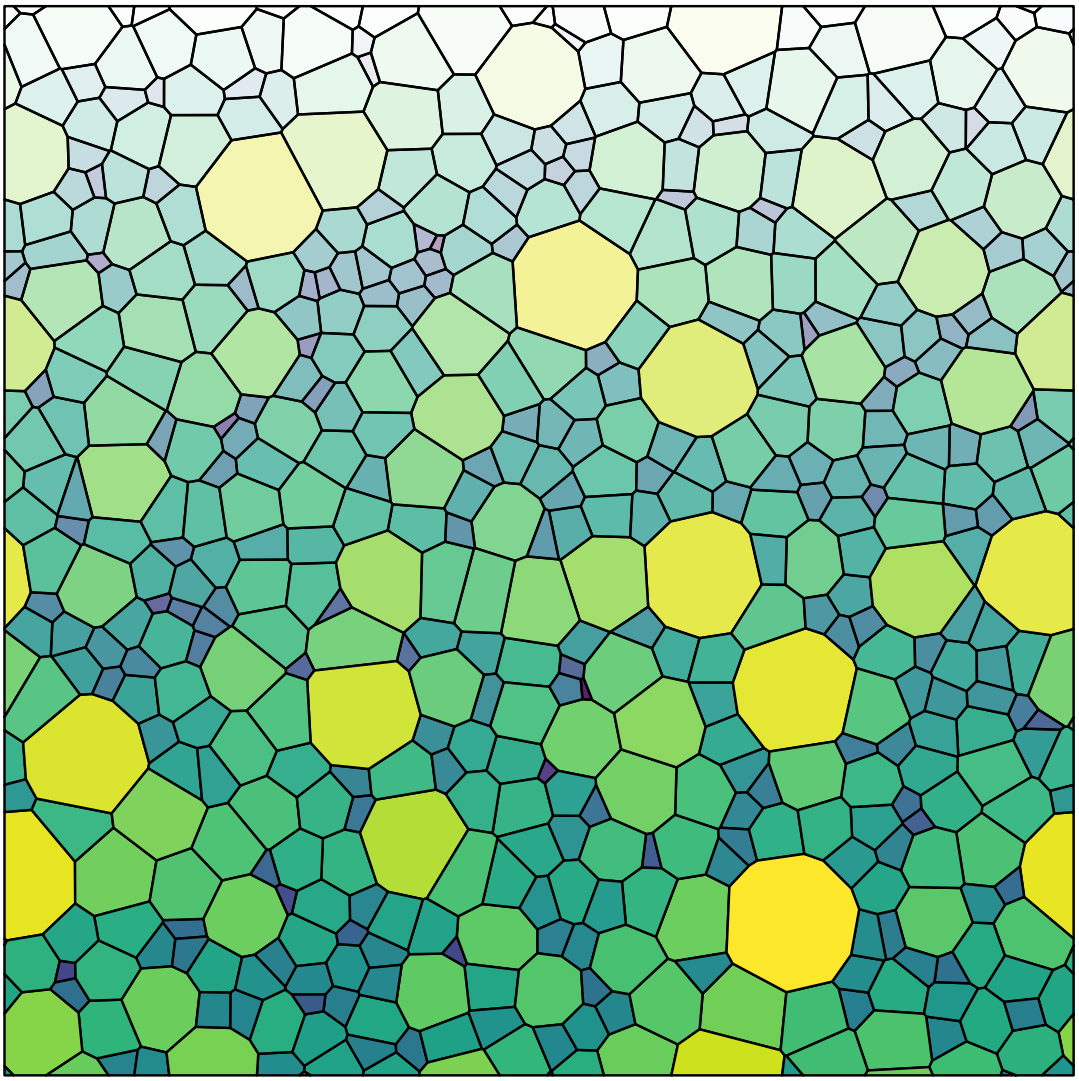
Estimating the cell volume distribution of a 3D Laguerre-Voronoi diagram from a planar section

With applications to materials science

Thomas van der Jagt

Technische Universiteit Delft



TU Delft

Estimating the cell volume distribution of a 3D Laguerre-Voronoi diagram from a planar section

With applications to materials science

by

Thomas van der Jagt

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday October 29, 2021 at 16:00.

Student number:	4305809	
Project duration:	January 1, 2021 – October 29, 2021	
Thesis committee:	Dr. M. Vittorietti,	Università degli studi di Palermo, supervisor
	Prof. dr. G. Jongbloed,	TU Delft
	Dr. A. Cipriani	TU Delft

TU Delft

Abstract

For modelling microstructures of materials the Voronoi diagram is one of the most commonly used models. In this thesis we study a generalization of Voronoi diagrams known as the Laguerre-Voronoi diagram. In particular, we consider the stereological problem of estimating the 3D cell volume distribution of such a diagram from one of its 2D planar sections. This problem is not in general solvable for all Laguerre-Voronoi diagrams. We consider a specific class of Laguerre-Voronoi diagrams generated using an algorithm which allows to control the volume distribution of its cells. Inspired by related stereological problems, an estimator for the distribution of areas observed in a planar section is proposed. This estimator can be used for estimating the cell volume distribution from a sample of observed areas in a planar section of a 3D Laguerre-Voronoi diagram. Given that the problem is motivated by a materials science application we consider Laguerre-Voronoi diagrams with a lognormal cell volume distribution, which is commonly used in this field. The simulations show that the proposed method works well in the sense that on average the estimated parameters of the lognormal cell volume distribution are close to the actual parameter values. While the focus is on the lognormal distribution, generalization of the estimator considering other parametric distributions is briefly discussed. The thesis is concluded with a case study: the 3D grain volume distribution of a real steel microstructure is estimated from observed areas obtained from 2D image data.

Keywords: stereology, Laguerre-Voronoi, tessellation, inverse problem, microstructure.

Acknowledgments

First and foremost, I would like to thank my supervisor Martina Vittorietti for the guidance during the past nine months. Despite having to discuss the progress of the thesis via online meetings I always felt motivated to keep working and exploring new lines of research after each meeting. Your comments, suggestions and the relevant literature you shared with me have greatly benefited the quality of this thesis. I would also like to thank Geurt Jongbloed for the helpful discussions regarding the progress of the project as well as for highlighting important details that should not be overlooked regarding the problem we study in this thesis.

Thanks to Carola Celada-Casero for organizing meetings with researchers in the field of materials science. Everyone who participated in these meetings helped me better understand the questions that are relevant to materials science and provided useful feedback on preliminary results. Finally, I would like to thank my family and friends for supporting me throughout this period. This thesis concludes my time as a Master's student at the TU Delft and I hope the reader finds this work interesting and learns something new about the fascinating intricacies of Laguerre-Voronoi diagrams.

Contents

List of Figures

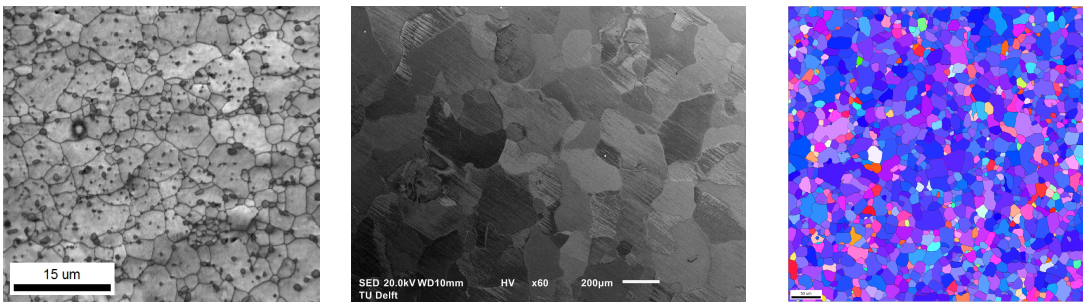
List of Tables

1

Introduction

Suppose you are provided with all the machinery you could dream of to produce metals in all sorts of shapes and sizes. If you would like to sell the metals you produce, it is important that you are able to provide some guarantees of the mechanical properties of these metals. After all, say that some car manufacturer uses your metals, then they would like to know whether these metals are up to the task. Especially during times that the structure of a car would have to endure extreme stresses or even a collision, it becomes important to know what forces these materials can handle. Not for all components the strength of the material is the most important property; other components could have as a requirement other properties such as flexibility. We can see that as a manufacturer of metals, it is desirable to be able to produce them with predetermined mechanical properties.

When observing a metal under a microscope or using other observational techniques, we may see something similar such as the images in Figure 1.1. Important elements of the so-called microstructure of the metal can be distinguished. One of the most important elements in metal microstructures are the grains, also called cells. In Figure 1.1 (a)-(b) it is possible to see that the grains can be described as polygonal shapes delimited by some lines called grain boundaries. In Figure 1.1 (c) grains are instead distinguished by colors that represent their orientation, which is another important feature of a grain.



(a) (b) (c)

Figure 1.1: Images of metal microstructures, taken using various imaging techniques. Pictures taken by (a) Javier Hidalgo, (b) Wei Li, (c) Carola Celada-Casero.

From the microstructure of a metal, we wish to get insights about the mechanical properties of the materials. Among all the possible grain features, in this thesis the focus is on the grain size. The size of the grains influences mechanical properties of the material, such as its behaviour under external stress [1, 2]. Now we arrive at the problem at hand, while we are interested in the 3D volume distribution of the grains in the material, in practice we only observe 2D information in a cross section of the material (such as in Figure 1.1). In materials science it is common to only consider the average 2D grain size. Besides considering 2D grain size instead of 3D grain size it is common that the distributional property of grain sizes is not taken into consideration. As shown in [3] multiple features of the grain size distribution, besides average grain volume, affect the hardness of a particular kind of steel known as Interstitial Free (IF). Therefore, we are interested in deriving or estimating the 3D volume distribution of the grains from the 2D sectional areas of the grains we observe.

The problem of estimating the 3D volume distribution from a 2D material's slice is definitely not trivial. For getting insights we use a mathematical model to represent a metal microstructure, the model which we consider is called Laguerre-Voronoi diagram. The Laguerre-Voronoi diagram is a generalization of the Voronoi diagram. The review paper [4] provides an overview of methods for modelling microstructures in materials and Voronoi diagrams are described as the state of the art geometrical method for generating so-called polycrystalline microstructures. In the next chapter the definition of a Laguerre-Voronoi diagram and its relation with the Voronoi diagram are given. As in the case of the microstructure of a metal, Voronoi diagrams and Laguerre-Voronoi diagrams consist of cells. The problem we would like to solve is estimating the volume distribution of the cells in a Laguerre-Voronoi diagram from the areas we observe in a cross section. There are several reasons why the resolution of this problem is of interest:

- It allows for a better understanding of the effect of the grain volume distribution on the mechanical behaviour. For example, a tensile test can be performed on a material to record its ability to resist to stress. In order to relate the results of this test to the grain volume distribution we first need to know this distribution. If a relationship is found between the grain volume distribution and some mechanical property then this information can be used in the production of metals. Since the size of the grains can be modified with mechanical processes, loosely speaking, it is possible to tune their size to obtain a material with the desired mechanical properties.

