ball

$A(d, q)$ , the set of  $q$ -dimensional affine subspaces of  $\mathbb{R}^d$

$\mathcal{A}(d, q)$ , the  $\sigma$ -algebra of  $A(d, q)$ , as constructed later in the paper

$\mathcal{G} := A(2, 1)$ , the set of lines in the real plane

$SO_d := \{\nu \in \mathbb{R}^{d \times d} \mid \nu\nu^\top = I_d \text{ and } \det \nu = 1\}$ , identify  $SO_2 = \{\nu_\beta := e^{i\beta} \in \mathbb{C} \mid \beta \in [0, 2\pi)\}$

$G_d := \{\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d, x \mapsto \nu x + b \mid \nu \in SO_d, b \in \mathbb{R}^d\}$ , the set of Euclidean motions

$\mathcal{P}_f^d$ , the set of finite subsets of  $\mathbb{Z}^d$

$\mathcal{P}_f := \mathcal{P}_f^2$

Note that we will always identify  $\mathbb{R}^2$  with  $\mathbb{C}$  ( $(a, b) \leftrightarrow a + bi$ ) for a more convenient notation.

### 2.2 Expressions

Throughout the paper let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. If for  $A \in \mathcal{F}$  we have  $\mathbb{P}(A) = 1$  we will say that „ $A$  holds  $\mathbb{P}$ -a.s.“, or short „ $A$  holds a.s.“ ( $A$  holds almost surely). Another short expression we will use is that if for a set of logical statements  $(A_t)_{t \in I}$  with  $I \in \{\mathbb{N}, \mathbb{R}\}$  we say „ $A_t$  holds for large  $t$ “ it shall mean that there exists a  $T \in I$  such that  $A_t$  holds for all  $t > T$ . Here we mean that a logical statement holds if and only if the statement is true.

### 2.3 Graphs

Let  $d \in \mathbb{N}$ . We will be interested in the undirected graph  $(\mathbb{Z}^d, E)$  with its canonical graph structure, which is two vertices (or points)  $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathbb{Z}^d$  form an edge (e.g.  $\{x, y\} \in E$ ) if and only if there exists exactly one  $i \in \{1, \dots, d\}$  such that  $|x_i - y_i| = 1$  and  $x_j = y_j$  for all  $j \neq i$ . For a point  $x \in \mathbb{Z}^d$  its set of *neighbours* is

defined as

$$N(x) := \{y \in \mathbb{Z}^d \mid \{x, y\} \in E\}.$$

For a set  $A \subset \mathbb{Z}^d$  the *outer boundary*  $\partial A$  of  $A$  is defined as

$$\partial A := \{y \in \mathbb{Z}^d \setminus A \mid \exists x \in A : \{x, y\} \in E\}$$

and the closure  $\bar{A}$  of  $A$  as

$$\bar{A} := A \cup \partial A.$$

Instead of  $(\mathbb{Z}^d, E)$  we will write  $\mathbb{Z}^d$  from now on.

## 2.4 Random walks

**Definition 2.4.1.** For our space of interest  $\mathbb{Z}^d$  we will always use the discrete  $\sigma$ -algebra which is the power set of  $\mathbb{Z}^d$ . A family  $(S_n)_{n \in \mathbb{N}_0}$  of measurable functions  $S_n : \Omega \rightarrow \mathbb{Z}^d$  is called a *random walk on  $\mathbb{Z}^d$*  (*starting at  $x \in \mathbb{Z}^d$* ) if and only if  $S_0 = x$  a.s. and

$$\mathbb{P}(S_n = y \mid S_{n-1} = z) = \frac{1}{|N(z)|} = \frac{1}{2d}, \quad \text{for all } y \in N(z) \text{ and } z \in \mathbb{Z}^d.$$

Note that  $|N(z)| = 2d$  for all  $z \in \mathbb{Z}^d$  since every point has two neighbours in the direction of every dimensional component. We can therefore conclude easily that  $\mathbb{P}(S_n = y \mid S_{n-1} = z) = 0$  for all  $y \notin N(z)$  and  $z \in \mathbb{Z}^d$ . For  $x, y \in \mathbb{Z}^d$  we introduce the short notation

$$\mathbb{P}_x(S_n = y) := \mathbb{P}(S_n = y \mid S_0 = x).$$

So a random walk can be understood as a particle starting from some point  $x$  and moving randomly on the grid choosing its next step uniformly from its neighbours. For the following let  $(S_n)_{n \in \mathbb{N}}$  be a random walk on  $\mathbb{Z}^d$  starting at  $x \in \mathbb{Z}^d$ .

**Definition 2.4.2.** Let  $A \subset \mathbb{Z}^d$ . We define the *hitting times* of  $A$  by

$$T_A := \min\{n \geq 0 \mid S_n \in A\} \text{ and } T_A^+ := \min\{n \geq 1 \mid S_n \in A\},$$

and  $T_y := T_{\{y\}}$  and  $T_y^+ := T_{\{y\}}^+$  for  $y \in \mathbb{Z}^d$ .

**Definition 2.4.3.** A random walk with origin in  $x \in \mathbb{Z}^d$  is called *recurrent* if

$$\mathbb{P}_x(T_x^+ < \infty) = 1$$

and *transient* if

$$\mathbb{P}_x(T_x^+ < \infty) < 1.$$

## 2 Preliminaries

**Lemma 2.4.1.** A random walk on  $\mathbb{Z}^d$  is recurrent if  $d \leq 2$  and transient if  $d \geq 3$ .

*Proof.* Proofs of this result are presented in [9, Satz 5.1] or [8, Korollar 2.6.6].  $\square$

**Lemma 2.4.2.** The following two statements are equivalent:

- (i)  $\mathbb{P}_x(T_x^+ < \infty) = 1$  for all  $x \in \mathbb{Z}^2$
- (ii)  $\mathbb{P}_x(T_A^+ < \infty) = 1$  for all  $x \in \mathbb{Z}^2$  and  $A \subset \mathbb{Z}^2$

*Proof.* The direction (ii) to (i) is clear. For the other direction choose  $x \in \mathbb{Z}^2$  and  $A \subset \mathbb{Z}^2$ . We know that in general for any point  $y \in \mathbb{Z}^2$  we have  $\mathbb{P}_x(T_y^+ < \infty) > 0$  for a random walk in  $\mathbb{Z}^d$  for any dimension  $d \in \mathbb{N}$ . By Lemma 2.4.1 we know that a random walk on  $\mathbb{Z}^2$  is recurrent. For that case it is proved in [8, Satz 2.6.9] that for  $y \in \mathbb{Z}^2$  even  $\mathbb{P}_x(T_y^+ < \infty) = 1$  holds. If we choose  $y \in A$  we get

$$1 = \mathbb{P}_x(T_y^+ < \infty) \leq \mathbb{P}_x(T_A^+ < \infty),$$

which completes the proof.  $\square$

After having introduced some basics we will turn our focus in the next chapter to the main definition of stochastic processes we will discuss in this paper.

## 3 Incremental aggregation

### 3.1 Definition

In this paper we will look at stochastic processes on the set of finite subsets of  $\mathbb{Z}^d$ , where we start with the one point set  $\{0\}$  and incrementally add a point of the boundary of the current cluster according to some distribution. Note that with 0 with will denote the zero vector. What we get is a randomly point by point growing connected cluster which we will call *incremental aggregation*. Define

$$\mathcal{P}_f^d := \{A \subset \mathbb{Z}^d \mid A \text{ is finite}\}$$

the set of finite subsets of  $\mathbb{Z}^d$ . Furthermore we will be interested in distributions on those sets, so for  $A \in \mathcal{P}_f^d$  we define

$$\mathcal{D}_A := \{\mu : \mathbb{Z}^d \rightarrow [0, 1] \mid \mu(y) = 0 \text{ for all } y \notin A \text{ and } \sum_{y \in A} \mu(y) = 1\}$$

the set of distributions on  $A$ . For  $\mathcal{P}_f^d$  we consider the discrete  $\sigma$ -algebra and call a measurable function

$$F : \Omega \rightarrow \mathcal{P}_f^d$$

a *random finite set*. Further for  $A \in \mathcal{P}_f^d$  we call a measurable function

$$y_A : \Omega \rightarrow \mathbb{Z}^d$$

a *random point in  $A$*  if  $\mathbb{P}(y_A = y) = 0$  is satisfied for all  $y \notin A$ . Now we define incremental aggregation as follows.

**Definition 3.1.1.** Let  $\mu = (\mu_A)_{A \in \mathcal{P}_f^d}$  be a family of distributions, therefore  $\mu_A \in \mathcal{D}_A$  for all  $A \in \mathcal{P}_f^d$ . An *incremental aggregation (with distribution  $\mu$ )* is a family of random finite sets  $(\mathcal{E}_n)_{n \in \mathbb{N}}$  which evolves as follows. The process starts with one point  $\mathcal{E}_1 = \{0\}$  (define  $y_1 := 0$ ) at the origin of  $\mathbb{Z}^d$ . Knowing the process  $\mathcal{E}_n$  at time  $n$ , let  $y_{n+1}$  be a random point in  $\partial\mathcal{E}_n \in \mathcal{P}_f^d$  with distribution

$$\mathbb{P}(y_{n+1} = y \mid \mathcal{E}_n) := \mu_{\partial\mathcal{E}_n}(y) \quad \text{for } y \in \mathbb{Z}^d. \quad (3.1)$$

We then define  $\mathcal{E}_{n+1} := \mathcal{E}_n \cup \{y_{n+1}\}$  and the limit cluster as  $\mathcal{E}_\infty := \bigcup_{n \in \mathbb{N}} \mathcal{E}_n$ . This naturally defines a Markov chain.

**Remark 3.1.1.** We want to quickly argument here that all  $\mathcal{E}_n$  are indeed measurable functions.  $\mathcal{E}_1 = \{0\}$  is constant and therefore measurable. It is not difficult to show that for  $A \in \mathcal{P}_f^d$  we have

$$\mathcal{E}_{n+1}^{-1}(A) = \bigcup_{y \in A} \mathcal{E}_n^{-1}(A \setminus \{y\}) \cap y_{n+1}^{-1}(\{y\})$$

for any  $n \in \mathbb{N}$  and we can therefore conclude the measurability for all  $\mathcal{E}_n$  by induction.

### 3 Incremental aggregation

**Remark 3.1.2.** It is worth to mention that since incremental aggregations are Markov chains the distribution of a new added point  $y_{n+1}$  is not depending on the order of how previous points where added to the cluster.

Properties of an incremental aggregation obviously base strongly on the properties of its distribution  $\mu$ . In the following we prepare some additional definitions which we will use later on.

**Definition 3.1.2.** We call a distribution family  $(\mu_A)_{A \in \mathcal{P}_f^d}$  an *outside distribution* if and only if for all connected sets  $A \in \mathcal{P}_f^d$  we have that

$$\mu_{\bar{A}}(x) = \begin{cases} \mu_{\partial A}(x), & x \in \partial A, \\ 0, & x \in A. \end{cases}$$

Recall that  $\bar{A} = A \cup \partial A$ . This definition shall capture the idea that particles coming from „outside“ to a set  $\bar{A}$  shall not be able to move into the „inside“  $A$  of that set without passing its boundary  $\partial A$  first. For both incremental aggregations we discuss in this paper this will be the case.

**Definition 3.1.3.** For  $y \in \mathbb{Z}^d$  let

$$\Phi_y : \mathbb{Z}^d \rightarrow \mathbb{Z}^d, x \mapsto x + y$$

be a translation function on  $\mathbb{Z}^d$ . Then we call the family of distributions  $(\mu_A)_{A \in \mathcal{P}_f^d}$  *translation invariant* if and only if for all  $y \in \mathbb{Z}^d$

$$\mu_A(z) = \mu_{\Phi_y(A)}(\Phi_y(z))$$

for all  $z \in A$  and  $A \in \mathcal{P}_f^d$ .

## 3.2 Notion of fractal dimension and growth rate

The notion of fractal dimension is usually used for sets with uncountable cardinality like continuous curves or surfaces. One way of defining a fractal dimension for curves like for example the coast line of New Zealand (see Figure 3) would be to look at the asymptotic relation between the minimal number of grid boxes (squares) which we need to cover the cost line and the side length of these boxes. The following definition is motivated by [2, Page 160]. If, for  $\varepsilon > 0$ ,  $N(\varepsilon)$  is the minimum number of boxes with side length  $\varepsilon$  which we need to cover the coast line, then the so called *box dimension*  $d_b$  is roughly the constant such that  $N(\varepsilon)$  grows like  $\varepsilon^{-d_b}$  as  $\varepsilon$  tends to zero, so

$$d_b := -\liminf_{\varepsilon \rightarrow 0} \frac{\ln(N(\varepsilon))}{\ln(\varepsilon)}. \quad (3.2)$$

This definition makes sense in many contexts, for example the box-dimension of straight line segments is 1 and of squares is 2 and so on, so in those cases equal to the topological

### 3 Incremental aggregation

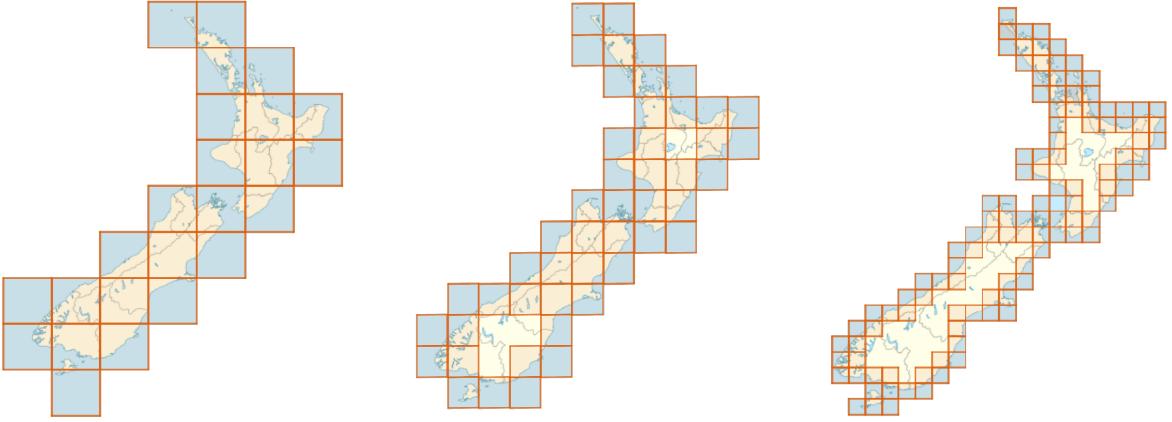


Figure 3: Box-covering of New Zealand's outer cost with decreasing box sizes  $\varepsilon$

dimension. As with incremental aggregations we are dealing with locally finite point sets, this approach of defining a fractal dimension for our clusters is not sensible. It is not difficult to show, that the box-dimension of any finite set is 0. A helpful detail about the situation with finite sets as the ones we are looking at is that each point of the cluster can actually be interpreted and identified with a unique box of side length 1 since the cluster is living on the grid  $\mathbb{Z}^d$ . We will precise that in Chapter 5. So instead of decreasing the sizes of the boxes which we cover our set of interest with, we leave the size of the boxes constant and increase the size of our set by adding points and looking at the limit cluster  $\mathcal{E}_\infty$ . Defining the radius of a finite set  $A \in \mathcal{P}_f^d$  (with  $|\cdot|$  the Euclidean norm) by

$$\text{rad}(A) := \max_{x \in A} |x|, \quad (3.3)$$

we can identify the relation between the geometrical sizes of the boxes and the cluster as follows. For  $n \in \mathbb{N}$  define random variables  $\varepsilon_n := \frac{1}{\text{rad}(\mathcal{E}_n)}$ . Before  $\varepsilon_n$  would have expressed the geometrical relation between one box and the whole cluster which we describe now by the fraction  $\frac{1}{\text{rad}(\mathcal{E}_n)}$  since a box size is now constantly 1 and the clusters geometrical size we can interpret by  $\text{rad}(\mathcal{E}_n)$ . The number of boxes with size 1 which we need to cover  $\mathcal{E}_n$  is always  $n$ , so  $N(\varepsilon_n) = n$  for all  $n \in \mathbb{N}$  and therefore we can rewrite the definition of the box-dimension (3.2) by replacing  $\varepsilon_n$  with  $\frac{1}{\text{rad}(\mathcal{E}_n)}$  and define the (*discrete*) *fractal dimension* of  $\mathcal{E}_\infty$  as the random variable

$$d_f := \liminf_{n \rightarrow \infty} \frac{\ln(n)}{\ln(\text{rad}(\mathcal{E}_n))}. \quad (3.4)$$

This way of defining a fractal dimension for incremental aggregations strongly correlates with the growth rate of the aggregation which shall indicate how the radius of the cluster evolves while increasing the particle number. We can define the growth rate by looking

### 3 Incremental aggregation

for the smallest exponent  $\alpha$  such that there exists a constant  $c > 0$  with

$$\text{rad}(\mathcal{E}_n) \leq cn^\alpha$$

for large  $n$ . Rewriting this we come to the equivalent inequality

$$\frac{\ln(\text{rad}(\mathcal{E}_n))}{\ln(n)} - \frac{\ln(c)}{\ln(n)} \leq \alpha$$

for large  $n$  and we could finally define the growth rate  $\alpha_f$  of an incremental aggregation as the smallest value satisfying this inequality, so the random variable

$$\alpha_f := \limsup_{n \rightarrow \infty} \frac{\ln(\text{rad}(\mathcal{E}_n))}{\ln(n)}. \quad (3.5)$$

We therefore simply get that

$$d_f = \frac{1}{\alpha_f} \quad \text{a.s.}, \quad (3.6)$$

which includes the cases  $d_f = \infty$  or  $\alpha_f = \infty$  ( $\frac{1}{0} := \infty$ ), which motivates choosing the limes inferior in the Definition (3.4). Since  $\text{rad}(\mathcal{E}_n) \leq n$  a.s. for all  $n \in \mathbb{N}$ , we get

$$\alpha_f \leq 1 \quad \text{a.s.}$$

and therefore

$$d_f \geq 1 \quad \text{a.s.}$$

In  $\mathbb{Z}^d$  we can fill the cube with center 0 and side length  $n^{\frac{1}{d}}$  with an order of  $n$  points. Therefore we can fill a ball with radius  $n^{\frac{1}{d}}$  with an order of  $n$  points as well, and since  $|\mathcal{E}_n| = n$  there exists a constant  $c > 0$  such that  $cn^{\frac{1}{d}} \leq \text{rad}(\mathcal{E}_n)$  a.s. for all  $n \in \mathbb{N}$ . Hence

$$d_f \leq \liminf_{n \rightarrow \infty} \frac{\ln(n)}{\ln(cn^{\frac{1}{d}})} = d \quad \text{a.s..}$$

So in total we get the trivial bounds

$$1 \leq d_f \leq d \quad \text{a.s.} \quad (3.7)$$

and

$$\frac{1}{d} \leq \alpha_f \leq 1 \quad \text{a.s.}$$

for any incremental aggregation in  $\mathbb{Z}^d$ .