- An accurate representation of materials microstructures is useful for mechanical behaviour simulations. Laguerre-Voronoi diagrams and other mathematical models serve as so-called Representative Volume Elements (RVEs), used as virtual samples of virtual mechanical experiments. Such RVEs acts as "digital twins" of the real material in such a virtual experiment. Therefore, it is useful if we can construct a Laguerre-Voronoi diagram which accurately describes the microstructure of a material using just the 2D data obtained from a cross section of this material.

- The resolution of this problem may yield a general method which can be used for other models of materials microstructures. The use of Laguerre-Voronoi diagrams might not be appropriate for representing some metal microstructures, instead another model is used. In those cases we still want to be able to determine the grain volume distribution from 2D data.

In order to properly approach the problem we first need to define Laguerre-Voronoi diagrams and discuss their properties. Once we know how to use these diagrams we provide a more mathematically precise statement of the problem. We perform various simulations in the search

of a relationship between cell volumes and cross section areas. Finally, we propose a solution to the problem and test it on 2D data of a cross section of a real material to verify that it works well enough for practical use.

In the remaining sections of this chapter we provide a motivation for the use of the Laguerre-Voronoi diagram as a model for materials microstructures. We conclude this chapter with an overview of the contents in this thesis.

1.1 Motivating the use of Laguerre-Voronoi diagrams

In this section we briefly mention some relevant literature to further motivate the choice of the Laguerre-Voronoi diagram. For a general overview of the statistical analysis of microstructures in materials we refer to [5]. While it does not discuss Laguerre-Voronoi diagrams specifically, it does cover the Voronoi diagram in the context of microstructure modelling. Previously, we mentioned that Voronoi diagrams may be considered the state of the art geometrical method for generating so-called polycrystalline microstructures. This may raise the question why we consider Laguerre-Voronoi diagrams instead of Voronoi diagrams. Compared to Voronoi diagrams, Laguerre-Voronoi diagrams are more flexible models, as they give control over the volume distribution of the cells.

In [6] a 3D Laguerre-Voronoi diagram is fitted to tomographic data of a foam. What is especially interesting is that a comparison is made with the Voronoi diagram. In this particular study it was found that the geometrical properties of cells of the Laguerre-Voronoi diagram corresponded better with the tomographic data in comparison to the cells of the Voronoi diagram. Another paper which similarly concluded that the Laguerre-Voronoi diagrams they considered provided a more accurate description of geometrical features in real materials compared to Voronoi diagrams is [7]. This suggests that the additional freedom which the Laguerre-Voronoi diagram provides over the Voronoi diagram may also result in a model that is a more accurate representation of real materials.

As mentioned in the introduction, the Laguerre-Voronoi diagram is of interest from the point of view of materials science as it can be used in simulations for the estimation of mechanical properties. When using a Laguerre-Voronoi diagram to represent the microstructure of some material it is useful to know that finite element methods may be applied on these diagrams to estimate mechanical properties of the material. Finite element methods may for example be used to perform a plasticity test on a virtual material by solving a partial differential equation which describes the forces exerted on this virtual material. In [4] the authors describe how finite element methods may be used for various models of materials (RVEs). In [8] it is shown how a Laguerre-Voronoi diagram may yield a mesh which is appropriate for finite element methods. In the same paper an algorithm is discussed for fitting a Laguerre-Voronoi diagram to a material when 3D information of the material is available, e.g. grain volumes and their centers of mass.

1.2 Overview of the thesis

In this section a brief overview of the thesis is given. We begin with introducing the Laguerre-Voronoi diagram and their properties in chapter 2. In chapter 3 we provide a more precise description of the problem statement and we highlight difficulties associated with the problem. Of particular importance is that we show that we need to restrict the problem to a particular class

of Laguerre-Voronoi diagrams. In chapter 4 we investigate the use of an algorithm for generating Laguerre-Voronoi diagrams. The chapter is concluded with a precise description of how 3D Laguerre-Voronoi diagrams and planar sections are computed in all further simulations. In chapter 5 one particular aspect of a Laguerre-Voronoi diagram is investigated, namely its weights. In particular the relationship between weights and a close packing of spheres is considered. In chapter 6 we discuss properties of planar sections of 3D Laguerre diagrams, and we investigate whether these properties are useful for the problem at hand. In chapter 7 an estimator is proposed for the distribution of areas in cross sections of the Laguerre-Voronoi diagrams we consider, which also depends on the cell volume distribution. In the following chapter, chapter 8, the use of the estimator is further motivated by comparing it to analytical solutions of related problems. Motivated by their use in materials science we consider 3D Laguerre-Voronoi diagrams with a lognormal cell volume distribution in chapter 9. We use the estimator proposed in chapter 7 for estimating a lognormal cell volume distribution from areas observed in a planar section. Finally, in chapter 10 the proposed estimator is applied to real-world data. The grain volume distribution is estimated from 2D image data of a sample of steel. In the appendices additional derivations may be found as well as some details of code implementations.

2

Preliminaries

In this chapter we provide some necessary definitions and notation which is used throughout this thesis. We begin with an introduction to the Laguerre-Voronoi diagram and its properties. Since we are interested in estimating the volume distribution of a Laguerre-Voronoi diagram from a planar section we also discuss the main features of planar sections. This chapter is concluded with an overview of well-known parametric probability distributions used for representing the grain volume distribution in metal microstructures.

2.1 Laguerre-Voronoi diagrams

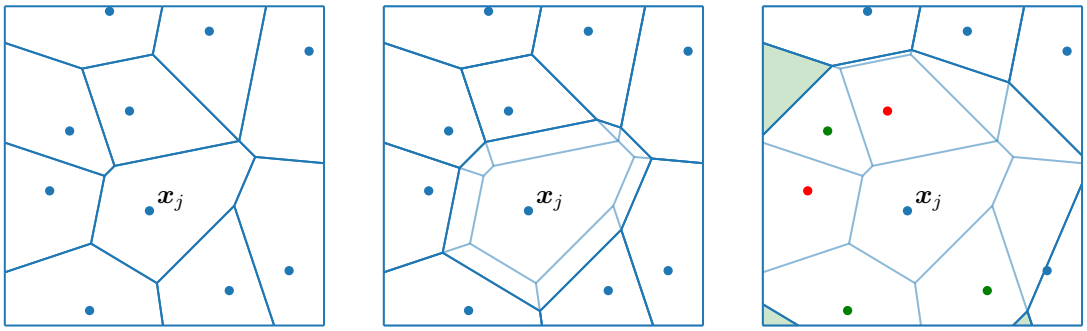
In this section we introduce the Laguerre-Voronoi diagram which can be seen as a generalization of the well-known Voronoi diagram. If the reader is not familiar with the Voronoi diagram this will not pose an issue, as we will see how these diagrams are related. The Laguerre-Voronoi diagram is known by many names, including but not limited to: Laguerre tessellation, Laguerre diagram, radical Voronoi diagram and power diagram. We usually refer to it simply as a Laguerre diagram. Similar to the Voronoi diagram, a Laguerre diagram provides a method to partition a domain into convex cells. Let us now simply start by providing the definition of a Laguerre diagram:

Definition 1 (Laguerre-Voronoi diagram)**.** [9, p. 128]. Given a convex domain $\Omega \subset \mathbb{R}^d$, n distinct generator points: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \ldots, w_n \in \mathbb{R}$. The Laguerre-Voronoi diagram $\{L_i\}_{i=1}^n$ generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$ is defined by:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x} - \boldsymbol{x}_i||^2 - w_i \leq ||\boldsymbol{x} - \boldsymbol{x}_j||^2 - w_j \ \ \forall j \in \{1, \ldots, n\}\}.$$

Given that we choose a domain Ω, the Laguerre diagram is fully defined by its generator points and their weights. The classical Voronoi diagram can be retrieved by simply setting all weights to be equal: $w_1 = w_2 = \cdots = w_n$. Compared to Voronoi diagrams, Laguerre diagrams provide additional flexibility by associating a weight with each generator point. These weights provide control over the size of the cells. While the definition is valid for arbitrary dimensions d we will focus only on two-dimensional and three-dimensional Laguerre diagrams ($d = 2$ or $d = 3$).

In order to better understand the construction and properties of a Laguerre diagram a simple example is shown. Ten generator points are placed in a unit square. Equal weight is assigned to



(a) (b) (c)

Figure 2.1: (a): Classical Voronoi diagram, all weights are equal. (b): A Laguerre diagram obtained by increasing the weight w_j corresponding to the generator point $\boldsymbol{x}_j$. (c) A Laguerre diagram obtained by further increasing the weight w_j. Red generator points belong to empty cells. Green generator points are not contained within their cell, they correspond to cells with green areas.

each generator point, leading to the classical Voronoi diagram shown in Figure 2.1 (a). Then, we increase the weight w_j which corresponds to the generator point labelled with $\boldsymbol{x}_j$. This causes the associated cell L_j to become larger as can be seen in Figure 2.1 (b). Finally, the weight w_j is further increased such that L_j becomes larger which causes some cells to vanish, which is shown in Figure 2.1 (c). Besides the vanishing cells we also observe that some generator points are not located within their cell.

In the previous example some properties of Laguerre diagrams can be observed. We now provide a more elaborate list of properties of Laguerre diagrams, see reference [9, p. 129-131] and [10, p. 81, 82]:

L.1 Cells of a Laguerre diagram may be empty.

L.2 Given n distinct generator points: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \ldots, w_n \in \mathbb{R}$, the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$ is the same as the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1 + c), \ldots, (\boldsymbol{x}_n, w_n + c)$ for any $c \in \mathbb{R}$.

L.3 The generator point of a cell is not necessarily contained within its cell.

L.4 The cells are space-filling: $\cup_{i=1}^{n} L_i = \Omega$, and cells only overlap at their boundaries with neighboring cells. We may say that a Laguerre diagram tessellates Ω.

L.5 The cells are convex polyhedra if $d = 3$ and convex polygons if $d = 2$.

Properties **L.1** and **L.3** do not hold for the Voronoi diagram, which has no empty cells and all generator points are contained within their cells.

Coming back to the materials science application we we may start thinking of how a Laguerre diagram could provide an accurate model for a microstructure. Laguerre diagrams can be used to describe a metal at the microscopic level, where the cells represent the grains in the metal. One thing that may seem problematic is the shape of the boundary cells. To some extent this

can also be seen in Figure 2.1. The cells which intersect with the boundary of the domain have a somewhat different shape compared to cells in the center of the domain, in fact their shape is highly influenced by the shape of the domain. To deal with artificial boundary effects Laguerre diagrams with periodic boundary conditions are used. We define the periodic distance and the Laguerre diagram with periodic boundary conditions for box shaped domains. We provide the definition of 2D Laguerre diagrams with periodic boundary conditions; the definition for the 3D case is analogous.

Definition 2 (Laguerre diagram with periodic boundary conditions)**.** [11]. Given $l_1 > 0$, $l_2 > 0$ we consider the box-shaped domain:

$$\Omega = [0, l_1] \times [0, l_2].$$

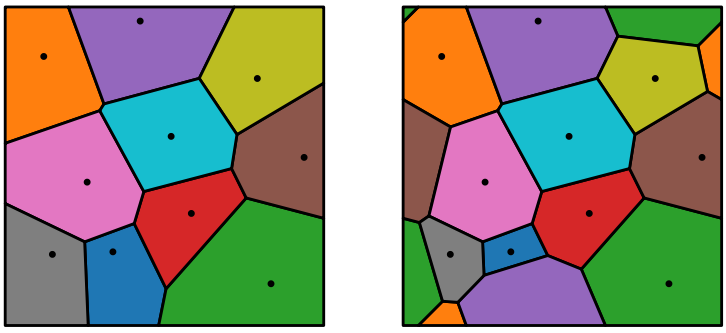
Given $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ the periodic distance between these points is defined as:

$$||\boldsymbol{x} - \boldsymbol{y}||_{\text{per}} = \min\{||\boldsymbol{x} - \boldsymbol{y} + (il_1, jl_2)|| : i, j \in \mathbb{Z}\}.$$

Given n distinct generator points: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \ldots, w_n \in \mathbb{R}$. The Laguerre diagram $\{L_i\}_{i=1}^n$ with periodic boundary conditions, generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$ is defined by:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x} - \boldsymbol{x}_i||_{\text{per}}^2 - w_i \leq ||\boldsymbol{x} - \boldsymbol{x}_j||_{\text{per}}^2 - w_j \ \ \forall j \in \{1, \ldots, n\}\}.$$

In Figure 2.2 we visualize the difference between a 2D Laguerre diagram with periodic boundary conditions and a Laguerre diagram without periodic boundary conditions. It is important to realize that when using periodic boundary conditions we may observe different fragments of a cell in different locations. Take for example the dark green cell in Figure 2.2 (b), it consists of four different fragments. In a (brick-shaped) 3D Laguerre diagram with periodic boundary conditions a cell could consist of at most eight fragments (one fragment in each of the eight corners).



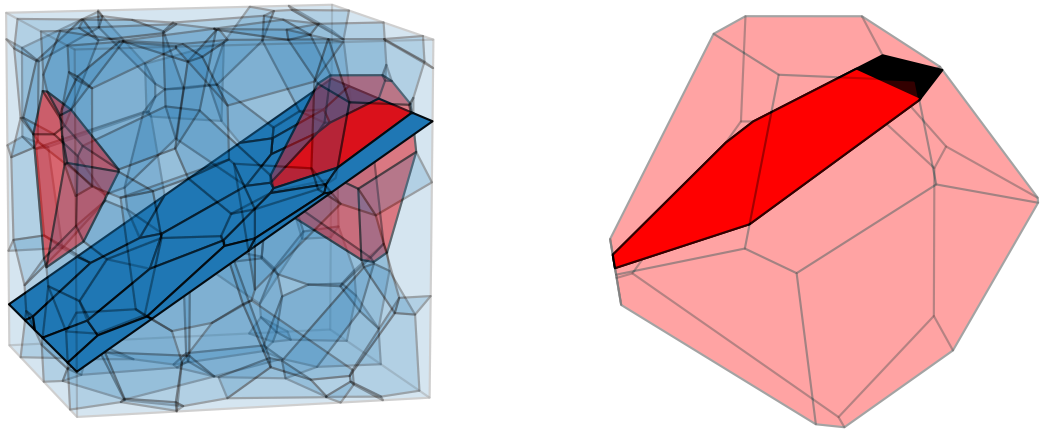
(a) (b)

Figure 2.2: A comparison of a 2D Laguerre diagram with or without periodic boundary conditions. Both diagrams have the same generator points and weights. Colors indicate unique different cells. (a): without periodic boundary conditions. (b): with periodic boundary conditions.

Finally, we need to discuss planar sections of 3D Laguerre diagrams, usually just referred to as cross sections. The cross section of a 3D Laguerre diagram is a 2D Laguerre diagram [10, p. 86]. This is not the case for Voronoi diagrams, in the sense that the cross section of a 3D Voronoi diagram is not a 2D Voronoi diagram but a 2D Laguerre diagram [9, p. 131, 132]. In chapter 6 we investigate whether the properties of these 2D Laguerre diagrams are useful for solving our problem.

As mentioned before we use 3D Laguerre diagrams with periodic boundary conditions. It is important to realize that because of the periodic boundary conditions, a cell may consist of multiple fragments. When studying 2D cross sections, we need to keep in mind that the periodicity refers to the 3D structure. Suppose a cell L_i intersects with the cross section plane P, then we observe the area of $L_i \cap P$. Mathematically, area$(L_i \cap P)$ represents the total area, hence if multiple fragments of L_i appear in P the total section area of this cell is the sum of all the areas of its fragments in the plane P.

The direction of cross section planes also plays a very important role. It seems that considering arbitrary cross section planes could in some sense be problematic. If we allow inclined planes for the cross section we may in some sense observe artificial boundary effects. For clarifying the concept, let's consider an example shown in Figure 2.3. We visualize a 3D Laguerre diagram, with transparent cells such that we can clearly see the cross section plane. One particular cell is highlighted in red, it consists of two fragments and only one of the fragments appears in the cross section. If we combine the two fragments of the cell, we get a clearer visualization of the shape of the cell, Figure 2.3 (b). We see that in some sense a piece is missing in the cross section



(a) (b)

Figure 2.3: Image to highlight that periodic boundary conditions can still lead to artificial boundary effects when taking cross sections with inclined planes. (a): A 3D Laguerre diagram with an inclined cross section plane, a single cell in the plane is highlighted in red. (b): The same cell that is highlighted in red, the two fragments which can be seen in (a) are combined to visualize the actual shape of the cell. The black fragment does not appear in the cross section.

of this cell, which we obtain by extending the cross section plane through this combined cell and which is highlighted in black. We could say that the cross section we observe for this cell is too small given the size of the cell. Such missing pieces will not be observed for cells which are not located at the boundary of the domain.

While periodic boundary conditions resolve artificial boundary effects, its definition suggests that planes of the form $x = c$, $y = c$ or $z = c$ with $c \in \mathbb{R}$ are a better choice. By using these types of cross sections we do not have this missing pieces problem, this is a direct consequence of the definition of the periodic distance. Suppose we take a cross section of the form $z = c$. If we re-create the image in Figure 2.3 (b) for each cell then there would be nothing to highlight in black. As all fragments of cells in the cross section appear at the same z-height, combining the fragments of the cells would show that there are no missing pieces.

2.2 List of commonly used parametric distributions

In this section we introduce some notation and probability distributions which are used throughout this thesis. In particular we mention distributions which are considered for the cell volume distribution in metal microstructures and in Laguerre diagrams. The lognormal distribution and (generalized) gamma distribution are often considered in materials science for describing the grain size distribution. These are also useful for describing the cell volume distribution of the Poisson-Voronoi diagram [12]. We frequently abbreviate probability density function as pdf but it may also simply be referred to as density. The cumulative distribution function is often abbreviated as cdf. Because we consider distributions of volumes and areas we are almost exclusively dealing with distributions which are supported on $[0, \infty)$ or a subset thereof.

The density of the lognormal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ is given by:

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right), \quad x > 0. \tag{2.1}$$

We write $X \sim \text{Lognormal}(\mu, \sigma^2)$ to signify that X is a random variable with a lognormal distribution with parameters μ and σ. Let $X \sim \text{Lognormal}(\mu, \sigma^2)$, then the following holds:

$$\mathbb{E}(X^\alpha) = \exp\left(\alpha\mu + \frac{1}{2}\alpha^2\sigma^2\right), \quad \text{for } \alpha \in \mathbb{R}. \tag{2.2}$$

$$X^\alpha \sim \text{Lognormal}(\alpha\mu, \alpha^2\sigma^2), \quad \text{for } \alpha \neq 0. \tag{2.3}$$

While the generalized gamma distribution will only sporadically be referred to, we do mention its pdf and its moments. The pdf of the generalized gamma distribution with parameters $\alpha > 0$, $\beta > 0$, $c > 0$ is given by:

$$f(x|\alpha, \beta, c) = \frac{cx^{c\alpha-1}}{\beta^{c\alpha}\Gamma(\alpha)} \exp\left(-\left(\frac{x}{\beta}\right)^c\right), \quad x > 0.$$

Where Γ denotes the gamma function. We write $X \sim \text{GG}(\alpha, \beta, c)$ to signify that X is a random variable with a generalized gamma distribution with parameters α, β and c. We obtain the gamma distribution by setting $c = 1$. Let $X \sim \text{GG}(\alpha, \beta, c)$, then:

$$\mathbb{E}(X^r) = \beta^r \frac{\Gamma(\alpha + \frac{r}{c})}{\Gamma(\alpha)}, \quad \text{for } r \in \mathbb{R}. \tag{2.4}$$

In a few cases we consider an exponential distribution for the cell volume distribution. While this is not necessarily realistic for real materials, mathematically it is a convenient distribution as it only depends on a single parameter. The pdf of the exponential distribution with rate parameter $\lambda > 0$ is given by:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

We write $X \sim \text{Exp}(\lambda)$ to signify that X is a random variable with an exponential distribution with parameter λ.

Next, we consider the uniform distribution. This distribution is not used for cell volume distributions in this thesis. Instead, we consider using it for the locations of the generator points of a Laguerre diagram. The pdf of the uniform distribution over the interval $[a, b]$ with $a < b$ is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}.$$

We write $X \sim U([a, b])$ to signify that X is a uniformly distributed random variable on the interval $[a, b]$. The pdf of the uniform distribution over the d-dimensional domain $\Omega = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$ with $a_i < b_i$ is given by:

$$f(x_1, x_2, \ldots, x_d) = \begin{cases} \frac{1}{\prod_{i=1}^{d}(b_i - a_i)} & \text{if } (x_1, x_2, \ldots, x_d) \in \Omega \\ 0 & \text{otherwise} \end{cases}.$$

We write $\boldsymbol{X} \sim \boldsymbol{U}(\Omega)$ to signify that $\boldsymbol{X}$ is a uniformly distributed random variable over the domain Ω.

3

Problem statement

In the previous chapter we introduced the Laguerre diagram and described some of its properties. In this chapter we state the main problem faced in this thesis. As mentioned before we would like to estimate the 3D volume distribution of cells of a Laguerre diagram from a sample of observed areas in a 2D cross section. First, the necessary notation is introduced.