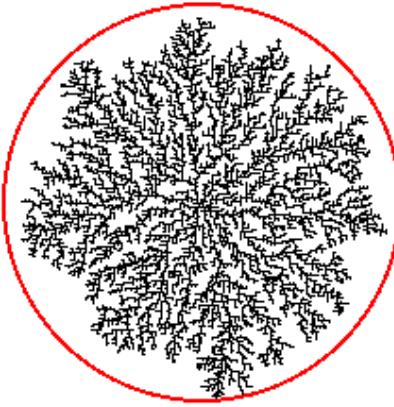


Figure 4: Cluster not „filled up“ to its radius

**Remark 3.2.1.** Another way of tackling the intuition for a fractal dimension for locally finite sets is the following which is motivated by [7, Part II, Section 1, Page 98] and [3, Page 83]. Let  $M \subset \mathbb{R}^d$  be a locally finite set. Let  $n \in \mathbb{N}$  and  $B_n$  be the ball with radius  $n$  and center 0. We could interpret the fractal dimension  $d_M$  of  $M$  as a value such that the cardinality  $|M \cap B_n|$  of  $M$  inside the ball  $B_n$  is roughly of order  $n^{d_M}$  for large  $n$ , so

$$d_M := \liminf_{n \rightarrow \infty} \frac{\ln(|M \cap B_n|)}{\ln(n)}. \quad (3.8)$$

If we apply that definition to our limit cluster  $\mathcal{E}_\infty$  it is not clear whether this definition is equivalent to (3.4). If we define the fractal dimension as in (3.4) we encounter the problem that when the cluster reaches some radius for the first time, then in many instances in the subsequent steps there will particles be added to the cluster without increasing its radius, so to say „filling“ the cluster up to that radius (see Figure 4). If the fractal dimension shall indicate an asymptotic relation between how many particles of the limit cluster  $\mathcal{E}_\infty$  lie in a ball of a radius letting the radius tend to infinity, the definition in (3.4) does not reflect that directly. Nevertheless it seems to make sense that these two situations are equivalent. If the distribution family  $\mu$  doesn't favorize outer lying particles for the next aggregation too much, creating long arms and therefore magnifying the problem we are talking about here is rather unprobable. Compared to the size of the cluster, arms will rather decrease in size which lets diminish the difference of Definitions (3.4) and (3.8) which we state more clearly in the following conjecture.

**Conjecture 3.2.1.** Let  $B_n$  be a ball of radius  $n \in \mathbb{N}$  with center 0 and with  $\mathcal{E}_\infty \cap B_n$  denote the set of particles in  $\mathcal{E}_\infty$  that lie in  $B_n$ . Then we have

$$d_f = \liminf_{n \rightarrow \infty} \frac{\ln(|\mathcal{E}_\infty \cap B_n|)}{\ln(n)} \quad \text{a.s.}.$$

### 3.3 A rigorous characterization for the fractal dimension

In this section we will proof an almost sure lower bound for the fractal dimension of an incremental aggregation if its distribution family  $(\mu_A)_{A \in \mathcal{P}_f^d}$  satisfies some conditions. Before coming to this statement we have to make some preparations in the following. We define two random functions

$$r : \mathbb{N} \rightarrow [0, \infty), \quad n \mapsto \text{rad}(\mathcal{E}_n)$$

and

$$T : [0, \infty) \rightarrow \mathbb{N}, \quad s \mapsto \min\{j \in \mathbb{N} \mid r(j) \geq s\}.$$

Recall that  $\text{rad}(\mathcal{E}_n)$  is the radius of the cluster at time  $n$  as defined in (3.3).

**Remark 3.3.1.** It is easy to show that for all  $\omega \in \Omega$  the functions  $r(\omega)$  and  $T(\omega)$  are increasing in  $n \in \mathbb{N}$  resp.  $s \in [0, \infty)$  and that for all  $n \in \mathbb{N}$  and  $s \in [0, \infty)$  we have  $T(\omega)(r(\omega)(n)) \leq n$  and  $r(\omega)(T(\omega)(s)) \geq s$ .

**Lemma 3.3.1.** For both random functions  $r$  and  $T$  we have that

$$r(\omega)(n) \rightarrow \infty \text{ for } n \rightarrow \infty$$

and

$$T(\omega)(s) \rightarrow \infty \text{ for } s \rightarrow \infty$$

for all  $\omega \in \Omega$ .

*Proof.* Since we are moving on the grid  $\mathbb{Z}^d$  we have that for a ball  $B_d(0, n)$  with radius  $n \geq 0$  the number  $N := |B_d(0, n) \cap \mathbb{Z}^d|$  is finite and for any  $\omega \in \Omega$  we get

$$r(\omega)(2N) \geq n.$$

Therefore for all  $\omega \in \Omega$  and any  $n \in \mathbb{N}$  we can find a  $M \in \mathbb{N}$  such that  $r(\omega)(M) \geq n$ , and since  $r(\omega)$  is increasing we get  $r(\omega)(n) \rightarrow \infty$  for  $n \rightarrow \infty$  for all  $\omega \in \Omega$ . Very similarly we can argue for  $T$ .  $\square$

**Lemma 3.3.2.** Let  $a > 0$  and  $h : [0, \infty) \rightarrow [0, \infty)$  be a bijective, multiplicative and increasing function. For  $c > 0$  define

$$A_c := \{\omega \in \Omega \mid \exists N = N(\omega) \in \mathbb{N} : r(\omega)(n) \leq ch(n) \text{ for all } n > N\}$$

and

$$D_c := \{\omega \in \Omega \mid \exists N = N(\omega) \in \mathbb{N} : T(\omega)(as) \geq ch^{-1}(s) \text{ for all } s > N\}.$$

Then the following are equivalent:

### 3 Incremental aggregation

$$(i) \exists c > 0 : \mathbb{P}(A_c) = 1$$

$$(ii) \exists c > 0 : \mathbb{P}(D_c) = 1$$

*Proof.*  $\Rightarrow$ : Choose  $c > 0$  such that  $\mathbb{P}(A_c) = 1$ . Take  $\omega \in A_c$  and choose  $N \in \mathbb{N}$  such that  $r(\omega)(n) \leq ch(n)$  for all  $n > N$ . Hence there exists  $\tilde{c} > 0$  such that  $\tilde{c}h^{-1}(r(\omega)(n)) \leq n$  for all  $n > N$ . By Lemma 3.3.1 we can choose  $M \in \mathbb{N}$  big enough such that  $T(\omega)(aM) > N$ , hence  $T(\omega)(as) \geq T(\omega)(aM) > N$  for all  $s > M$  since  $T(\omega)$  is increasing. Hence we can write  $\tilde{c}h^{-1}(r(\omega)(T(\omega)(as))) \leq T(\omega)(as)$  for all  $s > M$  and since  $r(\omega)(T(\omega)(as)) \leq as$  we finally get  $\tilde{c}h^{-1}(a)h^{-1}(s) = \tilde{c}h^{-1}(as) \leq T(\omega)(as)$  for all  $s > M$ , hence  $\omega \in D_{\tilde{c}h^{-1}(a)}$ , where we used the multiplicativity of  $h$  and that  $h^{-1}$  is decreasing. We therefore get  $A_c \subset D_{\tilde{c}h^{-1}(a)}$ , hence  $\mathbb{P}(D_{\tilde{c}h^{-1}(a)}) = 1$ .

$\Leftarrow$ : The proof for this direction works analogously and needs Lemma 3.3.1 as well.  $\square$

**Lemma 3.3.3.** Let  $n \in \mathbb{N}$ ,  $n \geq 2$  and  $T_1, \dots, T_n$  be independent geometrically distributed random variables with parameter  $0 < p < \frac{1}{2}$ . Let  $Y := T_1 + \dots + T_{n-1}$ , then for every  $a \in [2p, 1)$  we have

$$\mathbb{P}(Y \leq \frac{an}{p}) \leq \frac{1}{a}(ae^2)^n.$$

*Proof.* The moment generating function of  $Y$  is

$$\mathbb{E}[e^{tY}] = (pe^t)^{n-1}(1 - e^t(1-p))^{-(n-1)} = p^{n-1}(e^t - (1-p))^{-(n-1)}.$$

By Chebyshev for any random variable  $X$  we know the inequality

$$\mathbb{P}(X \geq x) \leq \inf_{t>0} \frac{\mathbb{E}[e^{tX}]}{e^{tx}}$$

and can therefore follow that for any  $t > 0$

$$\mathbb{P}(Y \leq \frac{an}{p}) = \mathbb{P}(-Y \geq -\frac{an}{p}) \leq \exp\left(\frac{ant}{p}\right)\mathbb{E}[e^{-tY}] = \exp\left(\frac{ant}{p}\right)p^{n-1}(e^t - (1-p))^{-(n-1)}.$$

Choose  $t = \ln(\frac{a(1-p)}{a-p})$  (note that  $t > 0$ ), then

$$\begin{aligned} \mathbb{P}(Y \leq \frac{an}{p}) &\leq \left(\frac{a(1-p)}{a-p}\right)^{\frac{an}{p}} p^{n-1}(1-p)^{-(n-1)} \left(\frac{p}{a-p}\right)^{-(n-1)} \\ &= \left(\frac{a(1-p)}{a-p}\right)^{\frac{an}{p}} (1-p)^{-(n-1)} (a-p)^{n-1} \\ &= \left(\frac{a}{a-p}\right)^{\frac{an}{p}} (1-p)^{\frac{an}{p}} (1-p)^{-(n-1)} (a-p)^{n-1} \\ &\leq \left(\frac{a}{a-p}\right)^{\frac{an}{p}} (1-p)^{n-1} (1-p)^{-(n-1)} (a-p)^{n-1} \\ &\leq \left(1 + \frac{p}{a-p}\right)^{\frac{2(a-p)n}{p}} a^{n-1} \\ &\leq \frac{1}{a}(ae^2)^n. \end{aligned}$$

$\square$

### 3 Incremental aggregation

The following lemma gives a tool to handle conditioned probabilities and will be helpful later for developing splittings of probabilities similar to the law of total probability.

**Lemma 3.3.4.** Let  $A, B \in \mathcal{F}$  with  $\mathbb{P}(B) > 0$ . Further let  $c > 0$  and  $(C_i)_{i \in \mathbb{N}} \subset \mathcal{F}$  be a sequence of pairwise disjoint sets with  $\mathbb{P}(B \cap C_i) > 0$ ,  $\mathbb{P}(A | B \cap C_i) \geq c$  for all  $i \in \mathbb{N}$  and  $B \subset \bigcup_{i \in \mathbb{N}} C_i$ . Then

$$\mathbb{P}(A | B) \geq c.$$

*Proof.* We have

$$\begin{aligned} \mathbb{P}(A | B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B \cap \bigcup_{i \in \mathbb{N}} C_i)}{\mathbb{P}(B)} \\ &= \sum_{i \in \mathbb{N}} \frac{\mathbb{P}(A \cap B \cap C_i)}{\mathbb{P}(B)} = \sum_{i \in \mathbb{N}} \frac{\mathbb{P}(B \cap C_i)}{\mathbb{P}(B)} \frac{\mathbb{P}(A \cap B \cap C_i)}{\mathbb{P}(B \cap C_i)} \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}(C_i | B) \mathbb{P}(A | B \cap C_i) \geq c \sum_{i \in \mathbb{N}} \mathbb{P}(C_i | B) \\ &= c. \end{aligned}$$

□

**Definition 3.3.1.** For  $x \in \mathbb{Z}^d$  and  $r \in [1, \infty)$  define

$$\mathcal{P}_r^x := \{A \in \mathcal{P}_f \mid x \in A, r = \max_{y \in A} |x - y| \text{ and } A \text{ is connected}\}.$$

**Lemma 3.3.5.** Let  $q > 0$  and  $(\mu_A)_{A \in \mathcal{P}_f^d}$  be translation invariant. If there exists a constant  $c > 0$  such that for all  $r \in [1, \infty)$  we have

$$\mu_A(0) \leq cr^{-q}, \quad \text{for all } A \in \mathcal{P}_r^0,$$

then there also exists a constant  $\tilde{c} > 0$  such that for all  $y \in \mathbb{Z}^d$  and  $s, r \in [1, \infty)$  with  $s \geq r$  we have

$$\mu_A(y) \leq \tilde{c}r^{-q}, \quad \text{for all } A \in \mathcal{P}_s^y.$$

*Proof.* Let  $y \in \mathbb{Z}^d$  and choose  $s, r \in [1, \infty)$  with  $s \geq r$ . Then with the translation invariance of  $(\mu_A)_{A \in \mathcal{P}_f^d}$  we get

$$\mu_A(0) = h_{\Phi_y(A)}(\Phi_y(0)) = h_{\Phi_y(A)}(y) \leq cs^{-q} \leq cr^{-q} \quad \text{for all } A \in \mathcal{P}_s^0,$$

and since  $A \in \mathcal{P}_s^0 \Leftrightarrow \Phi_y(A) \in \mathcal{P}_s^{\Phi_y(0)}$  we get

$$\mu_A(y) \leq cr^{-q} \quad \text{for all } A \in \mathcal{P}_s^y.$$

Since  $c$  was chosen independently of  $y, s$  and  $r$ , this completes the proof. □

At this point we are prepared to proof the following theorem.

### 3 Incremental aggregation

**Theorem 3.3.1.** Let  $q > 0$  and  $(\mu_A)_{A \in \mathcal{P}_f^d}$  be a translation invariant outside distribution. If there exists a constant  $C > 0$  such that for all  $r \in [1, \infty)$  we have

$$\mu_A(0) \leq Cr^{-q} \quad \text{for all } A \in \mathcal{P}_r^0,$$

then for the fractal dimension we have

$$d_f \geq 1 + q \quad \text{a.s..}$$

*Proof.* By (3.6) we have

$$d_f = \frac{1}{\alpha_f} \quad \text{a.s.}$$

where  $\alpha_f$  is the growth rate as defined in Section (3.5), and we will therefore show that

$$\alpha_f \leq \frac{1}{1+q} \quad \text{a.s..}$$

For  $c > 0$  define

$$A_c := \{\omega \in \Omega \mid \text{rad}(\mathcal{E}_n(\omega)) \leq cn^{\frac{1}{1+q}} \text{ for large } n\}$$

and

$$D_c := \{\omega \in \Omega \mid T(\omega)(2n) \geq cn^{1+q} \text{ for large } n\}.$$

If there is a constant  $c > 0$  such that

$$\mathbb{P}(A_c) = 1, \tag{3.9}$$

then we have

$$\alpha_f = \limsup_{n \rightarrow \infty} \frac{\ln(\text{rad}(\mathcal{E}_n))}{\ln(n)} \leq \limsup_{n \rightarrow \infty} \frac{\ln(cn^{\frac{1}{1+q}})}{\ln(n)} = \frac{1}{1+q} \quad \text{a.s..}$$

If we choose

$$h : [0, \infty) \rightarrow [0, \infty), x \mapsto x^{\frac{1}{1+q}},$$

and  $a = 2$ , then, by Lemma 3.3.2, we can show (3.9) if we find a constant  $c > 0$  such that

$$\mathbb{P}(D_c) = 1. \tag{3.10}$$

Note that  $h$  is bijective, multiplicative and increasing.

So now we will try to find a constant  $c > 0$  such that (3.10) holds. For  $n \in \mathbb{N}$  write  $\mathcal{E}_n = \{y_1, \dots, y_n\}$  according to Definition 3.1.1, where  $y_j$  is the  $j$ -th point added to the

### 3 Incremental aggregation

cluster. Let  $\beta > 0$  which will be determined later on. For  $n \in \mathbb{N}$  let  $\tilde{m}_n := \beta n^{1+q}$  and define

$$V_n := \{\omega \in \Omega \mid T(\omega)(2n) < \tilde{m}_n\}.$$

Further define the set of all possible random walk paths of length  $n$  with starting point in the set  $\tilde{B}_n := \{x \in \mathbb{Z}^d \mid n \leq |x| < n+1\}$  by

$$Z_n := \{[z] := (z_1, \dots, z_n) \in (\mathbb{Z}^d)^n \mid z_1 \in \tilde{B}_n, z_i \in N(z_{i-1}) \text{ for } i \geq 2 \text{ and } z_i \neq z_j \text{ for } i \neq j\}.$$

For  $[z] \in Z_n$  define the events

$$W_n([z]) := \{\omega \in \Omega \mid \exists j_1 < \dots < j_n \leq \tilde{m}_n \text{ such that } y_{j_i}(\omega) = z_i \text{ for all } i \in \{1, \dots, n\}\}$$

and the union of these events by

$$W_n := \bigcup_{[z] \in Z_n} W_n([z]).$$

With  $\tilde{B}_n$  we mean the „the boundary“ of  $B_n$ , the ball with center 0 and radius  $n$ . We will quickly prove that

$$V_n \subset W_n \text{ for all } n \in \mathbb{N}.$$

Let  $n \in \mathbb{N}$  and  $\omega \in V_n$ , then  $T(\omega)(2n) < \tilde{m}_n$ . For  $m_n := \max\{j \in \mathbb{N} \mid j \leq \tilde{m}_n\}$  we therefore have  $T(\omega)(2n) \leq m_n$ , hence  $\text{rad}(\mathcal{E}_{m_n}(\omega)) \geq 2n$ . Therefore, since  $\mathcal{E}_{m_n}(\omega)$  is connected, there must exist indices  $j_1 < \dots < j_n$  such that  $[z_0] := (y_{j_1}(\omega), \dots, y_{j_n}(\omega)) \in Z_n$  and  $[z_0] \subset \mathcal{E}_{m_n}(\omega)$ , since  $\max_{x \in [z_0]} |x| \leq 2n$ . Therefore  $j_n \leq m_n \leq \tilde{m}_n$  and therefore  $\omega \in W_n([z_0]) \subset W_n$  which proofs the inclusion.