While the cell volume distribution was mentioned before, we now also introduce the cross sectional area distribution. Given a cross section of a Laguerre diagram, the distribution of areas of cells which appear in the section is referred to as the cross sectional area distribution. A more precise description of how cross sections are taken is given in the next chapter.

- Let F_V, f_V be the cdf and pdf respectively, of the cell volume distribution of a 3D Laguerre diagram.
- Let F_A, f_A be the cdf and pdf respectively, of the cross sectional area distribution of a 3D Laguerre diagram.

Ideally, we would like to solve the following two problems:

1. **Direct problem**: Suppose that we have a Laguerre diagram with a known cell volume distribution F_V. What is the cross sectional area distribution F_A?
2. **Inverse problem**: Suppose we know that the sectional area distribution F_A describes the distribution of areas in a cross section of a Laguerre diagram. What is the cell volume distribution F_V of this Laguerre diagram?

It is important to see that for the direct problem we need to determine 2D information given 3D information and for the inverse problem we need to determine 3D information given 2D information. The field which deals with the problem of estimating 3D quantities based on 2D information and measurements is known as stereology. In [13] the practical purpose of stereology is described as the extraction of 3D quantitative information from 2D microscope images. Stereology does not concern itself with reconstructing the 3D geometry of the material, but only with the estimation of certain parameters, such as average grain volume in a material. Stereological techniques are for example used in neuroscience, pathology, geology, materials science and clinical medicine.

In the next section we discuss what an inverse problem is. We also consider some challenges that come with inverse problems and in particular in our inverse problem. We argue that we

cannot consider arbitrary Laguerre diagrams and we need to restrict ourselves to a particular class of Laguerre diagrams. In the final section of this chapter we briefly go over some related inverse problems, as well as proposed solutions to these problems.

3.1 Inverse problems

Let's begin with an intuitive explanation of an inverse problem. As in [14], the following could be a description of an inverse problem. Suppose we are interested in some probability distribution P, the problem is that we cannot observe samples from P, we can only observe samples from Q which is related to P via some relation. If we assume the distribution P to be known, and we try to determine the distribution Q then that is the corresponding direct problem. Examples of inverse problems in [14] are a deconvolution problem and the famous Wicksell's corpuscle problem which is also a stereological problem.

Essentially, Wicksell's corpuscle problem considers the case where some body is filled with spheres of various sizes and we would like to determine the size distribution of the spheres from a cross section. Let us simply recall the original statement of Wicksell's corpuscle problem as stated in the original paper [15, p. 86, 87]:

> In an opaque body there are suspended a large number of spherical corpuscles of different sizes, the density of the centres and the distribution of the sizes being the same in all parts of the body. This body is split in two by a plane section. To express the distribution of the diameters of the corpuscles in terms of the distribution of the diameters of the circular contours found in the plane section.

While Wicksell considers the problem in terms of the diameters of spheres, we consider it in terms of the radii of the spheres. Let f be the probability density function associated with the radii of the 3D spheres. Let ϕ be the probability density function of the radii of the apparent 2D circles (the circles we observe in the plane section). Wicksell gives the following relationship:

$$\phi(x) = \frac{x}{\mathbb{E}(R)} \int_x^\infty \frac{f(r)}{\sqrt{r^2 - x^2}} \mathrm{d}r. \tag{3.1}$$

Here, $\mathbb{E}(R)$ is the average radius of the 3D spheres:

$$\mathbb{E}(R) = \int_0^\infty r f(r) \mathrm{d}r.$$

In terms of the notation in the introduction of this section this means that we can identify ϕ with the distribution Q and we can identify f with the distribution P. Wicksell [15] shows that (3.1) can be solved for $f(r)$:

$$f(r) = -\frac{2\mathbb{E}(R)r}{\pi} \int_r^\infty \left(\frac{\mathrm{d}}{\mathrm{d}x} \frac{\phi(x)}{x} \right) \frac{1}{\sqrt{x^2 - r^2}} \mathrm{d}x.$$

This shows that ϕ has to have some level of smoothness for this solution to exist. This suggests that we cannot just take an arbitrary probability density function ϕ and expect that there exists a pdf f which corresponds to ϕ as in (3.1). While the inverse problem we consider is not quite the same as the Wicksell problem, some comparisons can be made. In the Wicksell problem all particles are shaped as spheres and in our problem the particles are shaped as convex polyhedra. As in the case of the Wicksell problem we should also expect that f_A cannot be an arbitrary pdf,

via a similar reasoning as to why this is the case for ϕ. The statement of the Wicksell problem also highlights an important condition. Given a Laguerre diagram with a known cell volume distribution, as in the Wicksell case we would like the distribution of the volumes of the cells throughout the diagram to be roughly the same in all parts of the diagram. A Laguerre diagram does not satisfy this property if cells of similar volumes are grouped together. A consequence of this would be that a cross section of such a Laguerre diagram could reveal very different information depending on the location of the cross section. If the volumes of the cells are randomly distributed throughout the diagram then the location of the cross section should not matter.

We want to stress that the provided description of the direct and inverse problem is insufficient in its current form. First of all, we indicated earlier that we need to be precise in describing how cross sections are taken. In the next chapter another issue is highlighted, we observe via examples that it is possible for a Laguerre diagram to have cells of similar size clustered together. In the same chapter we investigate the use of an algorithm for generating Laguerre diagrams and how it should be used to prevent such clusters. We conclude the chapter by precisely describing how we compute Laguerre diagrams and cross sections for all further analyses. Hence, we restrict ourselves to a particular class of Laguerre diagrams.

For inverse problems an important issue is identifiability, does the inverse problem have a unique solution? For our problem specifically we would like to know whether the cell volume distribution F_V can be uniquely determined from the cross sectional area distribution F_A. In other words, we would like to know whether two different volume distributions could give rise to the same cross sectional area distribution. While we cannot give a statement regarding the identifiability when considering general volume distributions F_V we will see in chapter 9 that when considering a lognormal volume distribution we have an identifiable problem, for the class of Laguerre diagrams we consider.

3.2 Related problems and algorithms

While we have decided to look at Laguerre diagrams, and the relationship between the cell volume distribution and the cross sectional area distribution a similar problem has been solved for the Poisson-Voronoi diagram. The Poisson-Voronoi diagram is a Voronoi diagram for which the generator points are a realization of a homogeneous Poisson point process with intensity parameter λ. A scaling argument is derived in [12], using this argument and via simulations it is shown that the volume distribution of cells in a Poisson-Voronoi diagram can be estimated accurately for λ. This can be used to solve the inverse problem of estimating the volume distribution of cells in a Poisson-Voronoi diagram from a cross section. See for example [16] for an estimator of λ which only depends on the average observed area of cells in the cross section. As a result, it is possible to estimate the volume distribution of a Poisson-Voronoi diagram by estimating λ using the observed areas in a cross section. The Poisson-Voronoi case requires the estimation of only a single parameter, we clearly have much more parameters to deal with.

In [17] a similar problem is considered. The Laguerre diagrams are generated via a random close packing of spheres. A random close packing of spheres is obtained via an algorithm which efficiently positions non-overlapping spheres in some domain, filling as much space with spheres as possible. The algorithm used in their simulations iteratively pours hard spheres into a box, gravity is simulated such that the spheres fall down and the algorithm continues until the box is

considered full. The volumes of the spheres follow a chosen lognormal distribution. A Laguerre diagram is then constructed as follows. The generator points of the cells were taken to be the centers of the spheres and the weights were chosen as $w_i = r_i^2$ where r_i is the radius of the sphere with center $\boldsymbol{x}_i$. This particular choice means that each sphere is contained within the associated cell. The authors then use the fact that the volumes of spheres follow a lognormal distribution if and only if their radii follow a lognormal distribution. The proposed method is empirical and it consists of multiple steps. Regression methods are used to link a fitted distribution of cross sectional areas to parameters describing the lognormal distribution of the radii of the hard spheres. While this approach seems to work quite well in the performed simulations, it is not necessarily generalizable to other cell volume distributions and it does not provide a clear relationship between the cell volumes and the cross sectional areas. This approach does raise the question whether the Laguerre diagrams we consider can also be related to a packing of spheres, we investigate this in chapter 5.

Another method which is used to determine a particle volume distribution from cross sectional areas is the Saltikov method [18, 19]. This method assumes particles to be spherical. It estimates the distribution of the 3D spheres radii as a discrete distribution, which is sometimes referred to as a finite histogram. Applying this method to Laguerre diagrams seem inappropriate given that cells of a Laguerre diagram are polyhedra. In [20] an extension to the Saltikov method is proposed which fits a lognormal distribution to the finite histogram.

Inspired by some of the solutions proposed in the above mentioned problems, in this thesis an alternative approach is proposed for the case of Laguerre diagrams. An alternative approach appears necessary because of properties of the Laguerre diagram.

4

Computing Laguerre diagrams with a prescribed volume distribution

Now that the Laguerre diagram has been introduced properly and we discussed the difficulties associated with the direct and inverse problem, we need to make some choices. We have mentioned that the weights of a Laguerre diagram provide more flexibility over Voronoi diagrams, but this flexibility comes with a cost. As we are interested in estimating the volume distribution of the cells in a Laguerre diagram from areas of cells which appear in cross sections, finding results for arbitrary Laguerre diagrams is too ambitious. We need to decide how exactly we choose the generator points and weights and we need to investigate whether the distribution of the volumes of the cells throughout the diagram is roughly the same in all parts of the Laguerre diagram. At the very least, we would like a Laguerre diagram such that the location of the cross section plane does not matter too much. If different cross sections of the same Laguerre diagram would reveal completely different information it may not be possible to estimate the cell volume distribution. In this chapter we attempt to find a method for generating a Laguerre diagram which satisfies these conditions.