For a sequence of events  $(A_n)_{n \in \mathbb{N}}$  recall that

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n \in \mathbb{N}} \bigcup_{i \geq n} A_i = \{\omega \in \Omega \mid \omega \in A_n \text{ for infinitely many } n \in \mathbb{N}\}.$$

We will use the Lemma of Borel-Cantelli on the sequence of events  $(W_n)_{n \in \mathbb{N}}$ . If we can show that

$$\sum_{n \in \mathbb{N}} \mathbb{P}(W_n) < \infty, \tag{3.11}$$

then with  $V_n \subset W_n$  for all  $n \in \mathbb{N}$  and Borel-Cantelli we get

$$\mathbb{P}(\limsup_{n \rightarrow \infty} V_n) \leq \mathbb{P}(\limsup_{n \rightarrow \infty} W_n) = 0.$$

Since

$$(\limsup_{n \rightarrow \infty} V_n)^C = \{\omega \in \Omega \mid \exists N \in \mathbb{N} \text{ s.t. } \omega \in V_n^C \text{ for all } n > N\} = D_\beta$$

### 3 Incremental aggregation

we can then conclude that  $\mathbb{P}(D_\beta) = 1$  and have finished the proof. So we want to show (3.11).

For  $n \in \mathbb{N}$ ,  $[z] \in Z_n$  and  $i \in \{1, \dots, n\}$  we define random variables

$$\tau_i : \Omega \rightarrow \mathbb{N}^\infty, \tau_i(\omega) = j : \Leftrightarrow \begin{cases} y_j(\omega) = z_i, & j < \infty, \\ z_i \notin \mathcal{E}_\infty(\omega), & j = \infty, \end{cases}$$

so  $\tau_i$  is either the index  $j$  such that the  $j$ -th added point is  $z_i$ , or infinity if the limit cluster  $\mathcal{E}_\infty$  doesn't contain  $z_i$ .  $\tau_i$  is measurable because for  $j \in \mathbb{N}$  we have

$$\tau_i^{-1}(j) = \{y_j = z_i\} \in \mathcal{F}$$

and for  $j = \infty$  we have

$$\begin{aligned} \tau_i^{-1}(\infty) &= \{z_i \notin \mathcal{E}_\infty = \bigcup_{k \in \mathbb{N}} \mathcal{E}_k\} \\ &= \{z_i \notin \mathcal{E}_k \text{ for all } k \in \mathbb{N}\} \\ &= \bigcap_{k \in \mathbb{N}} \{z_i \notin \mathcal{E}_k\} \\ &= \bigcap_{k \in \mathbb{N}} \bigcap_{l=1}^k \{y_l \neq z_i\} \in \mathcal{F}. \end{aligned}$$

Further for  $i \in \{1, \dots, n-1\}$  we define waiting times

$$\sigma_i : \Omega \rightarrow \mathbb{N}^\infty, \omega \rightarrow \begin{cases} \tau_{i+1}(\omega) - \tau_i(\omega), & \text{if } \tau_{i+1}(\omega) < \infty \text{ and } \tau_i(\omega) < \infty, \\ \infty, & \text{else,} \end{cases}$$

so  $\sigma_i$  is the waiting time between adding  $z_i$  and  $z_{i+1}$  to the cluster, if both are added. We quickly argue that  $\sigma_i$  is measurable as well. For  $j \in \mathbb{N}$  we have

$$\sigma_i^{-1}(j) = \{\tau_{i+1} - \tau_i = j\} = \bigcup_{k \in \mathbb{N}} \{\tau_{i+1} = k\} \cap \{\tau_i = k - j\} \in \mathcal{F}$$

and

$$\sigma_i^{-1}(\infty) = \{\tau_{i+1} = \infty\} \cup \{\tau_i = \infty\} \in \mathcal{F},$$

since  $\tau_{i+1}$  and  $\tau_i$  are measurable.

Now let  $n \in \mathbb{N}$  and  $[z] \in Z_n$ . Let  $i \in \{1, \dots, n-1\}$  and define the event

$$U_{[z]}^i := \{\tau_1 \leq \dots \leq \tau_i\}.$$

Note that  $\mathbb{P}(U_{[z]}^i) > 0$  and also note again, that all  $\tau_i$  are defined based on one  $[z]$ . We will now prove that the distribution of  $\sigma_i$  conditioned on  $U_{[z]}^i$  is dominated by that of a geometrically distributed random variable with parameter

$$p_n := c_1 n^{-q} \tag{3.12}$$

### 3 Incremental aggregation

for some constant  $c_1 > 0$  which is the same constant for all  $i \in \{1, \dots, n-1\}$  and for all  $n \in \mathbb{N}$ . We want to use Lemma 3.3.5 here for which we need to create terms where the distribution  $\mu$  appears. For that we need probabilities where we condition on a given cluster which we will develop now. We first define some helpful sets. For  $m \in \mathbb{N}$  define

$$C_m := \{\mathcal{E}_m(\omega) \mid \omega \in \Omega\}$$

the set of possible clusters of size  $m$  and the disjoint union of them as  $C := \bigcup_{m \in \mathbb{N}} C_m$ . Further define another disjoint partition of  $C$  into sets

$$E_m := \{\mathcal{E} \in C \mid m-1 \leq \text{rad}(\mathcal{E}) < m\}$$

for  $m \in \mathbb{N}$  such that  $C = \bigcup_{m \in \mathbb{N}} E_m$  as well. Since  $|z_1| \geq n$  and considering  $U_{[z]}^i$ , we need a minimal amount of time steps such that it is possible for the cluster to contain  $z_i$ . So there exists a  $n_i \in \mathbb{N}$  such that

$$\mathbb{P}(\tau_i = j, U_{[z]}^i) \begin{cases} > 0, & \text{for } n_i \leq j \leq \infty, \\ = 0, & \text{for } j < n_i. \end{cases}$$

Further if we choose  $n_i \leq j < \infty$  and considering  $U_{[z]}^i$  again, then  $\mathcal{E}_{j-1}$  must have some minimal radius if we want it to be possible that the next added point is  $z_i$ , similarly argumented as above. So there exists a  $m_i \in \mathbb{N}$  such that

$$\mathbb{P}(\mathcal{E}_{j-1} = \mathcal{E}, \tau_i = j, U_{[z]}^i) \begin{cases} > 0, & \mathcal{E} \in \bigcup_{m \geq m_i} E_m^i, \\ = 0, & \mathcal{E} \in C \setminus \bigcup_{m \geq m_i} E_m^i, \end{cases}$$

where  $E_m^i := \{\mathcal{E} \in E_m \mid \{z_1, \dots, z_{i-1}\} \subset \mathcal{E} \text{ and } z_i \in \partial\mathcal{E}\}$  with  $m \in \mathbb{N}$ . Since  $n \leq |z_1| < n+1$  we actually have exactly  $m_i = n$ . Having made that clear we can use the law of total probability twice in the following. For  $k \in \mathbb{N}$  we get

$$\begin{aligned} \mathbb{P}(\sigma_i > k \mid U_{[z]}^i) &= \sum_{n_i \leq j \leq \infty} \mathbb{P}(\tau_i = j \mid U_{[z]}^i) \mathbb{P}(\sigma_i > k \mid \tau_i = j, U_{[z]}^i) \\ &= \mathbb{P}(\tau_i = \infty \mid U_{[z]}^i) \mathbb{P}(\sigma_i > k \mid \tau_i = \infty, U_{[z]}^i) \\ &\quad + \sum_{n_i \leq j < \infty} \mathbb{P}(\tau_i = j \mid U_{[z]}^i) \mathbb{P}(\sigma_i > k \mid \tau_i = j, U_{[z]}^i) \\ &= \mathbb{P}(\tau_i = \infty \mid U_{[z]}^i) + \sum_{n_i \leq j < \infty} \mathbb{P}(\tau_i = j \mid U_{[z]}^i) \mathbb{P}(\sigma_i > k \mid \tau_i = j, U_{[z]}^i), \end{aligned}$$

note that  $\mathbb{P}(\sigma_i > k \mid \tau_i = \infty, U_{[z]}^i) = 1$  since  $\{\tau_i = \infty\} \subset \{\sigma_i = \infty\} \subset \{\sigma_i > k\}$  for all  $k \in \mathbb{N}$ . For shorter expressions write  $\Gamma_{ij} := \{\mathcal{E}_{j-1} = \mathcal{E}\} \cap \{\tau_i = j\} \cap U_{[z]}^i$ . Further for  $n_i \leq j < \infty$  we get

$$\mathbb{P}(\sigma_i > k \mid \tau_i = j, U_{[z]}^i) = \sum_{m_i \leq m} \sum_{\mathcal{E} \in E_m^i} \mathbb{P}(\mathcal{E}_{j-1} = \mathcal{E} \mid \tau_i = j, U_{[z]}^i) \mathbb{P}(\sigma_i > k \mid \Gamma_{ij}).$$

### 3 Incremental aggregation

We now state that there exists a constant  $c_1 > 0$  such that for any  $n_i \leq j < \infty$ ,  $m_i \leq m$  and  $\mathcal{E} \in E_m^i$  we have

$$\mathbb{P}(\sigma_i > k \mid \Gamma_{ij}) \geq (1 - c_1 n^{-q})^k \quad (3.13)$$

for all  $k \in \mathbb{N}$ , which we prove by induction. For  $k = 1$  by Lemma 3.3.5 there exists a  $\tilde{C} > 0$  such that

$$\begin{aligned} \mathbb{P}(\sigma_i > 1 \mid \Gamma_{ij}) &\geq \mathbb{P}(y_{j+1} \neq z_{i+1} \mid \Gamma_{ij}) \\ &= 1 - \mu_{\partial(\mathcal{E} \cup \{z_i\})}(z_{i+1}) \\ &= 1 - \mu_{\overline{\mathcal{E} \cup \{z_i\}}}(z_{i+1}) \\ &\geq 1 - \tilde{C} n^{-q} \\ &= (1 - \tilde{C} n^{-q})^1. \end{aligned}$$

where we used that  $\mu$  is an outside distribution. Note that we can use Lemma 3.3.5 because  $\text{rad}(\overline{\mathcal{E} \cup \{z_i\}}) \geq n$  since  $z_1 \in \mathcal{E} \cup \{z_i\}$  as we conditioned on  $U_{[z]}^i$ . Note that the harmonic measure doesn't care about the order of how points where added to form the current cluster as mentioned in Remark 3.1.2. Note that by Lemma 3.3.5 the chosen  $\tilde{C}$  has a strong universal property such that it doesn't depend on  $n_i \leq j < \infty$ ,  $m_i \leq m$  or  $\mathcal{E} \in E_m$ . Now let the statement be true for some  $k = l$ . Then again by Lemma 3.3.5 we have

$$\begin{aligned} \mathbb{P}(\sigma_i > l + 1 \mid \Gamma_{ij}) &= \mathbb{P}(\sigma_i > l, y_{l+1} \neq z_{i+1} \mid \Gamma_{ij}) \\ &= \mathbb{P}(\sigma_i > l \mid \Gamma_{ij}) \mathbb{P}(y_{l+1} \neq z_{i+1} \mid \Gamma_{ij}, \sigma_i > l) \\ &\stackrel{(+) }{\geq} (1 - \tilde{C} n^{-q})^l \mathbb{P}(y_{l+1} \neq z_{i+1} \mid \Gamma_{ij}, \sigma_i > l) \\ &\stackrel{(++)}{\geq} (1 - \tilde{C} n^{-q})^l (1 - \tilde{C} n^{-q}) \\ &= (1 - \tilde{C} n^{-q})^{l+1}, \end{aligned}$$

where in (+) we used the induction assumption. In order to show (++) we would need to split the probability with the law of total probability by conditioning on all clusters such that  $\Gamma_{ij}$  is fulfilled and the next  $l$  added points are not equal to  $z_{i+1}$ . We would then have the same inequality as in the induction beginning using Lemma 3.3.5 and reput together the conditioned probabilites to get this inequality here. With  $c_1 := \tilde{C}$  we have therefore found a constant such that (3.13) holds for all  $k \in \mathbb{N}$  and this constant  $c_1$  is independent of  $n_i \leq j < \infty$ ,  $m_i \leq m$  or  $\mathcal{E} \in E_m$ .

Since  $\{\tau_i = j\} \cap U_{[z]}^i \subset \bigcup_{m_i \leq m} \bigcup_{\mathcal{E} \in E_m^i} \{\mathcal{E}_{j-1} = \mathcal{E}\}$  by construction of  $E_m^i$  and  $m_i$  we can use Lemma 3.3.4 and get for  $n_i \leq j \leq \infty$

$$\mathbb{P}(\sigma_i > k \mid \tau_i = j, U_{[z]}^i) \geq (1 - c_1 n^{-q})^k$$

for all  $k \in \mathbb{N}$ . Because  $U_{[z]}^i \subset \bigcup_{n_i \leq j \leq \infty} \{\tau_i = j\}$  by construction of  $n_i$  we can use Lemma 3.3.4 again and finally get

$$\mathbb{P}(\sigma_i > k \mid U_{[z]}^i) \geq (1 - c_1 n^{-q})^k$$

### 3 Incremental aggregation

for all  $k \in \mathbb{N}$ . We therefore have shown what we stated in (3.12).

Note again that we needed the constant  $c_1$  to not depend on any  $j$ ,  $m$  or  $\mathcal{E}$  which it does not by its universal property given by Lemma 3.3.5. The universality is so strong that it does not even depend on  $i \in \{1, \dots, n-1\}$  or  $n \in \mathbb{N}$ .

We now continue to show (3.11). Choose  $\beta$  such that  $2d\beta c_1 e^2 < 1$  und choose  $N \in \mathbb{N}$  such that  $\beta c_1 \geq 2p_n$  for all  $n > N$ . If we then define  $a := \beta c_1$  we get  $\tilde{m}_n = \frac{an}{p_n}$  and with  $Y := \sum_{i=1}^{n-1} \sigma_i$  we can use Lemma 3.3.3 to get

$$\mathbb{P}(\tau_n \leq \tilde{m}_n \mid U_{[z]}^n) \leq \mathbb{P}(\tau_n - \tau_1 \leq \tilde{m}_n \mid U_{[z]}^n) = \mathbb{P}(Y \leq \frac{an}{p_n} \mid U_{[z]}^n) \leq \frac{1}{a}(ae^2)^n$$

for all  $n > N$ . Since  $W_n([z]) \subset \{\tau_n \leq \tilde{m}_n\} \cap U_{[z]}^n$  we get

$$\mathbb{P}(W_n([z])) \leq \mathbb{P}(W_n([z]) \mid U_{[z]}^n) \leq \mathbb{P}(\tau_n \leq \tilde{m}_n \mid U_{[z]}^n) \leq \frac{1}{a}(ae^2)^n$$

for all  $n > N$ . Counting the elements in  $Z_n$  we have less than or equal to  $c_2 n$  points in  $\tilde{B}_n$  for some constant  $c_2 > 0$  as starting points and at most  $(2d)^{n-1}$  possibilities for the next  $n-1$  steps for a path of length  $n$ . So  $|Z_n| \leq c_2 n (2d)^{n-1}$  and therefore

$$\mathbb{P}(W_n) \leq \sum_{[z] \in Z_n} \mathbb{P}(W_n([z])) \leq c_2 n (2d)^{n-1} \frac{1}{a} (ae^2)^n = \frac{c_2}{2da} n (2dae^2)^n \quad \text{for all } n > N$$

and finally

$$\sum_{n>N} \mathbb{P}(W_n) \leq \sum_{n>N} \frac{c_2}{2d\beta c_1} n (2d\beta c_1 e^2)^n =: r_\beta.$$

Now note again, that  $c_1$ ,  $c_2$  and  $\beta$  are not depending on  $n$ . Since  $\beta$  was chosen such that  $2d\beta c_1 e^2 < 1$ ,  $r_\beta$  is finite and therefore in total we get

$$\sum_{n \in \mathbb{N}} \mathbb{P}(W_n) < \infty,$$

which completes the proof. □

## 4 External diffusion limited aggregation

### 4.1 Definition

The following model of an incremental aggregation uses a very natural family of distributions called the *harmonic measures*. The idea is to let a new particle start a random walk from „far away“ and let it stick at the cluster where it hits it the first time. In a physical context this can be understood as diffusion of particles sticking and forming a cluster based on random movements. This incremental aggregation is called *external diffusion limited aggregation* and we define its distribution in the following which is motivated by [3, Chapter 2, definition 2.1].

**Definition 4.1.1.** (*harmonic measure*) Let  $A \subset \mathbb{Z}^d$ . The hitting probability of  $A$  is the function

$$H_A : \mathbb{Z}^d \times A \rightarrow [0, 1], \quad (x, y) \mapsto H_A(x, y) := \mathbb{P}_x(S_{T_A^+} = y).$$

For a fixed element  $x \in \mathbb{Z}^d$  the function  $H_A(x, \cdot)$  defines a measure on  $A$  with total mass  $\mathbb{P}_x(T_A^+ < \infty)$  and it can be adapted to a probability measure by conditioning the random walk to hit  $A$  in finite time. Define

$$\bar{H}_A : \mathbb{Z}^d \times A \rightarrow [0, 1], \quad (x, y) \mapsto \bar{H}_A(x, y) := \mathbb{P}_x(S_{T_A^+} = y \mid T_A^+ < \infty),$$

so for fixed  $x \in \mathbb{Z}^d$  the function  $\bar{H}_A(x, \cdot)$  defines a probability measure on  $A$ . In [3, Theorem 2.1.3] it is proved, that for finite sets  $A \in \mathcal{P}_f^d$  the limit

$$\lim_{|x| \rightarrow \infty} \bar{H}_A(x, y) =: h_A(y)$$

exists for each  $y \in A$ . The function  $h_A : A \rightarrow [0, 1]$  is called the *harmonic measure of  $A$* . For an element  $y \in A$ ,  $h_A$  can be interpreted as the probability that a random walk starting at „infinity“ hits  $A$  the first time at  $y$ .

**Definition 4.1.2.** (*external diffusion limited aggregation*)

*External diffusion limited aggregation* (in  $\mathbb{Z}^d$ ), short *external DLA* or just *DLA*, is an incremental aggregation with the family of harmonic measures  $(h_A)_{A \in \mathcal{P}_f^d}$  as distribution.

**Remark 4.1.1.** If we look at the 2-dimensional case, by Lemma 2.4.2, we have that  $\mathbb{P}_x(T_A^+ < \infty) = 1$  for any  $x \in \mathbb{Z}^2$ , and therefore get  $\bar{H}_A = H_A$ . So in two dimensions the harmonic measure of  $A \subset \mathbb{Z}^2$  can be written as

$$h_A(y) = \lim_{|x| \rightarrow \infty} \mathbb{P}_x(S_{T_A^+} = y) \quad \text{for } y \in A.$$

### 4.2 Fractal dimension and growth rate of external DLA in $\mathbb{Z}^2$

There is not yet a rigorous proof on the exact fractal dimension of external DLA. There are rather few rigorously proved results and it seems that it is very hard to prove such

results for external DLA. Looking at computer simulations of DLA in  $\mathbb{Z}^2$  it seems that the clusters are relatively sparse and they appear to have a noninteger fractal dimension. In [5] DLA is observed in a magnetic aggregation context and empirically they find a fractal dimension of around 1.8. We will proof in the following an almost sure lower bound for the fractal dimension of external DLA for which we will use Theorem 3.3.1. Recall that  $\mathcal{P}_f := \mathcal{P}_f^2$ .

**Lemma 4.2.1.** The family of harmonic measures  $(h_A)_{A \in \mathcal{P}_f}$  in  $\mathbb{Z}^2$  is translation invariant.

*Proof.* Let  $y \in \mathbb{Z}^2$  and  $A \in \mathcal{P}_f$ . Recall that in  $\mathbb{Z}^2$  the harmonic measure of  $A \subset \mathbb{Z}^2$  can be written as

$$h_A(z) = \lim_{|x| \rightarrow \infty} \mathbb{P}_x(S_{T_A^+} = z)$$

for any  $z \in A$  (see Remark 4.1.1). Since the distribution of random walks in  $\mathbb{Z}^2$  is invariant under translation, we can conclude that

$$\begin{aligned} h_A(z) &= \lim_{|x| \rightarrow \infty} \mathbb{P}_x(S_{T_A^+} = z) \\ &= \lim_{|x| \rightarrow \infty} \mathbb{P}_{\Phi_y(x)}(S_{T_{\Phi_y(A)}^+} = \Phi_y(z)) \\ &= h_{\Phi_y(A)}(\Phi_y(z)) \end{aligned}$$

for all  $z \in A$ . □

**Theorem 4.2.1.** In  $\mathbb{Z}^2$  there exists a constant  $c > 0$  such that for all  $r \in [1, \infty)$

$$h_A(0) \leq cr^{-\frac{1}{2}} \quad \text{for all } A \in \mathcal{P}_r^0.$$

*Proof.* If  $A \in \mathcal{P}_r^0$  then  $0 \in A$  and  $r = \text{rad}(A)$  and for that case there is a proof presented for all dimensions  $d \in \mathbb{N}$  in [3, Theorem 2.5.2]. The proof is very technical and requires various results of other theorems and lemmas which are also to find in [3]. □

**Theorem 4.2.2.** For the fractal dimension of external DLA in  $\mathbb{Z}^2$  as defined in (3.4) we have

$$d_f \geq \frac{3}{2} \quad \text{a.s..} \tag{4.1}$$

*Proof.* We have shown that the harmonic measure is translation invariant and it is not difficult to understand that the harmonic measure is an outside distribution as defined in Definition 3.1.2. Together with Theorem 4.2.1 we can therefore use Theorem 3.3.1 with  $q = \frac{1}{2}$  to conclude that

$$d_f \geq 1 + \frac{1}{2} = \frac{3}{2} \quad \text{a.s..}$$

□

We have therefore shown a lower bound of 1.5 for the fractal dimension of external DLA in two dimensions. In the next chapter we will look at another incremental aggregation which distribution works much simpler than the harmonic measure. In two dimensions we will therefore even be able to calculate its exact fractal dimension.

## 5 Line hitting aggregation

### 5.1 Motivation

In the following we will look at a process with another simple idea for its distribution. For this process we will consider the 2-dimensional case  $\mathbb{Z}^2$ . The idea is to let particles move on straight lines coming from infinity and add them to the cluster where they hit it. A similar approach can be found in [4] as a so called ballistic model, where they let particles move on straight lines to the cluster, but just from a small range of directions. Our approach here extends this idea, as we will try to choose a random direction out of all directions in the plane. The difficulty here is to argue that this random choosing happens in a sensible and fair way considering the asymmetry of an already formed cluster. Anyway we will see that the fractal dimension of 2 which is guessed in [4] for their ballistic model matches the one we will prove rigorously here.

Obviously in most cases particles cannot move completely straight on  $\mathbb{Z}^2$ . Therefore we will consider points in  $\mathbb{Z}^2$  as the centers of unit squares in  $\mathbb{R}^2$  and let the particles move along straight lines in the plane  $\mathbb{R}^2$ . We consider a line hitting a point in  $\mathbb{Z}^2$  if and only if it intersects its unit square as defined in the following. Note that we identify  $\mathbb{R}^2$  with  $\mathbb{C}$  here. Also recall that  $\mathcal{P}_f$  is the set of finite subsets of  $\mathbb{Z}^2$  and  $\mathcal{G}$  is the set of lines in the plane.

**Definition 5.1.1.** Define

$$\mathbb{C}_{sq} := \left\{ \left[ k - \frac{1}{2}, k + \frac{1}{2} \right] + \left[ l - \frac{1}{2}, l + \frac{1}{2} \right] i \subset \mathbb{C} \mid k, l \in \mathbb{Z} \right\}, \quad (5.1)$$

note that  $\mathbb{C} = \bigcup_{s \in \mathbb{C}_{sq}} s$ . The canonical function

$$sq : \mathbb{Z}^2 \rightarrow \mathbb{C}_{sq}, \quad (k, l) \mapsto \left[ k - \frac{1}{2}, k + \frac{1}{2} \right] + \left[ l - \frac{1}{2}, l + \frac{1}{2} \right] i \quad (5.2)$$

is bijective and intuitively identifies points in  $\mathbb{Z}^2$  with squares in  $\mathbb{C}$  which is  $p$  is the center of the square  $sq(p)$  for all  $p \in \mathbb{Z}^2$ . In the following when using a point  $p \in \mathbb{Z}^2$  it will reference the point in  $\mathbb{Z}^2$  or the corresponding square in  $\mathbb{C}$  respecting the context. This bijection also naturally defines a graph structure on  $\mathbb{C}_{sq}$ , which is two squares  $s_1, s_2 \in \mathbb{C}_{sq}$  form an edge if and only if  $sq^{-1}(s_1)$  and  $sq^{-1}(s_2)$  form an edge in  $\mathbb{Z}^2$ . For the following we say a line  $g$  hits a point  $p \in \mathbb{Z}^2$  if and only if  $g \cap sq(p) \neq \emptyset$  (see in *Figure 6*). For a set  $A \in \mathcal{P}_f$  define  $sq(A) := \bigcup_{p \in A} sq(p)$ . In the following we discuss a parametrisation of lines which is completely motivated by [6, 2.1.1].

**Definition 5.1.2.** We denote this set of lines by  $\mathcal{G}$ . Every line can be uniquely determined by an angle  $\alpha \in [0, \pi)$  and a real number  $p \in \mathbb{R}$ . Let  $\langle \cdot, \cdot \rangle$  be the standard scalar product on  $\mathbb{R}^2$ , respectively used for values in  $\mathbb{C}$  as we identify  $\mathbb{R}^2$  with  $\mathbb{C}$  as  $\mathbb{R}$ -vector spaces. Let

$$e_\alpha := e^{\alpha i} = \cos(\alpha) + \sin(\alpha)i$$

## 5 Line hitting aggregation

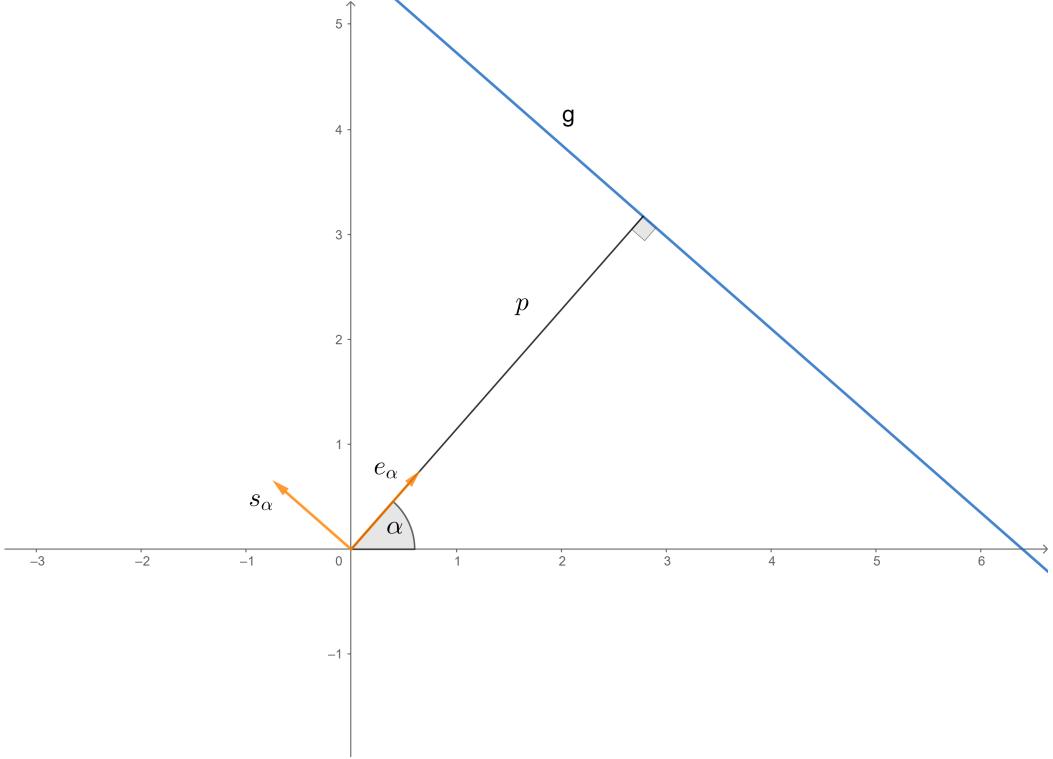


Figure 5: Line parameters  $\alpha$  and  $p$

and

$$s_\alpha := -\sin(\alpha) + \cos(\alpha)i$$

be the unit vectors  $1$  and  $i$  rotated by  $\alpha$  counterclockwise. Lets consider the representation  $x = \langle x, e_\alpha \rangle e_\alpha + \langle x, s_\alpha \rangle s_\alpha$  for  $x \in \mathbb{C}$ . Since  $e_\alpha$  and  $s_\alpha$  form a base of  $\mathbb{C}$  as a 2-dimensional  $\mathbb{R}$ -vector space, the parameters  $\langle x, e_\alpha \rangle$  and  $\langle x, s_\alpha \rangle$  are unique for each  $x$ . It thus is easy to realize that  $g_{\alpha,p} := \{x \in \mathbb{C} \mid \langle x, e_\alpha \rangle = p\}$  defines a line (compare with *Figure 5*) and that for every line  $g \in \mathcal{G}$  there is a unique pair of  $(\alpha, p)$  such that  $g = g_{\alpha,p}$ . In words,  $g_{\alpha,p}$  contains all points which have length  $p$  in direction of  $e_\alpha$ . With  $\Phi := [0, \pi) \times \mathbb{R}$  this naturally defines a bijection

$$\chi : \Phi \rightarrow \mathcal{G}, \quad (\alpha, p) \mapsto g_{\alpha,p}.$$

We take the trace  $\mathcal{B}_\Phi := \mathcal{B}^2 \cap \Phi$  of the Borel- $\sigma$ -algebra on  $\Phi$  and define the  $\sigma$ -algebra  $\mathfrak{G}$  on  $\mathcal{G}$  by  $\mathfrak{G} := \chi(\mathcal{B}_\Phi)$ . This works well since  $\chi$  is a bijection.

**Definition 5.1.3.** Let  $g = g_{\alpha,p} \in \mathcal{G}$  and  $A \in \mathcal{P}_f$ . We define

$$g^A := \{p \in A \mid g \cap sq(p) \neq \emptyset\}$$

which is the subset of all points in  $A$  which are hit by  $g$  (see *Figure 6*). For the following we suppose  $g^A \neq \emptyset$ . We will define a total ordered relation  $\triangleleft$  on  $g^A$  which shall be defined

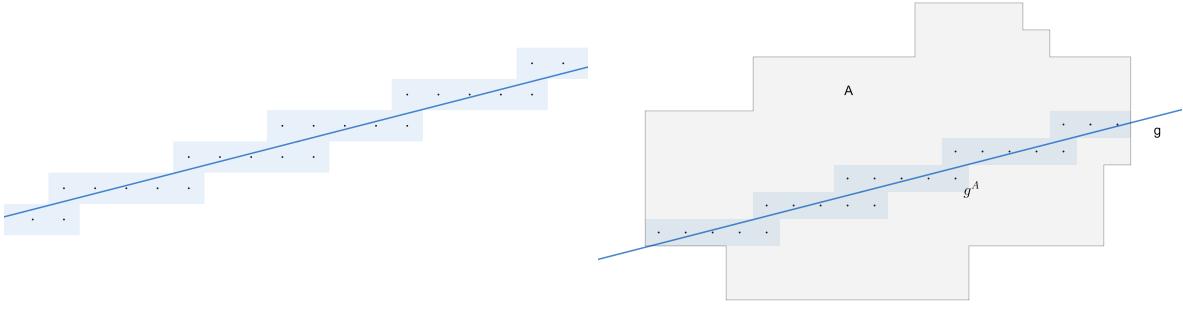


Figure 6: Line hitting squares

equivalently for all  $g \in \mathcal{G}$  and  $A \in \mathcal{P}_f$  with  $g^A \neq \emptyset$ . We choose two points  $x, y \in g^A$  and split the definition of the relation  $\triangleleft$  into four cases, depending on whether the line  $g$  goes from left-bottom to right-top, left-top to right-bottom, parallel to the  $x$ -axis or parallel to the  $y$ -axis. Denote the real part of  $x$  with  $\text{Re}(x)$  and the imaginary part with  $\text{Im}(x)$ .

*Case 1 :*  $g$  is parallel to the  $x$ -axis ( $\Leftrightarrow \alpha = \frac{\pi}{2}$ )

$$x \triangleleft y \Leftrightarrow \text{Re}(x) < \text{Re}(y)$$

*Case 2 :*  $g$  is parallel to the  $y$ -axis ( $\Leftrightarrow \alpha = 0$ )

$$x \triangleleft y \Leftrightarrow \text{Im}(x) < \text{Im}(y)$$

*Case 3 :*  $g$  is going from left-bottom to right-top ( $\Leftrightarrow \alpha \in (\frac{\pi}{2}, \pi)$ )

$$x \triangleleft y \Leftrightarrow \begin{cases} \text{Re}(x) < \text{Re}(y), & \text{if } \text{Re}(x) \neq \text{Re}(y), \\ \text{Im}(x) < \text{Im}(y), & \text{if } \text{Re}(x) = \text{Re}(y). \end{cases}$$

*Case 4 :*  $g$  is going from left-top to right-bottom ( $\Leftrightarrow \alpha \in (0, \frac{\pi}{2})$ )

$$x \triangleleft y \Leftrightarrow \begin{cases} \text{Re}(x) < \text{Re}(y), & \text{if } \text{Re}(x) \neq \text{Re}(y), \\ \text{Im}(x) > \text{Im}(y), & \text{if } \text{Re}(x) = \text{Re}(y). \end{cases}$$

It is easy to calculate that this defines a well-defined relation on  $g^A$ . In the following we will prove that this relation is totally ordered.

**Lemma 5.1.1.** For a line  $g = g_{\alpha,p} \in \mathcal{G}$  and  $A \in \mathcal{P}_f$  with  $g^A \neq \emptyset$  the relation  $\triangleleft$  on  $g^A$  is totally ordered.

*Proof.* We will only prove the case where  $g$  is going from left-bottom to right-top, which is *Case 3* of the definition. In this case we have  $\alpha \in (\frac{\pi}{2}, \pi)$ . Note, that the proof for *Case 4* will work very similar and in the case of  $g$  being parallel to one of the axes (*Case 1 or 2*), all properties for a totally ordered relation follow directly from the totally ordered relation  $<$  on  $\mathbb{R}$ . So let  $\alpha \in (\frac{\pi}{2}, \pi)$ .

*Antisymmetry* : For antisymmetry let  $x \triangleleft y$  and  $y \triangleleft x$ . Suppose  $\text{Re}(x) \neq \text{Re}(y)$ , then  $\text{Re}(x) < \text{Re}(y)$  and  $\text{Re}(y) < \text{Re}(x)$ , a contradiction because of the total order  $<$  in  $\mathbb{R}$ . So  $\text{Re}(x) = \text{Re}(y)$ . But then we have  $\text{Im}(x) < \text{Im}(y)$  and  $\text{Im}(y) < \text{Im}(x)$  and therefore also  $\text{Im}(x) = \text{Im}(y)$ , hence  $x = y$ .