Recalling the problem statement we are interested in the relationship between two distributions, the cell volume distribution and the cross sectional area distribution. The algorithm which we discuss in this chapter is therefore particularly appealing since it effectively allows to control one side of the equation. In fact, this algorithm proposed in [11] generates a Laguerre diagram with a cell volume distribution of choice. Both in 3D and in 2D, the algorithm allows to generate Laguerre diagrams with a predefined volume and area distribution, respectively. This algorithm is clearly a precious tool in simulations. When attempting to estimate the 3D cell volume distribution from a cross section we can verify our estimate as the true volume distribution is known. There are various parameters in this algorithm that have influence on the resulting Laguerre diagram. Besides providing control over the volume distribution of cells, the algorithm is used to produce Laguerre diagrams which are regularized in the sense that they are centroidal. The centroid of a cell in a Laguerre diagram is also known as its center of mass and is defined for a non-empty cell L_i as:

$$\boldsymbol{c}_i := \frac{1}{|L_i|} \int_{L_i} \boldsymbol{x} \mathrm{d}\boldsymbol{x}. \tag{4.1}$$

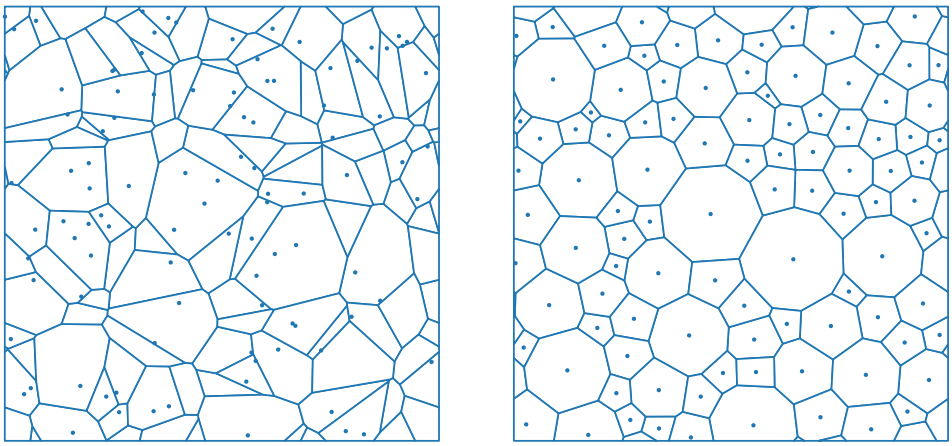
Where $|L_i|$ denotes the volume of L_i:

$$|L_i| = \int_{L_i} \mathrm{d}\boldsymbol{x}. \tag{4.2}$$

Given a Laguerre diagram $\{L_i\}_{i=1}^n$ generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$, which is assumed to only have non-empty cells. This Laguerre diagram is called centroidal when each generator point is equal to its centroid:

$$\boldsymbol{x}_i = \frac{1}{|L_i|} \int_{L_i} \boldsymbol{x} \mathrm{d}\boldsymbol{x} \quad \text{for all } i \in \{1, \ldots, n\}.$$

When we generate a centroidal Laguerre diagram for a chosen volume distribution with the aforementioned algorithm we will end up with a Laguerre diagram which does not have empty cells and (when numerical tolerances are chosen appropriately) will have the property that each cell contains its generator. The fact that each cell contains its generator is not particularly important for our purposes, we just mention it to give some sense of the locations of the generators. It is clear that the algorithm is useful because it provides control over the volume distribution, it is not obvious why we would want this regularization in the form of centroidality. The reason we consider centroidal Laguerre diagrams is that upon visual inspection of Laguerre diagrams we consider centroidal Laguerre diagrams to provide a better representation of the grain structure found in materials. The difference between a Laguerre diagram with and without regularization can be seen in Figure 4.1.



(a) (b)

Figure 4.1: Comparison of a 2D Laguerre diagram without regularization (a) to a regularized (centroidal) Laguerre diagram (b). These diagrams have periodic boundary conditions and both have the same area distribution.

Throughout this chapter we discuss several aspects regarding this algorithm for centroidal Laguerre diagrams. We conclude this chapter by precisely describing how we generate Laguerre diagrams in further analyses. For Laguerre diagrams generated by this particular procedure we want to estimate their volume distribution from a cross section.

4.1 Bourne et al.'s algorithm for a centroidal Laguerre diagram with a prescribed volume distribution

Before stating the algorithm some important results need to be discussed. Suppose that we take any set of distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in a domain Ω. If we would like to have a Laguerre diagram with volumes $m_1, \ldots, m_n$ it turns out that we can obtain the weights $w_1, \ldots, w_n$ such that these volumes are achieved by solving a smooth and unconstrained convex optimization problem. This is the case because the solution to this convex optimization are weights $w_1, \ldots, w_n$ such that for the resulting Laguerre diagram $\{L_i\}_{i=1}^n$ we have $|L_i| = m_i$ for all $i \in \{1, \ldots, n\}$. This is made precise in Theorem 3. Recall that $|L_i|$ denotes the volume of L_i, as in (4.2), and similarly $|\Omega|$ is the volume of Ω. Theorem 3 is stated for general convex domains Ω but in our simulations we only consider brick-shaped domains:

$$\Omega = [0, l_1] \times [0, l_2] \times [0, l_3]. \tag{4.3}$$

Theorem 3. *[21, 22, 11]. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ be distinct points. Let $m_1, \ldots, m_n$ be positive real numbers such that $|\Omega| = \sum_{i=1}^n m_i$. Define the function $g : \mathbb{R}^n \to \mathbb{R}$ by:*

$$g(\boldsymbol{w}) = \sum_{i=1}^n (m_i - |L_i(\boldsymbol{w})|) w_i + \sum_{i=1}^n \int_{L_i(\boldsymbol{w})} ||\boldsymbol{x} - \boldsymbol{x}_i||^2 \mathrm{d}\boldsymbol{x}, \tag{4.4}$$

where $\boldsymbol{w} = (w_1, \ldots, w_n)$. Let $\{L_i\}_{i=1}^n$ be the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$. This function g has the following properties:

1. *g is concave.*
2. *The gradient of g is given by:*

$$\frac{\partial g(\boldsymbol{w})}{\partial w_i} = m_i - |L_i(\boldsymbol{w})|.$$

3. *If $\boldsymbol{w}$ is a critical point of g, that is: $\nabla g(\boldsymbol{w}) = \boldsymbol{0}$, then the Laguerre diagram $\{L_i\}_{i=1}^n$ generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$ has cells with volumes $m_1, \ldots, m_n$.*

$$|L_i(\boldsymbol{w})| = m_i \quad \text{for all } i \in \{1, \ldots, n\}.$$

4. *g is twice differentiable.*

Remark. *Note that in Theorem 3 we write $L_i(\boldsymbol{w})$ instead of L_i to highlight that L_i depends on $\boldsymbol{w}$ and that the properties of g are not trivial as it may seem at first glance. Hereinafter we write $L_i(\boldsymbol{w})$ whenever we would like to highlight the dependence on the weights and otherwise we simply write L_i.*

Remark. *Theorem 3 holds for Laguerre diagrams with or without periodic boundary conditions. Suppose we have a domain of the form (4.3). In the periodic case we consider $\{\tilde{L}_i\}_{i=1}^n$ to be cells of a periodic Laguerre diagram (recall Definition 2). The function g is then defined as:*

$$g(\boldsymbol{w}) = \sum_{i=1}^n (m_i - |\tilde{L}_i(\boldsymbol{w})|) w_i + \sum_{i=1}^n \int_{\tilde{L}_i(\boldsymbol{w})} ||\boldsymbol{x} - \boldsymbol{x}_i||_{per}^2 \mathrm{d}\boldsymbol{x}.$$

We can turn the process of maximizing the concave function g into a convex minimization problem by defining $f := -g$ and then we may simply employ existing algorithms on the convex function f. The algorithm by Bourne et al. [11] iteratively maximizes g which is followed by updating the generator points for the next iteration. Specifically, by maximizing g we obtain a Laguerre diagram, then the centroids of its cells are computed and these are used as the generator points in the next iteration. The algorithm continues until we end up with a Laguerre diagram which has generator points which we consider to be close enough to the centroids of the Laguerre diagram while also having the desired cell volume distribution. In [11] the algorithm is also proven to converge via a monotonicity argument. The algorithm is described in Algorithm 1.

Algorithm 1 Lloyd-type centering algorithm by Bourne et al. [11]

Input: Initial (distinct) generator points $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)} \in \Omega$, target volumes $m_1, \ldots, m_n > 0$ such that $|\Omega| = \sum_{i=1}^n m_i$.

Output: A Laguerre diagram with cells with volumes $m_1, \ldots, m_n$ up to relative error ε_f, the diagram is centroidal up to tolerance ε.

1: $\boldsymbol{w}^{(0)} \leftarrow \mathbf{0}$
2: $k \leftarrow 1$
3: residual $\leftarrow \infty$
4: **while** $k <$ maxit **and** residual $> \varepsilon$ **do**
5: $\quad \boldsymbol{w}^{(k)} \leftarrow$ Use a convex optimization algorithm (such as Algorithm 2 or 3) with initial weights $\boldsymbol{w}^{(k-1)}$ and generator points $(\boldsymbol{x}_1^{(k)}, \boldsymbol{x}_2^{(k)}, \ldots, \boldsymbol{x}_n^{(k)})$ to find $\boldsymbol{w}^{(k)}$ such that $m_i = |L_i^{(k)}|$.
6: $\quad$ Compute centroids $(\boldsymbol{c}_1^{(k)}, \boldsymbol{c}_2^{(k)}, \ldots, \boldsymbol{c}_n^{(k)})$ of current Laguerre diagram $\{L_i^{(k)}\}_{i=1}^n$ via (4.1).
7: $\quad (\boldsymbol{x}_1^{(k+1)}, \boldsymbol{x}_2^{(k+1)}, \ldots, \boldsymbol{x}_n^{(k+1)}) \leftarrow (\boldsymbol{c}_1^{(k)}, \boldsymbol{c}_2^{(k)}, \ldots, \boldsymbol{c}_n^{(k)})$
8: $\quad$ residual $\leftarrow \sqrt{\frac{\sum_{i=1}^n m_i \left|\left|\boldsymbol{c}_i^{(k)} - \boldsymbol{x}_i^{(k)}\right|\right|^2}{|\Omega| \max_j l_j}}$
9: $\quad k \leftarrow k + 1$
10: **end while**
11: **return** $(\boldsymbol{x}_1^{(k-1)}, \boldsymbol{x}_2^{(k-1)}, \ldots, \boldsymbol{x}_n^{(k-1)})$, $\boldsymbol{w}^{(k-1)}$

We now clarify some of the notation and parameters in this algorithm. In Algorithm 1 the Laguerre diagram $\{L_i^{(k)}\}_{i=1}^n$ is the diagram generated by $(\boldsymbol{x}_1^{(k)}, w_1^{(k)}), (\boldsymbol{x}_2^{(k)}, w_2^{(k)}), \ldots, (\boldsymbol{x}_n^{(k)}, w_n^{(k)})$. The l_j's in line 8 are as in equation (4.3). The choice of the residual was taken as in [23], with error tolerance $\varepsilon = 0.001$ and with maxit $= 200$. Using a smaller value of ε often meant that Algorithm 1 only performed 1 or 2 iterations. Using only 1 iteration means there is no regularization, leading to Laguerre diagrams such as seen in Figure 4.1 (a). Therefore, we chose a smaller ε to ensure we end up with a regularized Laguerre diagram which we consider to be a better representation of the grain structure found in materials microstructures. Note that ε is related to the distance

Parameter	Value
ε	0.001
ε_f	0.01
maxit	200
l_j	as in (4.3)
$\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)}$	Is discussed in section 4.2

Table 4.1: Parameters as used in Algorithm 1.

between the generator points and centroids while ε_f is related to how close the cell volumes are to the m_i's. Another remark regarding the residual is that in the periodic case we may replace the Euclidean norm in the residual with the periodic norm. The only parameters we do not discuss in this section are the initial generator points $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)}$, this is covered in section 4.2. The parameters used in Algorithm 1 are summarized in Table 4.1.