*Transitivity* : For transitivity let  $x \triangleleft y$  and  $y \triangleleft z$ . We find four cases. In case  $\text{Re}(x) \neq \text{Re}(y)$  and  $\text{Re}(y) \neq \text{Re}(z)$  we get  $\text{Re}(x) < \text{Re}(z)$  by transitivity of  $<$ , hence  $x \triangleleft z$ . In case  $\text{Re}(x) \neq \text{Re}(y)$  and  $\text{Re}(y) = \text{Re}(z)$  we get  $\text{Re}(x) < \text{Re}(y) = \text{Re}(z)$ , therefore  $x \triangleleft z$ . In case  $\text{Re}(x) = \text{Re}(y)$  and  $\text{Re}(y) \neq \text{Re}(z)$  we get  $\text{Re}(x) = \text{Re}(y) < \text{Re}(z)$ , similar as the last case. In the last case  $\text{Re}(x) = \text{Re}(y) = \text{Re}(z)$  we get  $\text{Im}(x) < \text{Im}(y)$  and  $\text{Im}(y) < \text{Im}(z)$  and again by transitivity of  $<$  we get  $\text{Im}(x) < \text{Im}(z)$ , hence  $x \triangleleft z$  again.

*Connexity* : Connexity is given since for any two points  $x, y \in g^A$  we have either  $\text{Re}(x) \neq \text{Re}(y)$  or  $\text{Re}(x) = \text{Re}(y)$  and therefore either  $x \triangleleft y$  or  $y \triangleleft x$ .  $\square$

**Remark 5.1.1.** The relation  $\triangleleft$  on  $g^A$  basically orders the hitting points of  $g$  with  $A$  from left to right (or bottom to top in case of a line parallel to the  $y$ -axis). This order allows us to identify the outermost hitting points which are the minimum and maximum of  $g^A$  with respect to  $\triangleleft$ . To clarify, we define  $\min_{\triangleleft}(g^A) := x_0$  if and only if  $x_0 \triangleleft x$  for all  $x \in g^A, x \neq x_0$ , analogously  $\max_{\triangleleft}(g^A)$ . This means when moving on  $g$  facing  $A$  coming from infinity this order allows to know where the line  $g$  hits  $A$  first and where it hits  $A$  last. What we want to do next is to choose a line randomly and fair out of all lines which hit the current cluster. This is isn't obvious and we will have to develop and argue a most fair underlying distribution on lines in the next section where we will touch basic aspects of Integral Geometry.

## 5.2 Excursion: Integral geometry

In the next section we want to define an incremental aggregation for which distribution we need some concepts and results from integral geometry. We will want to fairly choose a random line out of all lines which intersect the current cluster. It is not obvious how to choose a line in a fair way since most of the time the cluster will be strongly non symmetric. We will introduce a possible solution for this problem through the parametrisation of lines as defined in 5.1.2 for the 2-dimensional case which goes hand in hand with a more general result of integral geometry which we will touch first.

### 5.2.1 General results

Let  $d \in \mathbb{N}$  and  $q \in \{0, \dots, d\}$ . In the general context we are in  $\mathbb{R}^d$  and consider  $q$ -dimensional affine subspaces, short  $q$ -flats. The set of  $q$ -flats in  $\mathbb{R}^d$  we denote with  $A(d, q)$ . Later we will be interested in choosing random lines in the plane (i.e. 1-flats in  $\mathbb{R}^2$ ). In order to get a probability measure on some set of  $q$ -flats, we first need a measure and a  $\sigma$ -algebra on  $A(d, q)$  in total. Recall that we denote the set of convex and compact subsets of  $\mathbb{R}^d$  with  $\mathcal{K}^d$ . Also recall that with  $G_d$  we denote the set of all Euclidean motions (more at 2.1).

**Definition 5.2.1.** For  $B \in \mathcal{B}^d$  define

$$[B]_{d,q} := \{F \in A(d, q) \mid F \cap B \neq \emptyset\}.$$

If the context is clear, we will only write  $[B]$  instead of  $[B]_{d,q}$ .

**Definition 5.2.2.** The  $\sigma$ -algebra  $\mathcal{A}(d, q)$  on  $A(d, q)$  is defined by

$$\mathcal{A}(d, q) := \sigma(\{[K] \mid K \in \mathcal{K}^d\}).$$

**Remark 5.2.1.** It can be shown that on  $A(d, q)$  there exists an unique  $G_d$ -invariant Radon measure  $\mu_q$  such that

$$\mu_q([B_d(0, 1)]) = \kappa_{d-q}, \quad (5.3)$$

where  $\kappa_n := \lambda_n(B_n(0, 1))$  is the  $n$ -dimensional Lebesgue measure of the  $n$ -dimensional unit ball for  $n \in \mathbb{N}$  and  $\kappa_0 := 1$ . We won't argue that further here. For more information see at [10].

### 5.2.2 Construction in the plane: Isotropic lines

For our special case we choose  $d = 2$  and  $q = 1$ , thus lines in the plane. We want to show in the following that the way we defined a  $\sigma$ -algebra on  $\mathcal{G}$  in Definition 5.1.2 is equivalent to the general Definition 5.2.2. To do that it is convenient to use a special generator set for the Borel- $\sigma$ -algebra  $\mathcal{B}_\Phi$ .

**Lemma 5.2.1.** Define

$$\mathcal{R}_+ := \{[\alpha, \beta] \times (0, b) \mid 0 \leq \alpha < \beta < \pi, b \geq 0\},$$

$$\mathcal{R}_- := \{[\alpha, \beta] \times [b, 0) \mid 0 \leq \alpha < \beta < \pi, b \leq 0\},$$

$$\mathcal{R}_0 := \{[\alpha, \beta] \times \{0\} \mid 0 \leq \alpha < \beta < \pi\},$$

and

$$\mathcal{R} := \mathcal{R}_+ \cup \mathcal{R}_- \cup \mathcal{R}_0.$$

Then  $\sigma(\mathcal{R}) = \mathcal{B}_\Phi$ .

## 5 Line hitting aggregation

*Proof.* We show that  $\sigma(\mathcal{R})$  contains all rectangles in  $\Phi$  of the form  $[\alpha, \beta] \times (a, b)$  with  $a, b > 0$ ,  $[\alpha, \beta] \times [a, b]$  with  $a, b < 0$  and  $[\alpha, \beta] \times [a, b]$  with  $0 \in [a, b]$ . First let  $a, b > 0$  and  $R = [\alpha, \beta] \times (a, b)$ , then

$$R = ([\alpha, \beta] \times (0, b)) \setminus ([\alpha, \beta] \times (0, a))$$

and therefore  $R \in \sigma(\mathcal{R}_+) \subset \sigma(\mathcal{R})$ . Similarly it works if  $a, b < 0$ . If  $0 \in [a, b]$  then we can write  $R = [\alpha, \beta] \times [a, b]$  with three components  $R = [\alpha, \beta] \times [a, 0) \cup [\alpha, \beta] \times \{0\} \cup [\alpha, \beta] \times (0, b]$  which lie in  $\mathcal{R}_-, \mathcal{R}_0$  and  $\mathcal{R}_+$  respectively. Therefore  $R \in \sigma(\mathcal{R})$  as well. By measure theory the above described rectangles form a generator set of  $\mathcal{B}_\Phi$ , which completes the proof.  $\square$

**Lemma 5.2.2.** We have  $\mathcal{A}(2, 1) = \mathfrak{G}$ .

*Proof.* We will consider generators of these  $\sigma$ -algebras. By Lemma 5.2.1 we know that  $\sigma(\mathcal{R}) = \mathcal{B}_\Phi$  and since  $\chi$  is a bijection, we have  $\chi(\sigma(\mathcal{R})) = \sigma(\chi(\mathcal{R}))$  and finally  $\mathfrak{G} = \sigma(\chi(\mathcal{R}))$ . For  $\tilde{A} := \{[K] \mid K \in \mathcal{K}^2\}$  we have by definition  $\mathcal{A}(2, 1) = \sigma(\tilde{A})$ .

$\subset$ : Let  $K \in \mathcal{K}^2$ . We will show that  $\chi^{-1}([K])$  is a closed set in  $\Phi$ . If that is the case we have  $\chi^{-1}([K]) \in \mathcal{B}_\Phi$ , therefore  $[K] \in \chi(\mathcal{B}_\Phi) = \mathfrak{G}$  and finally  $\mathcal{A}(2, 1) = \sigma(\tilde{A}) \subset \mathfrak{G}$ . To show that  $\chi^{-1}([K])$  is closed let  $(\alpha_0, p_0) \in \Phi \setminus \chi^{-1}([K])$ . Then  $\chi(\alpha_0, p_0) \notin [K]$  and therefore  $\chi(\alpha_0, p_0) \cap K = \emptyset$ . Since  $K$  is closed we can find small values  $\tilde{\alpha}, \tilde{p} > 0$  such that  $\chi(\alpha, p) \cap K = \emptyset$  for all  $(\alpha, p) \in [\alpha_0, \alpha_0 + \tilde{\alpha}] \times [p_0, p_0 + \tilde{p}] =: R$ . Hence we have  $R \subset \Phi \setminus \chi^{-1}([K])$ , so  $\Phi \setminus \chi^{-1}([K])$  is open. Hence  $\chi^{-1}([K])$  is closed.

$\supset$ : For this inclusion we will show that  $\chi(R) \in \mathcal{A}(2, 1)$  for all  $R \in \mathcal{R}$ . First let  $R = [\alpha, \beta] \times (0, b) \in \mathcal{R}_+$  for some  $b > 0$ . Define

$$S := \{pe_\gamma \in \mathbb{C} \mid (\gamma, p) \in R\}$$

Furthermore for  $n \in \mathbb{N}$  define

$$A_n := \{tns_\beta \mid t \in [0, 1]\} \text{ and } B_n := \{-tns_\alpha \mid t \in [0, 1]\},$$

the segments from 0 to  $ns_\beta$  and 0 to  $-ns_\alpha$  (Figure 7). We will show now that

$$\chi(R) = [\bar{S}] \setminus (\bigcup_{n \in \mathbb{N}} [A_n] \cup \bigcup_{n \in \mathbb{N}} [B_n]) =: \tilde{S},$$

where  $\bar{S}$  is the closure of  $S$  as a set in  $\mathbb{C}$  (note that  $\bar{S} = S \cup \{0\}$ ). Let  $(\gamma, p) \in R$ . Then  $pe_\gamma \in \chi(\gamma, p) \cap S$  and therefore  $\chi(\gamma, p) \in [\bar{S}]$ . Assume that there exists an  $n \in \mathbb{N}$  such that  $\chi(\gamma, p) \cap A_n \neq \emptyset$ . Then  $\beta + \frac{\pi}{2} - \gamma < \frac{\pi}{2}$ , hence  $\beta < \gamma$ , a contradiction. Similarly argument for any  $B_n$ , so we finally have  $\chi(\gamma, p) \notin [A_n]$  and  $\chi(\gamma, p) \notin [B_n]$  for any  $n \in \mathbb{N}$ . Hence  $\chi(\gamma, p) \in \tilde{S}$  and therefore  $\chi(R) \subset \tilde{S}$ .

Now let  $(\gamma, p) \in \Phi$  such that  $\chi(\gamma, p) \in \tilde{S}$ . Assume that  $\gamma \notin [\alpha, \beta]$  then with a similar argument as in the first inclusion it is easy to see that there must be an  $n \in \mathbb{N}$  such

## 5 Line hitting aggregation

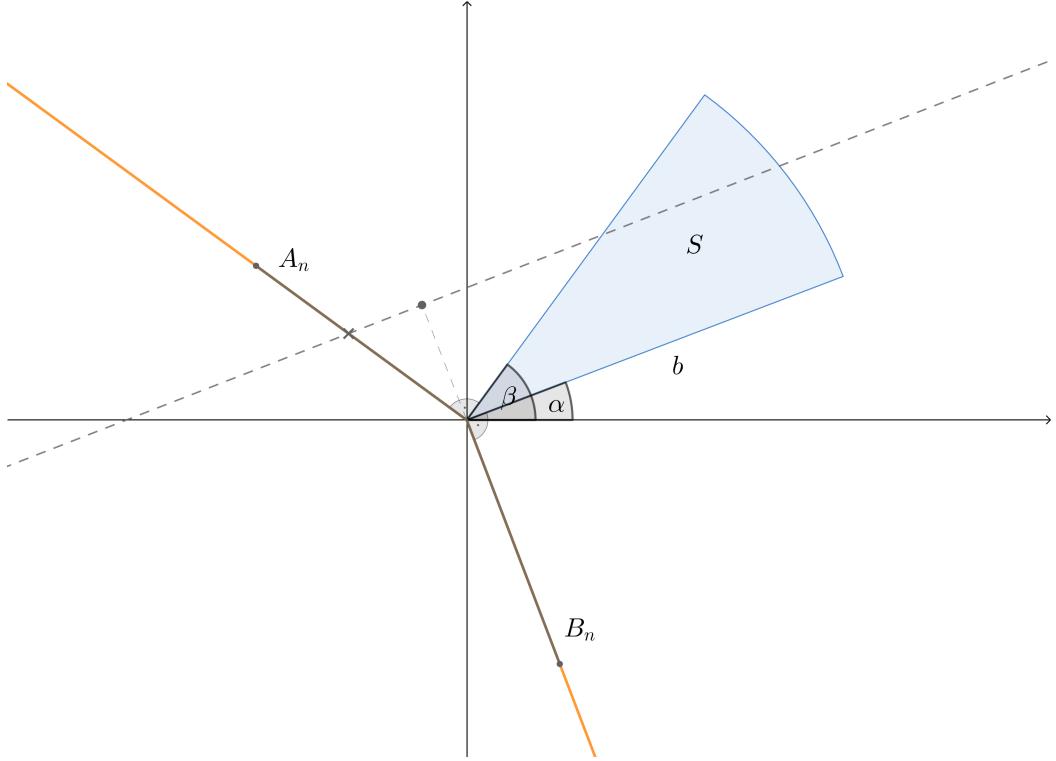


Figure 7:  $S$  and the sets  $A_n$  and  $B_n$

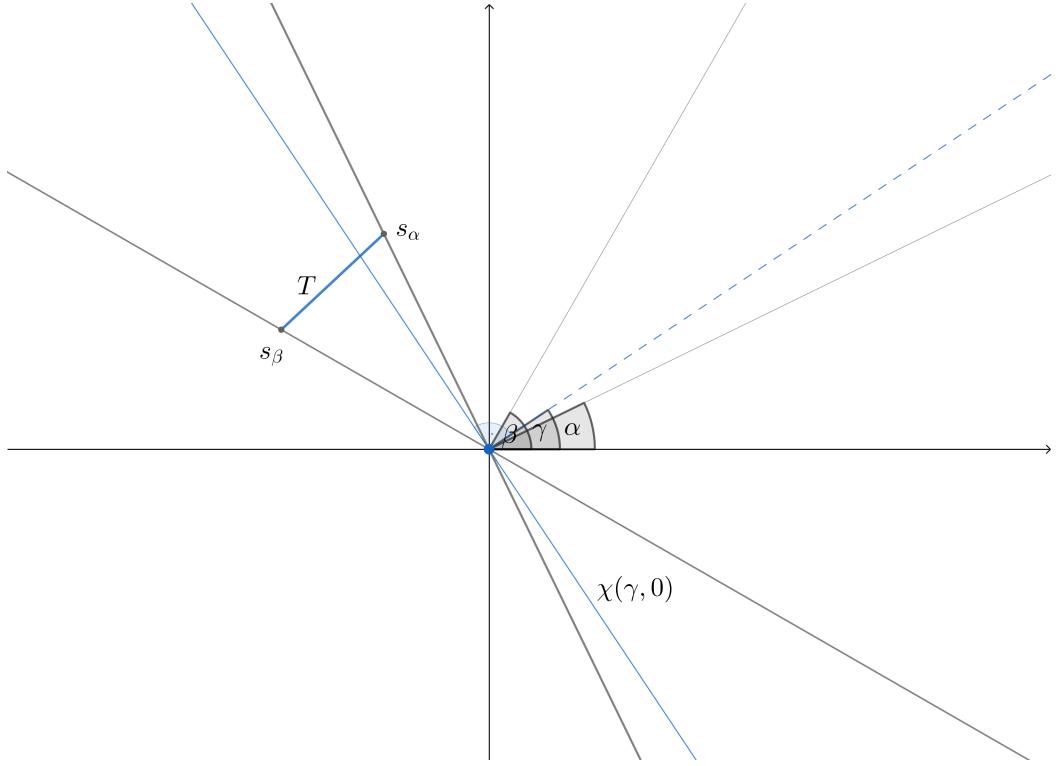
that  $\chi(\gamma, p) \cap A_n \neq \emptyset$  or  $\chi(\gamma, p) \cap B_n \neq \emptyset$ , a contradiction. Therefore  $\gamma \in [\alpha, \beta]$ . Now assume  $p \notin (0, b]$ . If  $p > b$  then  $\chi(\gamma, b) \cap B_b = \emptyset$  and since  $\bar{S} \subset B_b$  it is  $\chi(\gamma, p) \notin [\bar{S}]$ , a contradiction. If  $p < 0$  then, since the angle between the segments  $A_n$  and  $B_n$  opposite of  $S$  is strictly smaller than  $\pi$ ,  $\chi(\gamma, p)$  must intersect with  $A_n$  or  $B_n$  for some  $n \in \mathbb{N}$ , again a contradiction. Note that  $p \neq 0$  since  $\{0\} \subset [A_1]$ . Thus we have  $p \in (0, b]$ . Therefore we have  $(\gamma, p) \in R$  and finally  $\tilde{S} \subset \chi(R)$ .

It is left to show that  $\tilde{S} \in \mathcal{A}(2, 1)$ . All the segments  $A_n$  and  $B_n$  are compact and convex for all  $n \in \mathbb{N}$ , and since  $S$  is bounded and convex as a circle segment with angle smaller than  $\pi$ ,  $\bar{S}$  is compact and convex. Finally  $\tilde{S} \in \mathcal{A}(2, 1)$ .