In line 5 of Algorithm 1 we solve the convex optimization problem of minimizing $f = -g$. We note that this means that in practice we only find weights $\boldsymbol{w}^{(k)}$ such that we obtain cells with the desired volumes up to relative error ε_f:

$$\frac{\left|m_i - \left|L_i^{(k)}\right|\right|}{m_i} < \varepsilon_f \quad \text{for all } i \in \{1, \ldots, n\}. \tag{4.5}$$

This is the result of choosing the tolerance tol_f for the gradient algorithm that is used for the convex minimization. The gradient methods we employ and this tolerance tol_f will be introduced in a moment. Similarly, we can only guarantee that the resulting Laguerre diagram is centroidal up to the chosen tolerance ε.

In [23] an extensive analysis of the performance of various minimization algorithms is given, as well as a method to reduce the amount of Lloyd-type centering steps in Algorithm 1 via Anderson-acceleration. Anderson-acceleration is a general-purpose scheme for accelerating fixed-point methods. In their simulations, this acceleration scheme could in most cases slightly reduce the amount of iterations of Algorithm 1. As its impact was subtle we do not use this technique, as it adds to the complexity of the algorithm. The aforementioned Lloyd-type centering steps refer to setting the generator points for the next iteration to the centroids of the Laguerre diagram that was computed in the current iteration (line 7 of Algorithm 1). The Lloyd algorithm is a well-known algorithm which produces a centroidal Voronoi diagram.

The results in [23] showed that for Laguerre diagrams with a large number of cells (large n) the Barzilai-Borwein method [24] as described in Algorithm 2 tended to be the fastest minimization algorithm. In the case of diagrams with fewer cells the Newton method was often a better choice while the Malitsky-Mishchenko method [25] (Algorithm 3) was usually also a competitive option. In these algorithms $||.||_\infty$ denotes the usual supremum norm in $\mathbb{R}^n$. An important remark is that

Algorithm 2 Barzilai-Borwein method (BB1) [24]

Input: Initial weights: $\boldsymbol{w}_0 \in \mathbb{R}^n$, initial step size: $\alpha_0 > 0$ and a smooth convex function: $f : \mathbb{R}^n \to \mathbb{R}$.
Output: $\boldsymbol{w} \in \mathbb{R}^n$ such that $||\nabla f(\boldsymbol{w})||_\infty < \text{tol}_f$
1: residual $\leftarrow ||\nabla f(\boldsymbol{w}_0)||_\infty$
2: $j \leftarrow 0$
3: **while** $j < \text{maxit}_f$ **and** residual $> \text{tol}_f$ **do**
4: $\quad \boldsymbol{w}_{j+1} \leftarrow \boldsymbol{w}_j - \alpha_j \nabla f(\boldsymbol{w}_j)$
5: $\quad \alpha_{j+1} \leftarrow \frac{||\boldsymbol{w}_{j+1} - \boldsymbol{w}_j||^2}{(\nabla f(\boldsymbol{w}_{j+1}) - \nabla f(\boldsymbol{w}_j))^T (\boldsymbol{w}_{j+1} - \boldsymbol{w}_j)}$
6: $\quad$ residual $\leftarrow ||\nabla f(\boldsymbol{w}_{j+1})||_\infty$
7: $\quad j \leftarrow j + 1$
8: **end while**
9: **return** $\boldsymbol{w}_j$

Algorithm 3 Malitsky-Mishchenko method [25]

Input: Initial weights: $\boldsymbol{w}_0 \in \mathbb{R}^n$, initial step size: $\alpha_0 > 0$ and a smooth convex function: $f : \mathbb{R}^n \to \mathbb{R}$.
Output: $\boldsymbol{w} \in \mathbb{R}^n$ such that $||\nabla f(\boldsymbol{w})||_\infty < \text{tol}_f$

residual $\leftarrow ||\nabla f(\boldsymbol{w}_0)||_\infty$
$\theta_0 \leftarrow +\infty$
$j \leftarrow 0$
while $j < \text{maxit}_f$ **and** residual $> \text{tol}_f$ **do**
  $\boldsymbol{w}_{j+1} \leftarrow \boldsymbol{w}_j - \alpha_j \nabla f(\boldsymbol{w}_j)$
  $\alpha_{j+1} \leftarrow \min\left\{\sqrt{1+\theta_j}\alpha_j, \frac{||\boldsymbol{w}_{j+1}-\boldsymbol{w}_j||}{2||\nabla f(\boldsymbol{w}_{j+1})-\nabla f(\boldsymbol{w}_j)||}\right\}$
  $\theta_{j+1} \leftarrow \frac{\alpha_{j+1}}{\alpha_j}$
  residual $\leftarrow ||\nabla f(\boldsymbol{w}_{j+1})||_\infty$
  $j \leftarrow j + 1$
end while
return $\boldsymbol{w}_j$

the Barzilai-Borwein method as described in Algorithm 2 is not in general globally convergent. Globalization techniques do exist to address this issue but in the simulations performed by [23] the algorithm always converged. In our simulations there were in fact instances where the algorithm did not converge (within the chosen amount of maximum iterations), fortunately this was a rare occurrence. Specifically, it seemed that the higher the variance in the volume distribution, the more likely it was not to converge in a reasonable amount of iterations. Therefore, we also describe the Malitsky-Mishchenko method which is globally convergent for differentiable convex functions. We can benefit from the performance of the Barzilai-Borwein method in most cases, and when convergence is not achieved we use the Malitsky-Mishchenko method as a back-up.

In step 5 of Algorithm 1 we use Algorithm 2 or 3 to minimize the function f_k in iteration k which depends on the generator points $\boldsymbol{x}_1^{(k)}, \ldots, \boldsymbol{x}_n^{(k)}$. Let $\{L_i^{(k)}\}_{i=1}^n$ be the Laguerre diagram generated by $(\boldsymbol{x}_1^{(k)}, w_1), \ldots, (\boldsymbol{x}_n^{(k)}, w_n)$. The function $f_k : \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$f_k(\boldsymbol{w}) := -\sum_{i=1}^n (m_i - |L_i^{(k)}(\boldsymbol{w})|)w_i - \sum_{i=1}^n \int_{L_i^{(k)}(\boldsymbol{w})} ||\boldsymbol{x} - \boldsymbol{x}_i^{(k)}||^2 \mathrm{d}\boldsymbol{x}. \tag{4.6}$$

And therefore:

$$\nabla f_k(\boldsymbol{w}) = (|L_1^{(k)}(\boldsymbol{w})| - m_1, |L_2^{(k)}(\boldsymbol{w})| - m_2, \ldots, |L_n^{(k)}(\boldsymbol{w})| - m_n).$$

By Theorem 3 we have that f_k is convex and twice differentiable. For the algorithms we need to make a choice regarding some parameters. The parameters for both algorithms are shown in Table 4.2.

The choice of tol_f is motivated by the fact that the following inequality implies (4.5):

$$||\nabla f_k(\boldsymbol{w})||_\infty = \max_{i\in\{1,\ldots,n\}} \left| m_i - \left| L_i^{(k)}(\boldsymbol{w}) \right| \right| < \varepsilon_f \min_{i\in\{1,\ldots,n\}} m_i = \text{tol}_f. \tag{4.7}$$

The parameter maxit_f is chosen rather large. Technically we only need this parameter for the rare case that Algorithm 2 does not converge, it ensures that the algorithm terminates. Regarding

Parameter	Value
$\boldsymbol{w}_0$	$\boldsymbol{w}^{(k-1)}$, as stated in Algorithm 1
ε_f	0.01, as in [11]
tol_f	$\varepsilon_f \min_i m_i$, as in [11]
maxit_f	$10n$
α_0	$0.1 \cdot (l_1 l_2 l_3)^{-\frac{1}{3}}$, as in [23]

Table 4.2: Parameters as used in Algorithm 2 and 3.

the implementation of either of the gradient algorithms it is interesting to note that both only require evaluations of the gradient of f_k, we never actually have to compute f_k. More details on the implementation of all algorithms can be found in Appendix A. These algorithms are now implemented and available in the repository https://github.com/thomasvdj/vorostereology.

4.2 Choosing the initial generator points for Algorithm 1

The algorithm as introduced in the previous section still requires us to choose the initial generator points $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)}$. As mentioned at the start of this chapter, we are interested in Laguerre diagrams with the volumes of the cells being randomly distributed throughout the diagram. If we have a Laguerre diagram which has cells of similar volumes grouped together this diagram does not satisfy the requirement. In this section we highlight that we need to make sure that the initial generator points and the cell volumes are chosen independently from one another, and we will also see that even when we do this there are bad choices for the initial generator points. We remind that all Laguerre diagrams in this section have periodic boundary conditions.

In [11] some examples are shown that highlight the impact of some choices of the initial generator points on the resulting Laguerre diagram. They also show how to generate a Laguerre diagram with clusters of cells with a similar size, which is useful for representing microstructures of some materials. An important remark is that these clusters were achieved when choosing the initial generator points dependent on the chosen volumes. We now look at a similar example in 2D, consider the following discrete distribution:

$$\mathbb{P}(V = 1) = 0.8 \quad \text{and} \quad \mathbb{P}(V = 10) = 0.2 \tag{4.8}$$

This distribution is taken as the area distribution, which means we expect to see a large amount of small cells (with area 1) and a small amount of large cells (with area 10). Let us now describe the chosen parameters. Set $n = 200$, and let $m_1, \ldots, m_n$ be a an independent and identically distributed sample from the distribution (4.8). As a domain we take the square: $\Omega = [0, l]^2$ with $l = \sqrt{\sum_{i=1}^{n} m_i}$. Before using Algorithm 1 we need to choose the locations of the initial generator points. Inside this square domain we draw an invisible circle; its area is equal to the number of m_i's which are equal to 1. If i is such that $m_i = 1$ we put its generator inside this circle, if $m_i = 10$ instead, then we put its generator outside this circle. The result of this procedure is shown in Figure 4.2 (a), with the Laguerre diagram produced by Algorithm 1 in Figure 4.2 (b). This highlights that when the initial generator points are chosen dependent on the cell area (or volume in the 3D case) we may create a cluster of cells with the same size.