In total we get  $\mathcal{R}_+ \subset \mathcal{A}(2, 1)$ .  $\mathcal{R}_- \subset \mathcal{A}(2, 1)$  can be shown analogously. So for the last case let  $R = [\alpha, \beta] \times \{0\} \in \mathcal{R}_0$ . Define the line segment

$$T := \{(1-t)s_\alpha + ts_\beta \mid t \in [0, 1]\}.$$

Then  $\{0\} \cap [T] \in \mathcal{A}(2, 1)$ . We show that  $\chi(R) = \{0\} \cap [T]$ . Let  $(\gamma, 0) \in R$ . Then  $\chi(\gamma, 0)$  contains 0 and since  $\gamma$  is in between  $\alpha$  and  $\beta$ ,  $\chi(\gamma, 0)$  must intersect with  $T$  (Figure 8). For the other inclusion let  $(\gamma, p) \in \Phi$  such that  $\chi(\gamma, p) \in \{0\} \cap [T]$ . Since  $0 \in \chi(\gamma, p)$  it must be  $p = 0$  and since it intersects with  $T$  its angle must lay in between  $\alpha$  and  $\beta$ . All in all this completes the proof.  $\square$


 Figure 8: Line in  $\{0\} \cap [T]$ 

**Definition 5.2.3.** A  $\mathcal{F}$ - $\mathfrak{G}$ -measurable function  $g : \Omega \rightarrow \mathcal{G}$  is called a *random line*.

**Definition 5.2.4.** We define the measure  $\mu := \lambda_{2|\Phi} \circ \chi^{-1}$  on  $(\mathcal{G}, \mathfrak{G})$  where  $\lambda_{2|\Phi}$  is the 2-dimensional Lebesgue measure restricted to  $\Phi$ . We say a measure  $\nu$  on  $(\mathcal{G}, \mathfrak{G})$  is locally finite if for any  $K \in \mathcal{K}^2$  we have  $\nu([K]) < \infty$ .

**Lemma 5.2.3.**  $\mu$  is locally finite and  $G_2$ -invariant.

*Proof.* Let  $K \in \mathcal{K}^2$  and since  $K$  is compact choose  $r > 0$  such that  $K \subset B_r$ . Then we have  $[K] \subset [B_r]$  and

$$\mu([K]) \leq \mu([B_r]) = \lambda_{2|\Phi}(\chi^{-1}([B_r])) = \lambda_{2|\Phi}([0, \pi) \times [-r, r]) = 2\pi r < \infty,$$

hence  $\mu$  is locally finite. To show that  $\mu$  is  $G_2$ -invariant, that is Euclidean motion invariant, we must show it is translation and rotation invariant. First we clarify what exactly translation and rotation mean for lines. We denote  $x$  modulo  $r$  as  $(x)_r$ . For  $b \in \mathbb{C}$  and  $\beta \in [0, 2\pi)$  we define

$$T_b : \Phi \rightarrow \Phi, \quad (\alpha, p) \mapsto (\alpha, p + \langle e_\alpha, b \rangle) \tag{5.4}$$

and

$$D_\beta : \Phi \rightarrow \Phi, \quad (\alpha, p) \mapsto ((\alpha + \beta)_\pi, \delta((\alpha + \beta)_{2\pi})p), \tag{5.5}$$

## 5 Line hitting aggregation

where

$$\delta : [0, 2\pi) \rightarrow \{-1, 1\}, \quad \gamma \mapsto \begin{cases} 1, & \gamma \in [0, \pi) \\ -1, & \gamma \in [\pi, 2\pi) \end{cases}.$$

It is easy to see that both functions are well-defined. The parameter transformation  $T_b$  defines a translation by  $b$  of the corresponding line, and analogously  $D_\beta$  a rotation by  $\beta$ . We prove that quickly here. Let  $(\alpha, p) \in \Phi$ , then

$$\begin{aligned} x \in \chi(\alpha, p) + b &\Leftrightarrow x - b \in \chi(\alpha, p) \\ &\Leftrightarrow \langle e_\alpha, x - b \rangle = p \\ &\Leftrightarrow \langle e_\alpha, x \rangle = p + \langle e_\alpha, b \rangle \\ &\Leftrightarrow x \in \chi(\alpha, p + \langle e_\alpha, b \rangle) \\ &\Leftrightarrow x \in \chi(T_b(\alpha, p)) \end{aligned}$$

and therefore  $\chi(\alpha, p) + b = \chi(T_b(\alpha, p))$ . Hence  $T_b(\alpha, p)$  are indeed the parameters of the line translated by  $b$ . For the rotation we look at two cases. First let  $(\alpha + \beta)_{2\pi} \in [0, \pi)$ , then  $\delta((\alpha + \beta)_{2\pi}) = 1$  and  $(\alpha + \beta)_\pi = \alpha + \beta$  and therefore

$$D_\beta(\alpha, p) = (\alpha + \beta, p).$$

For the second case let  $(\alpha + \beta)_{2\pi} \in [\pi, 2\pi)$ , then we have  $\delta((\alpha + \beta)_{2\pi}) = -1$  and  $(\alpha + \beta)_\pi = \alpha + \beta - \pi$  and therefore

$$D_\beta(\alpha, p) = (\alpha + \beta - \pi, -p).$$

In the second case we have to carefully understand the parametrisation of  $\mathcal{G}$ , but finally we can see that  $D_\beta(\alpha, p)$  are indeed the parameters of the line rotated by  $\beta$ .

We will further show now, that  $\mu$  is invariant with respect to both these functions. Let  $A \in \mathfrak{G}$ ,  $b \in \mathbb{C}$  and  $\nu_\beta \in SO_2$  for some  $\beta \in [0, 2\pi)$ . We will understand  $A + b = \{g + b \mid g \in A\}$  and  $\nu_\beta A = \{\nu_\beta g \mid g \in A\}$  pointwise, and  $g + b = \{x + b \mid x \in g\}$  and  $\nu_\beta g = \{\nu_\beta x \mid x \in g\}$  pointwise as well. We furthermore define  $A_p := \{\alpha \in [0, \pi) \mid (\alpha, p) \in \chi^{-1}(A)\}$  and  $A_\alpha := \{p \in \mathbb{R} \mid (\alpha, p) \in \chi^{-1}(A)\}$  for  $(\alpha, p) \in \Phi$ . For the translation by  $b$  we get

$$\begin{aligned} \mu(A + b) &= \int_{\mathfrak{G}} \mathbb{1}_{A+b}(g) \mu(dg) \\ &= \int_{\mathfrak{G}} \mathbb{1}_A(g - b) \mu(dg) \\ &= \int_{\mathfrak{G}} \mathbb{1}_A(g - b) \lambda_{2|\Phi}(\chi^{-1}(dg)) \\ &= \int_{\chi^{-1}(\mathfrak{G})} \mathbb{1}_{\chi^{-1}(A)}(\chi^{-1}(g - b)) \lambda_{2|\Phi}(d(\chi^{-1}(g))) \\ &\stackrel{(5.4)}{=} \int_{\Phi} \mathbb{1}_{\chi^{-1}(A)}(\alpha, p - \langle e_\alpha, b \rangle) \lambda_2(d(\alpha, p)) \end{aligned}$$

## 5 Line hitting aggregation

$$\begin{aligned}
&= \int_0^\pi \int_{\mathbb{R}} \mathbb{1}_{A_\alpha}(p - \langle e_\alpha, b \rangle) \lambda_1(dp) \lambda_1(d\alpha) \\
&= \int_0^\pi \int_{\mathbb{R}} \mathbb{1}_{A_\alpha + \langle e_\alpha, b \rangle}(p) \lambda_1(dp) \lambda_1(d\alpha) \\
&= \int_0^\pi \lambda_1(A_\alpha + \langle e_\alpha, b \rangle) \lambda_{1|[\alpha, \pi]}(d\alpha) \\
&\stackrel{(+) \equiv}{=} \int_0^\pi \lambda_1(A_\alpha) \lambda_1(d\alpha) \\
&= \dots \\
&= \mu(A),
\end{aligned}$$

and for a rotation we get

$$\begin{aligned}
\mu(\nu_\beta A) &= \int_{\mathcal{G}} \mathbb{1}_{\nu_\beta A}(g) \mu(dg) \\
&= \int_{\mathcal{G}} \mathbb{1}_A(\nu_{-\beta} g) \lambda_{2|\Phi}(\chi^{-1}(dg)) \\
&= \int_{\chi^{-1}(\mathcal{G})} \mathbb{1}_{\chi^{-1}(A)}(\chi^{-1}(\nu_{-\beta} g)) \lambda_{2|\Phi}(d(\chi^{-1}(g))) \\
&\stackrel{(5.5)}{=} \int_{\Phi} \mathbb{1}_{\chi^{-1}(A)}((\alpha - \beta)_\pi, \delta((\alpha - \beta)_{2\pi})p) \lambda_{2|\Phi}(\alpha, p) \\
&= \int_{\Phi} \mathbb{1}_{\chi^{-1}(A)}((\alpha - \beta)_\pi, \delta((\alpha - \beta)_{2\pi})p) \lambda_{2|\Phi}(\alpha, p) \\
&= \int_0^{2\pi} \int_{\mathbb{R}} \mathbb{1}_{\chi^{-1}(A)}((\alpha - \beta)_\pi, \delta((\alpha - \beta)_{2\pi})p) \lambda_1(dp) \lambda_{1|[\alpha, \pi]}(d\alpha) \\
&= \int_0^{2\pi} \int_{\mathbb{R}} \mathbb{1}_{A_{(\alpha - \beta)_\pi}}(\delta((\alpha - \beta)_{2\pi})p) \lambda_1(dp) \lambda_{1|[\alpha, \pi]}(d\alpha) \\
&\stackrel{(+) \equiv}{=} \int_0^{2\pi} \int_{\mathbb{R}} \mathbb{1}_{A_{(\alpha - \beta)_\pi}}(p) \lambda_1(dp) \lambda_{1|[\alpha, \pi]}(d\alpha) \\
&= \int_{\mathbb{R}} \int_0^{2\pi} \mathbb{1}_{\chi^{-1}(A)}((\alpha - \beta)_\pi, p) \lambda_{1|[\alpha, \pi]}(d\alpha) \lambda_1(dp) \\
&= \int_{\mathbb{R}} \int_0^{2\pi} \mathbb{1}_{A_p}((\alpha - \beta)_\pi) \lambda_{1|[\alpha, \pi]}(d\alpha) \lambda_1(dp) \\
&\stackrel{(+) \equiv}{=} \int_{\mathbb{R}} \int_0^{2\pi} \mathbb{1}_{A_p}(\alpha) \lambda_{1|[\alpha, \pi]}(d\alpha) \lambda_1(dp) \\
&= \int_{\mathbb{R}} \int_0^{2\pi} \mathbb{1}_{\chi^{-1}(A)}(\alpha, p) \lambda_{1|[\alpha, \pi]}(d\alpha) \lambda_1(dp) \\
&= \dots \\
&= \mu(A).
\end{aligned}$$

where in (+) we used the translation and rotation invariance of the Lebesgue measure. This completes the proof.  $\square$

**Example 5.2.1.** Choose  $K := B_r \in \mathcal{K}^2$  to be the ball with radius  $r > 0$  and center 0. Then

$$\mu([K]) = \lambda_{2|\Phi} \circ \chi^{-1}([B_r]) = \lambda_{2|\Phi}([0, \pi] \times [-r, r]) = 2\pi r.$$

**Example 5.2.2.** Now choose  $K := [0, r] \times \{0\} \in \mathcal{K}^2$  to be a segment of length  $r > 0$ . Then

$$\begin{aligned}\mu([K]) &= \lambda_{2|\Phi} \circ \chi^{-1}([K]) \\ &= \int_{\Phi} \mathbb{1}_{\chi^{-1}([K])}(\alpha, p) \lambda_2(d(\alpha, p)) \\ &= \int_0^\pi \int_{-\infty}^\infty \mathbb{1}_{\chi^{-1}([K])}(\alpha, p) dp d\alpha \\ &= \int_0^{\frac{\pi}{2}} \int_0^{r \cos(\alpha)} 1 dp d\alpha + \int_{\frac{\pi}{2}}^\pi \int_{-r \cos(\pi - \alpha)}^0 1 dp d\alpha \\ &= \int_0^{\frac{\pi}{2}} r \cos(\alpha) d\alpha + \int_{\frac{\pi}{2}}^\pi r \cos(\pi - \alpha) d\alpha \\ &= r\left(\sin\left(\frac{\pi}{2}\right) - \sin(\pi - \pi) + \sin\left(\pi - \frac{\pi}{2}\right)\right) \\ &= r(1 - 0 + 1) = 2r.\end{aligned}$$

In Remark 5.2.1 we have mentioned that it can be shown that  $\mu$  is, up to a factor, the only Euclidean motion invariant measure on  $\mathcal{G}$ . Even without presenting the proof here, being locally finite and Euclidean motion invariant as we have shown it in Lemma 5.2.3 gives  $\mu$  very strong properties which we would expect of a sensible and fair measure on lines. Since it is locally finite, we can therefore for any  $K \in \mathcal{K}^2$  define a probability measure on  $\mathcal{G}$  by

$$\mathbb{P}_\mu^K(A) := \frac{\mu(A \cap [K])}{\mu([K])}, \quad A \in \mathfrak{G}.$$

For  $A \in \mathfrak{G}$  this describes the probability that a random line which hits  $K$  lies in  $A$ .

**Definition 5.2.5.** Let  $K \in \mathcal{K}^2$ . A random line  $g : \Omega \rightarrow \mathcal{G}$  is called *K-isotropic* if

$$\mathbb{P}(g \in A) = \mathbb{P}_\mu^K(A), \quad A \in \mathfrak{G}.$$

**Lemma 5.2.4.** Let  $M, K \in \mathcal{K}^2$  with  $M \subset K$ . Let  $f$  be a random  $K$ -isotropic and  $g$  be a random  $M$ -isotropic line. Then for all  $A \in \mathfrak{G}$  we have

$$\mathbb{P}(f \in A \mid f \in [M]) = \mathbb{P}(g \in A).$$

*Proof.* Note that since  $M \subset K$  it is  $[M] \subset [K]$ . For  $A \in \mathfrak{G}$  we therefore directly get

$$\begin{aligned}\mathbb{P}(f \in A \mid f \in [M]) &= \frac{\mathbb{P}(f \in A \cap [M])}{\mathbb{P}(f \in [M])} \\ &= \frac{\mu(A \cap [M] \cap [K])}{\mu([K])} \frac{\mu([K])}{\mu([M] \cap [K])} \\ &= \frac{\mu(A \cap [M])}{\mu([M])} \\ &= \mathbb{P}(g \in A).\end{aligned}$$

□

## 5 Line hitting aggregation

If we choose a simple convex set such as  $K = B_r$ , the ball around the origin with radius  $r$ , then choosing random  $K$ -isotropic lines becomes a very intuitive and easy realizable task as the following lemma shows.

**Lemma 5.2.5.** Let  $K = B_r \in \mathcal{K}^2$  and let  $(\alpha, p)$  be uniformly distributed in  $\tilde{\Phi} := [0, \pi) \times [-r, r] = \chi^{-1}([K]) \subset \Phi$ . Then  $\chi(\alpha, p)$  is a random  $K$ -isotropic line.

*Proof.* For  $A \in \mathfrak{G}$  we get

$$\begin{aligned}\mathbb{P}(\chi(\alpha, p) \in A) &= \mathbb{P}((\alpha, p) \in \chi^{-1}(A)) = \frac{\lambda_{2|\tilde{\Phi}}(\chi^{-1}(A) \cap \chi^{-1}([K])))}{\lambda_{2|\tilde{\Phi}}(\chi^{-1}([K])))} \\ &= \frac{\lambda_{2|\Phi}(\chi^{-1}(A \cap [K])))}{\lambda_{2|\Phi}(\chi^{-1}([K])))} = \frac{\mu(A \cap [K])}{\mu([K])}.\end{aligned}$$

□

**Remark 5.2.2.** Both Lemmas 5.2.4 and 5.2.5 give a help for realizing  $K$ -isotropic lines for complicated sets  $K$ . Lemma 5.2.4 tells us that if we are looking for a  $K$ -isotropic line, we can actually take a convex, compact set  $B$  which contains  $K$  and realize  $B$ -isotropic lines. If we realize such a line and it happens that it intersects  $K$ , we know that it is a realization of a  $K$ -isotropic line. And how to realize  $B$ -isotropic lines? Lemma 5.2.5 tells us that if we choose  $B = B_r$ , a ball with a big enough radius such that it contains  $K$ , then a  $B$ -isotropic line is realized by choosing the line parameters  $\alpha$  and  $p$  uniformly in  $[0, \pi)$  and  $[-r, r]$ , respectively. Finally we have a practicable process of choosing random  $K$ -isotropic lines, even if  $K$  happens to be very asymmetric and complicated. But note that  $K$  has to be convex for this construction. This gives the base to define a new incremental aggregation in the next section.

## 5.3 Definition

We are now able to choose a line randomly out of all lines hitting a cluster in a fair and agreeable way, which are  $K$ -isotropic lines as presented in Definition 5.2.5. All in all we have realized the mathematical structure to define the incremental aggregation as we have planned it in the motivation in the beginning of this chapter. Recall that for a bounded subset  $B \subset \mathbb{C}$  the convex hull  $\text{conv}(B)$  of  $B$  is defined to be the smallest convex set containing  $B$ , formally

$$\text{conv}(B) := \bigcap_{B \subset K \in \mathcal{K}^2} K \in \mathcal{K}^2.$$

Recall also the definition

$$[B] := \{g \in \mathcal{G} \mid g \cap B \neq \emptyset\}$$

for Borel sets  $B \in \mathcal{B}^2$  and

$$g^A := \{p \in A \mid g \cap sq(p) \neq \emptyset\}$$

for  $A \in \mathcal{P}_f$ .

## 5 Line hitting aggregation

**Definition 5.3.1.** (*Random line hitting distribution*) Let  $A \in \mathcal{P}_f$  and  $K := \text{conv}(sq(A))$ . For  $x \in \mathbb{Z}^2$  and  $g \in \mathcal{G}$  define

$$\gamma_A(g) := \begin{cases} 0, & \text{if } |g^A| = 0, \\ 1, & \text{if } |g^A| = 1, \\ \frac{1}{2}, & \text{if } |g^A| \geq 2, \end{cases}$$

and

$$\tilde{\mu}_A(x, g) := \begin{cases} \gamma_A(g), & \text{if } g^A \neq \emptyset \text{ and } x \in \{\min_{\triangleleft}(g^A), \max_{\triangleleft}(g^A)\}, \\ 0, & \text{else.} \end{cases} \quad (5.6)$$

For fixed  $A \in \mathcal{P}_f$  and  $g \in \mathcal{G}$ ,  $\tilde{\mu}_A(x, g)$  assigns probabilities for each value  $x \in A$  which is 0 if  $g^A$  is empty, also 0 if  $x$  is not minimum or maximum of  $g^A$ ,  $\frac{1}{2}$  if  $x$  is minimum or maximum and those two values are different and finally 1 if  $x$  is minimum and maximum, thus  $g^A = \{x\}$ . We then define the following distribution on  $A$  by

$$\mu_A(x) := \frac{1}{\mathbb{P}_{\mu}^K([sq(A)])} \int_{\mathcal{G}} \tilde{\mu}_A(x, g) \mathbb{P}_{\mu}^K(dg). \quad (5.7)$$

We quickly show that  $\mu_A \in \mathcal{D}_A$  (recall Definition 3.1). For all  $x \in \mathbb{Z}^2 \setminus A$  and  $g \in \mathcal{G}$  we have  $\tilde{\mu}_A(x, g) = 0$  and therefore  $\mu_A(x) = 0$ . Furthermore for all  $g \in \mathcal{G}$  we have

$$\sum_{x \in A} \tilde{\mu}_A(x, g) = \begin{cases} 0, & |g^A| = 0, \\ 1, & |g^A| = 1, \\ \frac{1}{2}, & |g^A| \geq 2. \end{cases} = \mathbb{1}_{[sq(A)]}(g)$$

and therefore

$$\sum_{x \in A} \mu_A(x) = \frac{1}{\mathbb{P}_{\mu}^K([sq(A)])} \int_{\mathcal{G}} \mathbb{1}_{[sq(A)]}(g) \mathbb{P}_{\mu}^K(dg) = \frac{\mathbb{P}_{\mu}^K([sq(A)])}{\mathbb{P}_{\mu}^K([sq(A)])} = 1.$$

Hence, the family of distributions  $(\mu_A)_{A \in \mathcal{P}_f}$  defines an incremental aggregation.

**Definition 5.3.2.** (*Line hitting aggregation*) An incremental aggregation with the random line hitting distribution we call *line hitting aggregation*, short *LHA*.

**Remark 5.3.1.** Let  $A \in \mathcal{P}_f$ . If  $A$  is connected as a subgraph of  $\mathbb{Z}^2$  then with  $K := \text{conv}(sq(A))$  it is easy to see that  $[sq(A)] = [K]$ , leading to  $\mathbb{P}_{\mu}^K([sq(A)]) = 1$ . Therefore we get

$$\begin{aligned} \mu_A(x) &= \frac{1}{\mathbb{P}_{\mu}^K([sq(A)])} \int_{\mathcal{G}} \tilde{\mu}_A(x, g) \mathbb{P}_{\mu}^K(dg) \\ &= \int_{[sq(x)]} \tilde{\mu}_A(x, g) \mathbb{P}_{\mu}^K(dg) + \int_{\mathcal{G} \setminus [sq(x)]} \tilde{\mu}_A(x, g) \mathbb{P}_{\mu}^K(dg) \\ &\leq \int_{[sq(x)]} 1 \mathbb{P}_{\mu}^K(dg) + 0 \\ &= \mathbb{P}_{\mu}^K([sq(x)]) \end{aligned}$$

for any  $x \in A$ .

## 5.4 Fractal dimension and growth rate of LHA in $\mathbb{Z}^2$

In this section we will try to give an idea of the fractal dimension of the line hitting aggregation.

**Theorem 5.4.1.** There exists a  $c > 0$  such that for all  $r \in [1, \infty)$  we have

$$\mu_A(0) \leq cr^{-1}, \quad \text{for all } A \in \mathcal{P}_r^0.$$

*Proof.* Let  $r \geq 1$ ,  $A \in \mathcal{P}_r^0$  and  $K := \text{conv}(sq(A))$ . Since  $A$  is connected we can use Remark 5.3.1 and the Examples 5.2.1 and 5.2.2 to get

$$\begin{aligned} \mu_A(0) &\leq \mathbb{P}_\mu^K([sq(0)]) \\ &= \frac{\mu([sq(0)])}{\mu([K])} \\ &\leq \frac{\mu([B_1])}{\mu([(0, r] \times \{0\})]} \\ &= \frac{2\pi}{2r} \\ &= \pi r^{-1}. \end{aligned}$$

Note that  $sq(0) \subset B_1$ . We used that since  $K$  is convex and has at least radius  $r$  as a set in  $\mathbb{C}$  it certainly contains a line segment of length  $r$ , and, since  $\mu$  is translation and rotation invariant, we can choose the segment  $[0, r] \times \{0\}$  for the computation of the denominator. Since choosing  $c := \pi$  does not depend on  $r$  or  $A$  this completes the proof.  $\square$

We are therefore prepared to proof the following theorem.

**Theorem 5.4.2.** For the fractal dimension of the line hitting aggregation we have

$$d_f = 2 \quad \text{a.s.}$$

*Proof.* It is not difficult to proof that the line hitting distribution is translation invariant and an outside distribution, just like the harmonic measure in external DLA. Recall the Definitions 3.1.3 and 3.1.2. Together with Theorem 5.4.1 we can therefore use Theorem 3.3.1 and choose  $q = 1$  to get

$$d_f \geq 1 + 1 = 2 \quad \text{a.s..}$$

Since  $d_f \leq 2$  a.s. by (3.7) we therefore get  $d_f = 2$  a.s..  $\square$

## 6 Python simulations

In this section we will present a simulation for the line hitting aggregation. The simulation is written in *python* and images are created with *pygame*. If you want to check and try the code yourself, feel welcomed to checkout this repository on github:

<https://github.com/tillmanntristanbosch/incremental-aggregations.git>

Each vertex in  $\mathbb{Z}^2$  can be naturally identified with values in  $\mathbb{C}$  and is represented in this way in this code. Every value in  $\mathbb{C}$  with integer components can be represented by a square as presented in Definition 5.1.1, and each square will be represented by exactly one pixel on a monitor screen, so every picture created by the simulation consists of pixels each naturally representing exactly one vertex in  $\mathbb{Z}^2$ . The aim of this simulation is to create realizations of LHA as close as possible to its mathematical definition.

Note that this software also contains a simulation for external DLA. This simulation doesn't claim to be implemented very close to its mathematical definition, but it is good enough to play around and create images with it, and get a vague visual comparison of the two aggregation models.

### 6.1 Code

In the following we will describe the code and claim that it is a precise implementation of LHA, excluding errors from python packages and machine calculation. In the simulation we are using imports for systemic purposes (**pygame**, **sys**, **json**, **time**, **datetime**, **os**, **shutil**) and for mathematical purposes (**random**, **math.sin**, **math.cos**, **math.pi**, **math.log**). Only the imports for mathematical purposes are used for the calculation of a realization of LHA. So on the side of import packages only these packages will have an effect on the mathematical precision of the realized process. Other than that every calculations precision is bounded to some extent by the precision of machine calculation. We will ignore a study on that in this simulation.

When using the software we will have the possibility to enter some parameters in the *parameters.json* file which include colorings of background and particle pixels, how many particles we want the cluster to have at the end and how big the created image shall be. An example of this file looks like this:

```
{
  "aggregation": "lha",
  "cluster_size": 10000,
  "background_color": "FFFFFF",
  "particle_color": ["000000"],
  "color_generation_size": 1000,
  "fractal_calculation_range": [0.2, 0.75],
  "image_width": 300,
  "image_height": 300
}
```

The value `color_generation_size` specifies how big one generation of a color shall be in the case that we choose more than one color to print the particles. So if we for example enter the values `["000000", "ff0000", "ffffff"]` for `"particle_color"` and `100` for `"color_generation_size"`, then the first 100 particles will be printed in color `"000000"` (black), the next 100 in `"ff0000"` (red), the next in `"ffffff"` (white) and the next again in black, and so on. This allows to visualize layers of the aggregation and produce beautiful rainbowlike colored realizations. Since for the simulation here we only use black particle color, the color generation size does not have any effect in this case. With the `fractal_calculation_range` we can have influence on how the approximation of the fractal dimension shall be calculated, more on that later. We can find more about setting parameters in the `README` file of the repository.

If correct parameters are set, you can run a shell script which in turn will run a single python file `main.py` which reads the entered parameters and executes first the mathematical process which needs the mathematical imports and second the data creation of the calculated process which needs the systemic imports as described above. The created data will contain an image of the created cluster, a json file with the parameters saved for later lookup and another json file containing information about the creation time, fractal dimension values and a list of all particle coordinates for potential later use.

We will have a look now on what happens on the mathematical side. We will show code directly here to make things clearer. Note that in the original code here and there you will find comments or other irrelevant differences, which we leave away here.

The main executed python file `main.py` looks like this:

```
if __name__ == "__main__":
    with open("parameters.json") as json_file:
        data = json.load(json_file)

    if data["aggregation"] == "lha":
        aggregation = lha.Line_Hitting_Aggregation()
    elif data["aggregation"] == "dla":
        aggregation = dla.External_DLA()
    else:
        raise Exception("Invalid aggregation parameter. Enter lha or dla.")

    aggregation.run_process(data["cluster_size"] - 1)
    cd.export_data(aggregation, data, separator, runtime)
```

Note again that we can also run an external DLA simulation but we do not claim that it is mathematically precise and therefore we do not discuss it here. The parameters in `parameters.json` are read and depending on which process (LHA or DLA) shall be simulated, the according object gets initiated and its `run_process` method gets executed. Both classes `Line_Hitting_Aggregation` and `External_DLA` base on the class `Incremental_Aggregation`. So the full initialization of LHA looks like:

```

class Incremental_Aggregation:
    def __init__(self):
        self.particles = [0]
        self.boundary_set = self.get_neighbours(0)
        self.cluster_radius = 0

class Line_Hitting_Aggregation(ia.Incremental_Aggregation):
    def __init__(self):
        super().__init__()

```

We initialize the particle set *particles* as a set containing 0, initialize the boundary set which in every iteration of the process will contain all neighbours of all particles in the particle set that are not in the particle set themselves and set the initial cluster radius to 0. As neighbours we only consider graph neighbours (up, down, left and right) as defined in 2.3.

After initializing a *Line\_Hitting\_Aggregation* object the method *run\_process* gets executed, which reads as follows:

```

def run_process(self, iterations):
    for k in range(iterations):
        line_hits_cluster = False

        while not line_hits_cluster:
            random_line = geom.Line(self.get_random_line())

            if self.line_hits_cluster(random_line):
                line_hits_cluster = True
                next_particle = self.get_next_particle(random_line)

                self.particles.append(next_particle)
                self.actualize_boundary_set()
                self.actualize_cluster_radius()

```

This method will directly run a loop with  $\text{data}[\text{"cluster\_size"}] - 1$  iterations. In each iteration exactly one particle will be added to the particle set, so that we end up with exactly  $\text{data}[\text{"cluster\_size"}]$  particles since we started with one particle.

If we call  $K$  the current cluster it is now important to recall again, how choosing random  $K$ -isotropic lines works. In 5.2.2 we have argued that a  $K$ -isotropic line can be realized by considering a ball  $B$  which contains the current cluster  $K$  and choose a  $B$ -isotropic line. If it happens that this line intersects with  $K$  then, by 5.2.4, we know that we have realized a random  $K$ -isotropic line. The method *get\_random\_line*

```

def get_random_line(self):
    alpha = pi * random.random()
    radius = self.cluster_radius + 2
    p = 2 * radius * random.random() - radius
    return (alpha, p)

```

produces a uniformly chosen pair of line parameters  $(\alpha, p)$ , and using  $cluster\_radius + 2$  here ensures that the resulting ball  $B$  does contain the cluster. The method `random.random()` produces an uniformly chosen value in  $[0, 1]$ . The fact that `random.random()` cannot choose the value 1 doesn't disturb the distribution of the line since  $\{1\}$  is a zero set for the 1-dimensional Lebesgue-measure. By 5.2.5, choosing these line parameters uniformly indeed realizes a random  $B$ -isotropic line.

After having realized a  $B$ -isotropic line by choosing its line parameters as described above, we want to check whether the line hits the cluster or not which is implemented in the method `line_hits_cluster`.

```
def line_hits_cluster(self, line):
    for k in range(len(self.particles)):
        if line.intersects_with_polygon(self.get_square(self.particles[-k])):
            return True
    return False
```

This method bases on the method `intersects_with_polygon` which evaluates whether a line intersects with a given general polygon or not (the implemented geometry classes `Line` and `Polygon` you can find in the file `geometry.py`). Here the polygon is a square given by the method

```
def get_square(self, pos):
    return geom.Polygon([pos + 1/2 * (1+1j), pos + 1/2 * (1-1j), pos + 1/2 * (-1-1j), pos + 1/2 * (-1+1j)])
```

which implements the Definitions 5.1.1 and 5.1.3.

If the condition is satisfied that a chosen line hits the cluster, we enter the if-condition and then choose the next particle with the method `get_next_particle`

```
def get_next_particle(self, line):
    hit_positions = self.get_boundary_hit_positions(line)
    min_position = self.get_min(line.alpha, hit_positions)
    max_position = self.get_max(line.alpha, hit_positions)
    return random.choice([min_position, max_position])
```

which first collects all particle squares in the boundary set which are hit by the line, calculates minimum and maximum of all those particles and then chooses uniformly between minimum and maximum with `random.choice`. The methods `get_min` and `get_max` base on the method `is_lower`

```

def is_lower(self, alpha, x, y):
    x_0, x_1 = x.real, x.imag
    y_0, y_1 = y.real, y.imag

    if alpha == pi/2: # Case 1
        return x_0 < y_0
    elif alpha == 0: # Case 2
        return x_1 < y_1
    elif pi/2 < alpha < pi: # Case 3
        if x_0 == y_0:
            return x_1 < y_1
        else:
            return x_0 < y_0
    elif 0 < alpha < pi/2: # Case 4
        if x_0 == y_0:
            return x_1 > y_1
        else:
            return x_0 < y_0

```

which implements the totally ordered relation on all squares which got hit by the line as defined in 5.1.3. Minimum and maximum are then canonically calculated with respect to this relation.

After having realized the next particle, values like the particle set, the boundary set and the cluster radius get updated and the next iteration starts. When all iterations are calculated, data like the image, the informations file and a copy of the parameters file are created which can be found in the folder *exports*. In the image file we can see the process printed in the colors you have specified in the parameters file. In the information file we can find date and time when the simulation was run, the running time of the simulation in seconds, the linear regression parameters, the last fractal dimension value and volume ratio as described in the next section, and finally the whole list of particles of the process for later use.

In the next section we look at some empirical values of LHA.

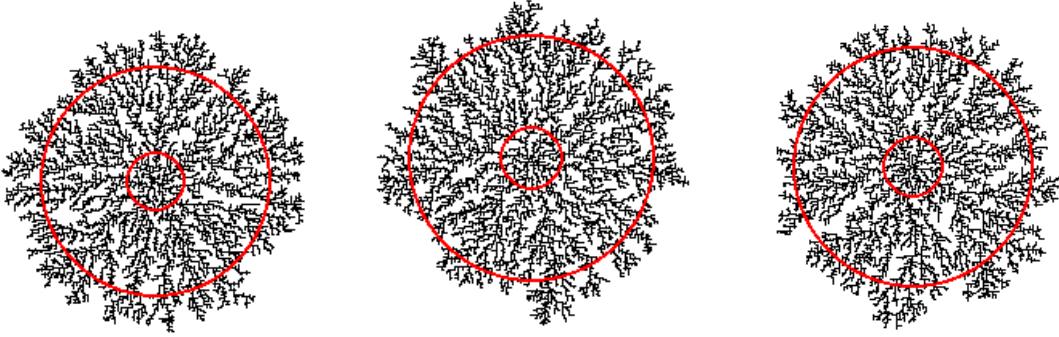
## 6.2 Empirical values of LHA

In this section we present what our simulations produce in terms of images and fractal dimension values. Since by Theorem 5.4.2 we already know that the fractal dimension of the line hitting aggregation is 2, our approximation for the fractal dimension here has the purpose to give an idea of the convergence behaviour towards the value 2. First it is worth to mention that it is not trivial to create a sensible approximation for the fractal dimension. If we go back to the definition

$$d_f = \liminf_{n \rightarrow \infty} \frac{\ln(n)}{\ln(\text{rad}(\mathcal{E}_n))}$$

we can start with the ansatz that

$$\text{rad}(\mathcal{E}_n) \approx cn^{\frac{1}{d_f}}$$


 Figure 9: circle radii range  $[0.2, 0.75]$  of the clusters radius

for some constant  $c > 0$  and therefore get

$$\ln(\text{rad}(\mathcal{E}_n)) \approx \frac{1}{d_f} \ln(n) + \ln(c). \quad (6.1)$$

Therefore we could be looking for the parameters  $\frac{1}{d_f}$  and  $\ln(c)$  of a linear regression of the two values  $\ln(n)$  and  $\ln(\text{rad}(\mathcal{E}_n))$  for a range of values  $n \in \{1, \dots, m\}$ ,  $m \in \mathbb{N}$ . In the range of up to 100000 particles we have noticed a strong instability of these parameters as functions of the maximum value  $m$ . We have come to the conclusion that for calculating approximations it is problematic that the radius of a cluster always stays constant for the time that new particles get added to the cluster where they don't increase the radius. Compare with Figure 4 and the discussion in Remark 3.2.1 to see that this cluster is not yet „filled up“ to its radius which creates something like a stair function in the linear plot of 6.1. To get more stable parameters it might make sense to come back to Conjecture 3.2.1 that

$$d_f = \limsup_{n \rightarrow \infty} \frac{\ln(|\mathcal{E}_\infty \cap B_n|)}{\ln(n)}.$$

With the ansatz

$$|\mathcal{E}_\infty \cap B_n| \approx c_1 n^{d_f} \quad (6.2)$$

for some constant  $c_1 > 0$  we would get

$$\ln(|\mathcal{E}_\infty \cap B_n|) \approx d_f \ln(n) + \ln(c_1). \quad (6.3)$$

Recall that  $\mathcal{E}_\infty \cap B_n$  is the set of all particles of the limit cluster  $\mathcal{E}_\infty$  which lie in  $B_n$ , the ball with radius  $n$  and center 0. Since it is not possible to realize  $\mathcal{E}_\infty$ , we have to use an approximation here. What we want to compare is the particle number of „filled up“ clusters with their radius. If we have a look at Figure 9 and let  $r$  be the radius of the cluster in the figure, then we can see that in a certain range of radii  $n \in [ar, br]$  with values  $0 \leq a < b \leq 1$  the clusters  $\mathcal{E}_\infty \cap B_n$  seem to be fairly filled up (this range is

## 6 Python simulations

described with the two red circles). This range of radii is what can be specified in the *parameters.json* file in the *fractal\_calculation\_range* field (see in the git repository). We found that it makes sense to not start the observation at 0 but a bit higher like  $a = 0.2$  to avoid distorted values occurring by the beginning of the process. On the other side we found empirically that  $b = 0.75$  seems to be a fair upper ratio where the clusters in most cases seem to be fairly filled up. Note that in the parameters file we should enter the values  $a$  and  $b$ , not  $ar$  and  $br$ . So a correct setting would be  $[0.2, 0.75]$ .