Now we keep the same generator points as before, but randomly shuffle the assigned areas (the m_i's) which is shown in Figure 4.2 (c). The resulting Laguerre diagram produced by Algorithm

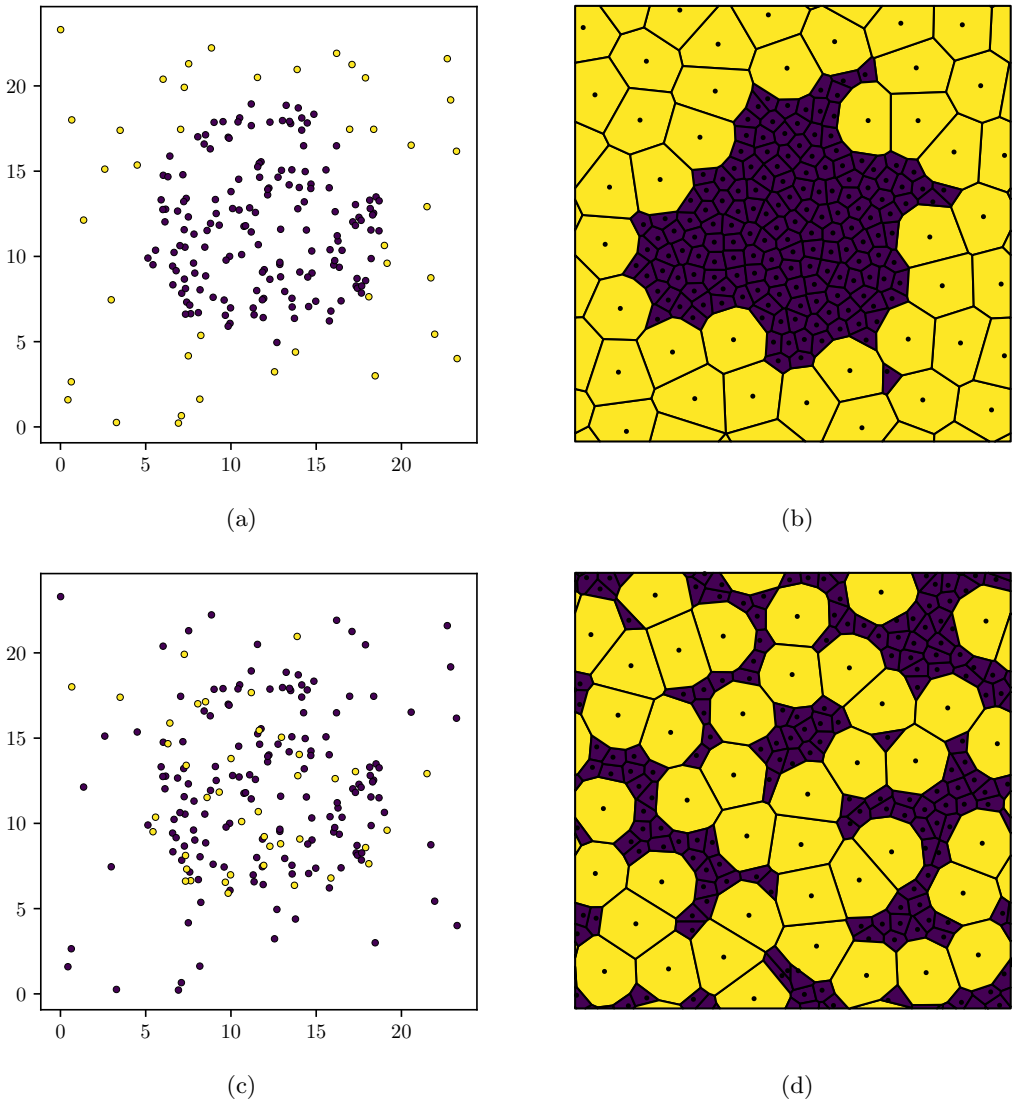


Figure 4.2: Example of choosing the initial generator points for Algorithm 1 dependent or independent of the cell areas. Purple cells have area 1 ($m_i = 1$), with corresponding purple generator points. Yellow cells have area 10 ($m_i = 10$), with corresponding yellow generator points. (a): Purple points are located inside a circle, yellow points are located outside this circle. (b): The resulting Laguerre diagram of the initial points in (a). (c): The same generator points as in (a) but the colors (and therefore the assigned areas) are now randomly shuffled. (d): The resulting Laguerre diagram of the initial points in (c).

1 is visualized in Figure 4.2 (d). This shows that if the initial generator points and cell volumes are chosen independently we do not necessarily see a cluster. As clusters are undesirable for our purposes we will from now take the initial generator points independently from the cell sizes. Cell size meaning its area for 2D Laguerre diagrams and its volume in the 3D case. We do remark that some materials may have clusters or bands of similarly sized grains, materials with multiple phases for example. For such materials it is useful to know that Laguerre diagrams with clusters can be generated with Algorithm 1.

Remark. *We usually define* $l = \sqrt[3]{\sum_{i=1}^{n} m_i}$*, where the* m_i*'s are sampled from a volume distribution (3D case). Then, as a domain the cube with side length* l*:* $\Omega = [0, l]^3 = [0, l] \times [0, l] \times [0, l]$ *is taken. As a result there is a dependence between* Ω *and the cell volumes. Therefore, we may choose the initial generator points independent from the cell volumes, conditional on the domain.*

While a single example does not reveal all the properties of Algorithm 1, in this particular case we see that the algorithm does not necessarily move the generator points a lot. If this is in fact a property of the algorithm then it seems reasonable to spread the initial generator points fairly evenly over the domain, and to randomly assign a volume m_i to each point. In the next subsection we explore a few possible choices for the initial generator points.

4.2.1 Analyzing the influence of the initial generator points via hypothesis testing

In this subsection we investigate how the choice of the initial generator points affects the resulting Laguerre diagram as well as cross sections of these diagrams. Specifically, we look at the distribution of areas we observe in a cross section for these Laguerre diagrams. We refer to this distribution as the cross sectional area distribution. Via hypothesis testing we assess whether two samples of cross sectional areas, each corresponding to a different cross section, may be assumed to originate from the same (unknown) distribution. Besides the hypothesis testing we use visualizations of these Laguerre diagrams and cross sections to assess whether a particular choice of initial generator points is useful for our purposes.

Consider the following simulation setting. The discrete distribution in (4.8) is used as the volume distribution of the (3D) Laguerre diagram. We take $n = 4000$ cells, and sample $m_1, \ldots, m_n$ from the aforementioned volume distribution. Define $l = \sqrt[3]{\sum_{i=1}^{n} m_i}$ and take as a domain the cube with side length l: $\Omega = [0, l]^3$. Then, the *same sample of the* m_i*'s* is used in combination with the following different choices for the initial generator points:

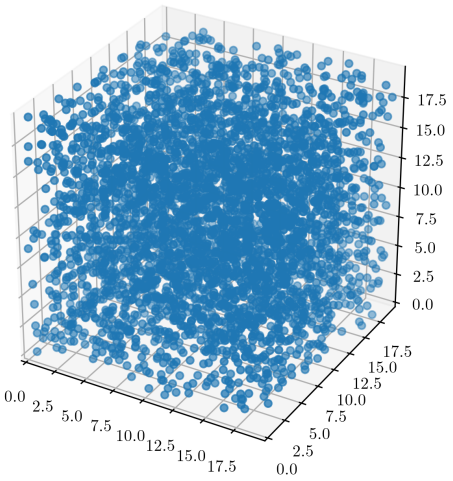
P.1 As a baseline we consider sampling the initial points from the uniform distribution over the domain: $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)} \overset{\text{iid}}{\sim} \boldsymbol{U}([0, l]^3)$.

P.2 The uniform distribution over a thin slice in the domain: $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)} \overset{\text{iid}}{\sim} \boldsymbol{U}([0, l] \times [0, l] \times [\frac{12}{25}l, \frac{13}{25}l])$.

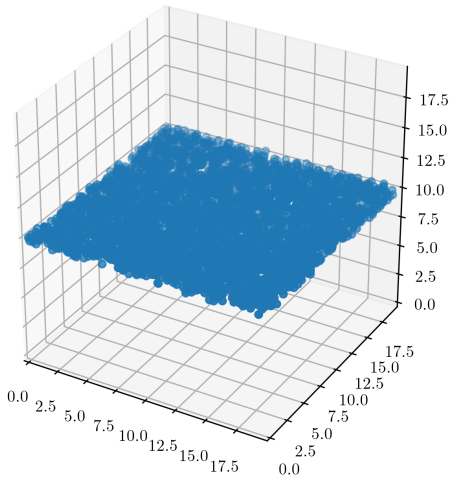
P.3 The uniform distribution over a small cube within the domain: $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \ldots, \boldsymbol{x}_n^{(0)} \overset{\text{iid}}{\sim} \boldsymbol{U}([\frac{9}{20}l, \frac{11}{20}l]^3)$.

P.4 The uniform distribution over a plane in the domain, we take: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \overset{\text{iid}}{\sim} \boldsymbol{U}([0, l]^2)$ and define $\boldsymbol{x}_j^{(0)} = (x_j, y_j, \frac{l}{2})$ for $j \in \{1, \ldots, n\}$.

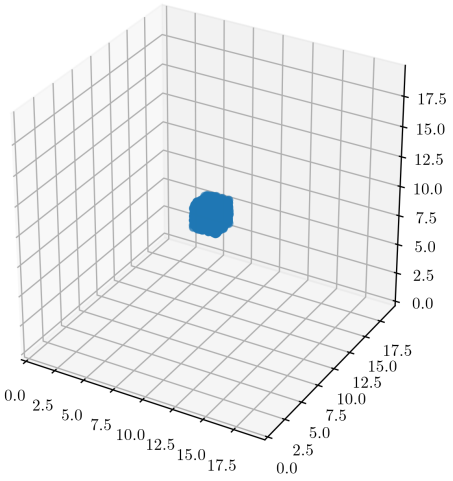
A visualization of these choices for the initial points can be seen in Figure 4.3. The next step is to generate the Laguerre diagrams for each choice of the initial generator points with Algorithm 1. A visualization of the resulting diagrams is shown in Figure 4.4.