So all in all we will look at radii in the specified range  $n \in [0.2r, 0.75r]$  and count all particles of the cluster which lie in  $B_n$ , which shall be an approximation of  $|\mathcal{E}_\infty \cap B_n|$ . After that we calculate the linear regression parameters  $\hat{d}_f$  and  $\hat{c}_1$  for  $d_f$  and  $\ln(c_1)$ , respectively, as described in (6.3) which happens in the following method:

```
def fractal_dimension_approximation_1(particles, ratio_range = [0,1]):
    radius = max([abs(x + y*1j) for x,y in particles])

    min_radius, max_radius = ratio_range[0] * radius, ratio_range[1] * radius

    radius_list, cardinal_list = [], []

    radius_value = min_radius
    while radius_value < max_radius:
        radius_list.append(radius_value)
        cardinal_list.append(len([(x,y) for (x,y) in particles if abs(x + y*1j) < radius_value]))
        radius_value += 1

    x_list = [log(x) for x in radius_list]
    y_list = [log(x) for x in cardinal_list]

    if len(x_list) > 0:
        params = polyfit(x_list, y_list, 1)
        return {"parameters": (params[0], params[1]),
                "x_list": x_list,
                "y_list": y_list}
```

You will find every of the following realized simulations in the folder *exports* of the git repository. Inside a folder you will find three types of files, one is the created image, one contains information about the process as described at the end of the last section and the third is a copy of the parameters file the simulation was preset with. More on that you can find in the README file of the repository. The names of the folder and files will have a certain format to identify each simulation clearly. The folder name

2021\_03\_02\_02\_37\_54\_lha\_100000

indicates that the folder contains a realized LHA simulation with 100000 particles which was created at 02. March 2021 02:37:54. Analogously

2021\_03\_02\_11\_02\_25\_dla\_5000

is the folder of a realized simulation of DLA. For the following observation we have run the line hitting aggregation from 10000 to 100000 twice at each 10000 step which makes

## 6 Python simulations

Simulation name	$\hat{d}_f (d_f)$	$\hat{c}_1 (\ln(c_1))$	$\hat{c}_v (c_v)$
2021_03_04_09_40_51_lha_10000	1.8601	0.8083	0.4198
2021_03_04_10_05_53_lha_10000	1.8679	0.7847	0.4222
2021_03_04_11_53_34_lha_20000	1.9135	0.5899	0.4006
2021_03_04_12_35_01_lha_20000	1.9273	0.5265	0.3976
2021_03_04_13_54_14_lha_30000	1.9071	0.6370	0.4016
2021_03_04_14_49_08_lha_30000	1.9191	0.5734	0.3965
2021_03_04_15_56_37_lha_40000	1.9518	0.4164	0.3885
2021_03_04_18_00_02_lha_40000	1.9173	0.5874	0.3941
2021_03_04_15_58_28_lha_50000	1.9311	0.5127	0.3863
2021_03_04_18_00_18_lha_50000	1.9381	0.4664	0.3811
2021_03_04_12_35_20_lha_60000	1.9283	0.5328	0.3868
2021_03_04_09_42_12_lha_60000	1.9187	0.5771	0.3861
2021_03_04_09_42_36_lha_70000	1.9257	0.5467	0.3853
2021_03_04_13_43_38_lha_70000	1.9184	0.5707	0.3812
2021_03_04_09_42_56_lha_80000	1.9135	0.6230	0.3906
2021_03_04_14_49_16_lha_80000	1.9463	0.4352	0.3788
2021_03_04_16_30_00_lha_90000	1.9420	0.4632	0.3805
2021_03_04_09_43_16_lha_90000	1.9344	0.5083	0.3834
2021_03_03_21_23_50_lha_100000	1.9065	0.6421	0.3805
2021_03_03_21_23_47_lha_100000	1.9344	0.4989	0.3782

Table 1: LHA Simulations (10000 - 100000 particles)

20 simulations in total. The simulations have produced the values in Table 1. A plot of the values can be found in Figure 10. The plot indicates a rather slow convergence of the empirical dimension values towards the theoretical value 2. Something similar was guessed by [4]. For a further empirical study on this convergence behaviour more simulations would be necessary.

Another interesting value to calculate is the following. Since we know that the fractal dimension for LHA is 2, we can rewrite (6.2) to get

$$|\mathcal{E}_\infty \cap B_n| \approx c_2 n^2 \quad (6.4)$$

for some constant  $c_2 > 0$  which can be transformed to

$$\frac{c_2}{\pi} \approx \frac{|\mathcal{E}_\infty \cap B_n|}{\pi n^2}.$$

We can see that  $\frac{c_2}{\pi}$  describes the relation between the volume of the cluster in the ball  $B_n$  and the volume of the ball  $B_n$  itself. Hence  $\frac{c_2}{\pi}$  can be interpreted as the *volume fraction* of the limit cluster  $\mathcal{E}_\infty$  which can be formally defined as

$$c_v := \lim_{n \rightarrow \infty} \frac{|\mathcal{E}_\infty \cap B_n|}{\pi n^2},$$

## 6 Python simulations

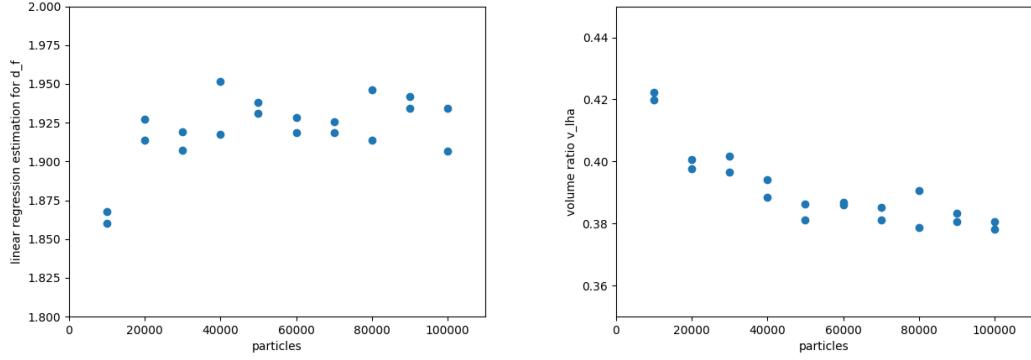


Figure 10: Linear regression parameters and volume ratio LHA

if the limit exists. If we look at the ansatz

$$\ln(|\mathcal{E}_\infty \cap B_n|) \approx 2 \ln(n) + \ln(c_2) \quad (6.5)$$

similar to linear regression we can calculate an in the sense of a minimal quadratic error optimal estimator  $\hat{c}_2$  for  $\ln(c_2)$  which satisfies

$$\hat{c}_2 = \underset{c \in \mathbb{R}}{\operatorname{argmin}} \sum_{n=m_1}^{m_2} (\ln(|\mathcal{E}_\infty \cap B_n|) - (2 \ln(n) + c))^2,$$

where  $m_1 := \min([0.2r, 0.75r] \cap \mathbb{N})$  and  $m_2 := \max([0.2r, 0.75r] \cap \mathbb{N})$ . It is not difficult to show that  $\hat{c}_2$  is given by

$$\hat{c}_2 = \frac{1}{m_2 - m_1 + 1} \sum_{n=m_1}^{m_2} \ln(|\mathcal{E}_\infty \cap B_n|) - 2 \ln(n).$$

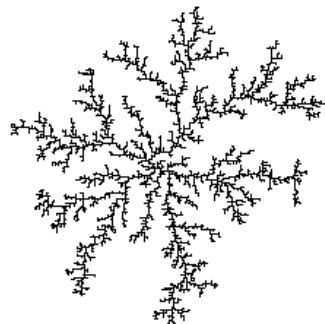
We can then obtain an estimator  $\hat{c}_v$  for the volume ratio  $c_v$  via

$$\hat{c}_v := \frac{e^{\hat{c}_2}}{\pi}.$$

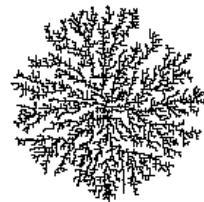
In Table 1 you can find the values for  $\hat{c}_v$  which have been produced by our simulations. Figure 10 indicates that the volume ratio decreases as the number of particles grows and possibly starts to stabilize at around  $c_v \approx 0.38$ . For a further study on the volume ratio again more simulations would be necessary.

### 6.3 Images

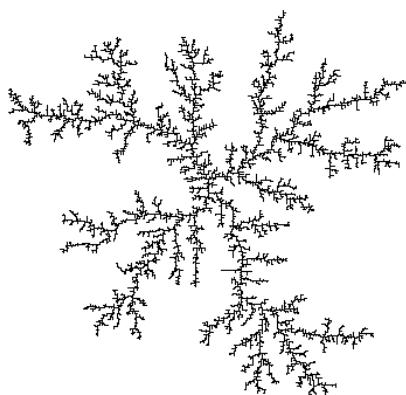
In the following we present some of the realized images of our LHA and external DLA simulations. More can be found in the repository in the folder *exports*.



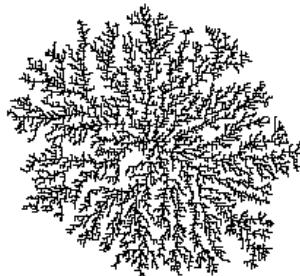
(a) 2021\_03\_03\_12\_25\_49\_\_image\_dla\_5000.png



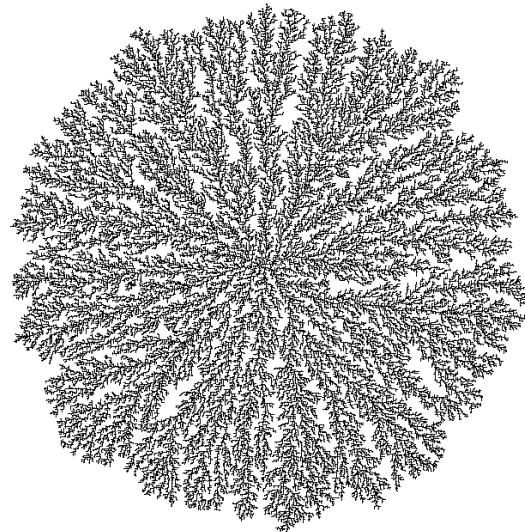
(b) 2021\_03\_06\_00\_23\_38\_\_image\_lha\_5000.png



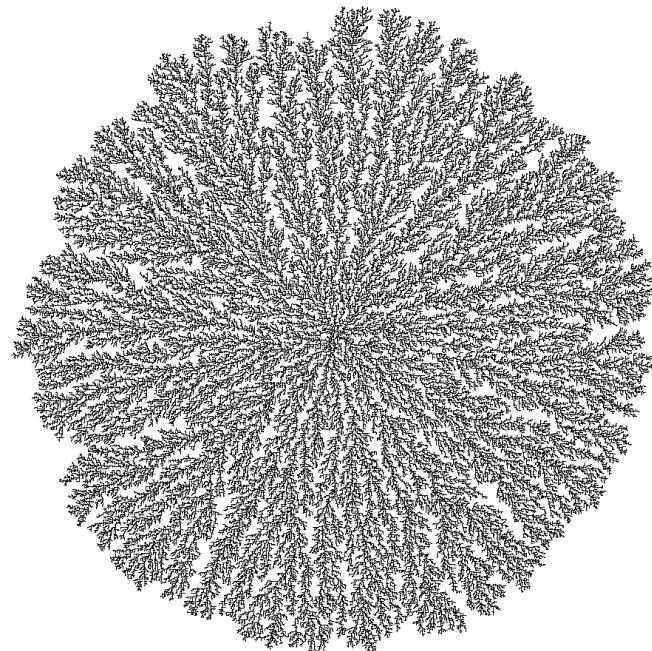
(a) 2021\_03\_19\_58\_26\_\_image\_dla\_10000.png



(b) 2021\_03\_23\_28\_53\_\_image\_lha\_10000.png



(a) 2021\_03\_03\_21\_23\_47\_\_image\_lha\_100000.png



(b) 2021\_03\_04\_00\_42\_19\_\_image\_lha\_200000.png

## 7 Review

In this paper we have looked at a general definition of incremental aggregations in  $\mathbb{Z}^d$  and the notion of a fractal dimension for such discrete clusters. We then moved our attention to  $\mathbb{Z}^2$  and found for this case in Theorem 3.3.1 a condition for proving a lower bound of the fractal dimension. We were able to apply this theorem for both incremental aggregations we looked at in this paper, external diffusion limited aggregation (external DLA) and the line hitting aggregation (LHA). For external DLA we found a lower bound of

$$d_f^{dla} \geq \frac{3}{2} \quad \text{a.s.}$$

and for the line hitting aggregation we were even able to find the exact dimension, which is

$$d_f^{lha} = 2 \quad \text{a.s..}$$

In the last chapter we provided simulations for both incremental aggregations. Especially the one for LHA was implemented as close to its mathematical definition as possible.

All in all did we here only touch the properties of incremental aggregations. The most known incremental aggregation you can find in literature is probably external DLA, but the general definition here in this paper opens a door to other interesting incremental aggregations.

I hope that I could raise interest in this topic through my work in this paper and I invite the reader to play around with the code I provided in a github repository (see at the beginning of Chapter 6). I feel delighted to watch at the beautiful realizations of the aggregations and I hope the reader can feel similar looking at them and realizing images him- and herself.

## References

- [1] T. A. Witten and L. M. Sander: *Diffusion-limited aggregation*, 1983  
DOI: <https://doi.org/10.1103/PhysRevB.27.5686>  
Download: 09.03.2021
- [2] Egbert Brieskorn: *Felix Hausdorff zum Gedächtnis I*, vieweg, 1996  
DOI: 10.1007/978-3-322-80276-7  
Download: 09.03.2021
- [3] Gregory F. Lawler: *Intersections of Random Walks*, Birkhäuser, 1996
- [4] D. Bensimon, B. Shraiman and S. Liang: *On the ballistic model of aggregation*, 1984  
DOI: [https://doi.org/10.1016/0375-9601\(84\)90701-1](https://doi.org/10.1016/0375-9601(84)90701-1)  
Download: 09.03.2021
- [5] J. Gonzalez-Gutierrez, J. L. Carrillo-Estrada and J. C. Ruiz-Suarez: *Aggregation and dendritic growth in a magnetic granular system*, 2013  
DOI: 10.1088/1742-5468/2013/12/P12015  
Download: 09.03.2021
- [6] Franz Sackmann: *Zufällige Geraden*, 2007  
Staatsexamensarbeit at University Karlsruhe (TH)
- [7] Ecaterina Sava-Huss, Uta Freiberg, Ben Hambly, Michael Hinz and Steffen Winter: *Fractal Geometry and Stochastics VI*, Preprint 2020  
Download: 09.03.2021  
[https://www.math.kit.edu/stoch/~winter/seite/pub/media/fgs6book\\_draft.pdf](https://www.math.kit.edu/stoch/~winter/seite/pub/media/fgs6book_draft.pdf)
- [8] Prof. Dr. Stefan Tappe: *Markov-Ketten*, 2018  
Lecture at Albert-Ludwigs-Universität Freiburg  
Download: 09.03.2021  
[https://www.stochastik.uni-freiburg.de/lehre/ss-2018/vorlesung-markov-ketten-ss-2018/markov\\_ketten\\_ss\\_2018.pdf](https://www.stochastik.uni-freiburg.de/lehre/ss-2018/vorlesung-markov-ketten-ss-2018/markov_ketten_ss_2018.pdf)
- [9] Norbert Henze: *Irrfahrten - Faszination der Random Walks*, Springer Spektrum, 2. Auflage, 2018
- [10] Rolf Schneider, Wolfgang Weil: *Stochastic and integral geometry*. Springer, Berlin, Heidelberg, 2008

## 8 Declaration

Herewith I assure that I created this paper self-reliant and on my own and that I did not use any material in this paper that I have not mentioned. The sources and contents I used in this paper I clearly labeled as such. I assure that I respected the statute for good and scientifical practice in the most current version of the Institute of Technology Karlsruhe, Germany.

Karlsruhe, 9. March 2021

A handwritten signature in black ink, appearing to read "T. Kossel".

This work is created with L<sup>A</sup>T<sub>E</sub>X and apart of the photographs all graphics in this work are created with Geogebra (<https://www.geogebra.org/calculator>) or Python package *pygame*.