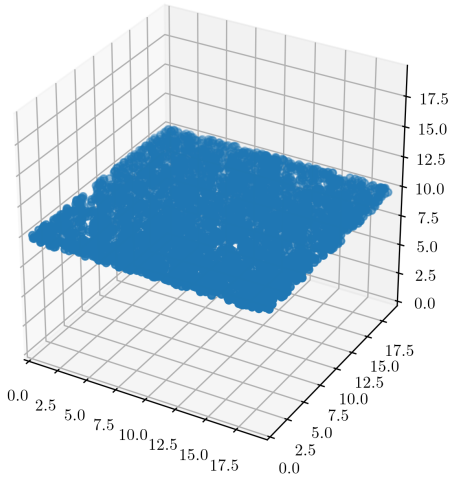
(a) **P.1**



(b) **P.2**

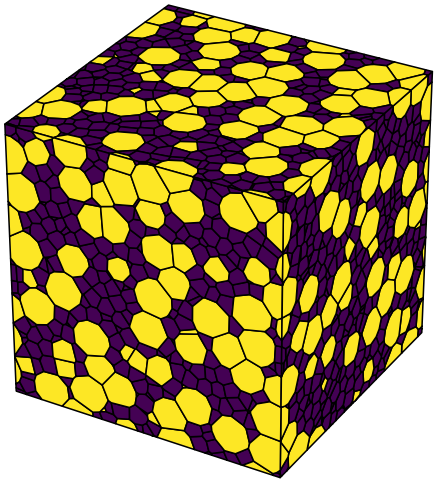


(c) **P.3**

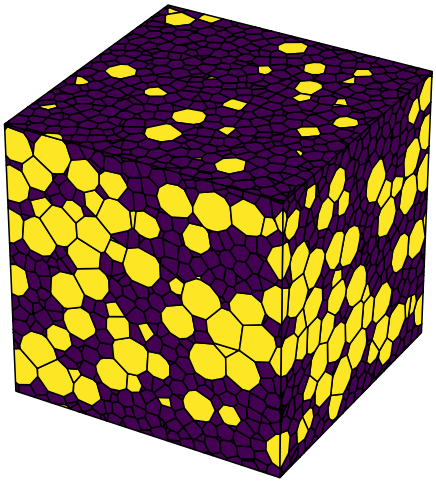


(d) **P.4**

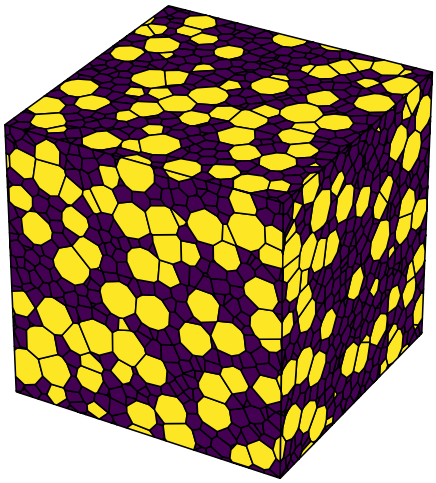
Figure 4.3: A visualization of different choices for the initial generator points of the Laguerre diagram. (a): The uniform distribution over the domain. (b): The uniform distribution over a thin slice. (c): The uniform distribution over a small cube within the domain. (d): The uniform distribution over a plane in the domain.



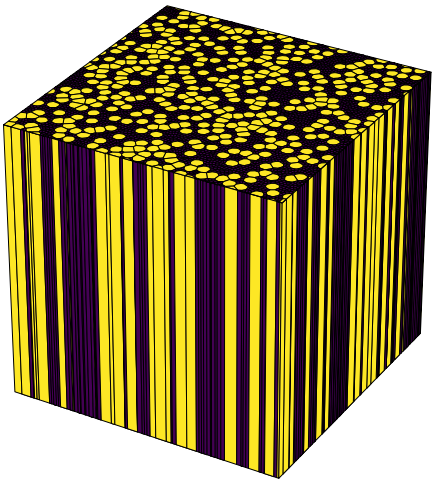
(a) **P.1**



(b) **P.2**

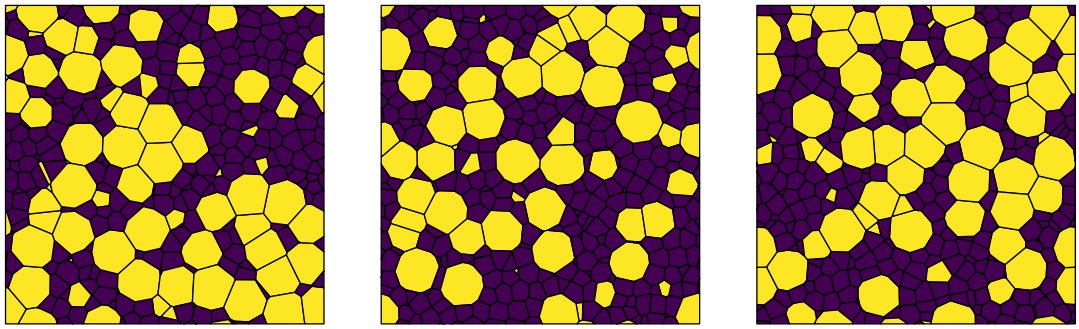


(c) **P.3**

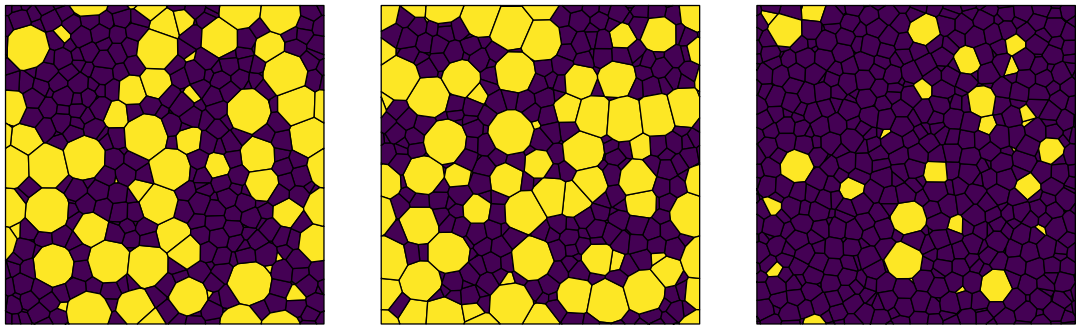


(d) **P.4**

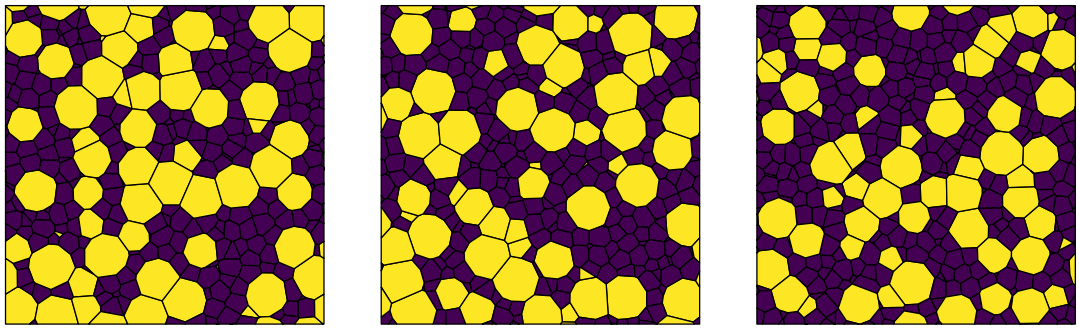
Figure 4.4: A visualization of Laguerre diagrams generated using Algorithm 1 for different choices of the initial generator points. Below each figure the chosen initial points are displayed, either **P.1**, **P.2**, **P.3** or **P.4**.



(a) **P.1**, $z = \frac{l}{3}$ (b) **P.1**, $z = \frac{2l}{3}$ (c) **P.1**, $z = l$



(d) **P.2**, $z = \frac{l}{3}$ (e) **P.2**, $z = \frac{2l}{3}$ (f) **P.2**, $z = l$



(g) **P.3**, $z = \frac{l}{3}$ (h) **P.3**, $z = \frac{2l}{3}$ (i) **P.3**, $z = l$

Figure 4.5: A visualization of cross sections taken of the Laguerre diagrams shown in Figure 4.4, taken at different z-levels. Below each figure the chosen initial points are displayed, either **P.1**, **P.2** or **P.3**. Below each figure we also provide the chosen height of the horizontal cross section: $z \in \{\frac{l}{3}, \frac{2l}{3}, l\}$.

Of the computed Laguerre diagrams three cross sections are taken. Specifically, cross sections of the form $z = c$ with $c \in \{\frac{l}{3}, \frac{2l}{3}, l\}$. Note that due to periodic boundary conditions, taking the cross section $z = l$ is the same as taking the cross section $z = 0$. The resulting cross sections for the diagrams corresponding to initial points **P.1**, **P.2** and **P.3** are shown in Figure 4.5.

For the Laguerre diagram with initial points **P.4** the height of the cross section does not matter, each of these sections are the same. Such a section is shown in Figure 4.6. From Figure 4.4 we observe that **P.4** is not the kind of Laguerre diagram we are looking for. If a cross section of the form $x = c$ was taken instead, this would result in a very different looking cross section. This diagram has a somewhat artificial looking structure which only has elongated cells. This result can be explained, by putting all the initial generator points in the plane $z = \frac{l}{2}$ the algorithm will first find the right weights such that all cells have the desired volume. The result of this is that the boundary between neighboring cells, which are associated with points $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ is perpendicular to the line which connects $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The next step of the algorithm is to compute the centroids and because the cells all have this elongated shape the centroids have z-component $\frac{l}{2}$. As this persists throughout all iterations of the algorithm, the generator points never move in the z-direction. Such a diagram may be useful for modelling materials with elongated grains, but it does not suit our purposes.

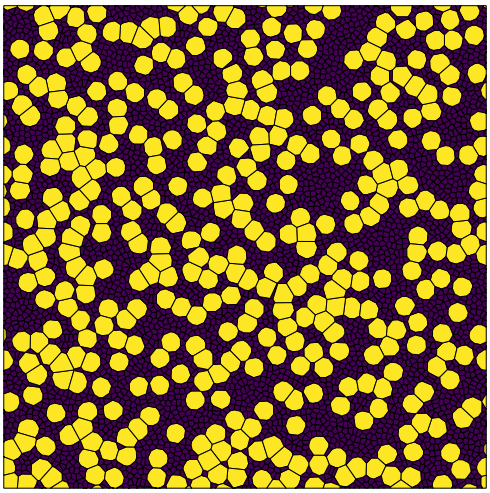


Figure 4.6: A visualization of a horizontal cross section of the Laguerre diagram which is visualized in Figure 4.4 with initial points **P.4**.

Another observation we can make from Figure 4.4 is that **P.2** seems to have resulted in a Laguerre diagram which has very few large grains near the top of the diagram, especially when compared to **P.1**. Table 4.3 shows the number of cells which were observed in each of the computed cross sections. This table also reflects that **P.2** might lead to a Laguerre diagram with clusters of similarly sized cells, as the cross section at $z = l$ seems to contain disproportionately many (small) cells, compared to the cross sections at other heights of the same diagram. As expected, Table 4.3 shows that when using initial points **P.4** we always observe all $n = 4000$ cells.

Initial points	$z = \frac{l}{3}$	$z = \frac{2l}{3}$	$z = l$
P.1	316	346	304
P.2	340	319	488
P.3	304	319	358
P.4	4000	4000	4000

Table 4.3: Number of observed cells in each cross section when considering the Laguerre diagrams shown in Figure 4.4. The number of observed cells is given for every combination of the options for the initial generator points and cross section heights.

While the images of the Laguerre diagrams and cross sections are useful we now take a more quantitative approach. As samples of areas in cross sections do not seem to follow known parametric distributions, we consider nonparametric tests. These tests also allow the comparison of differently sized samples, which is important since the number of cells we observe in cross sections is also random. Let us consider the k-sample Anderson-Darling test, as described in [26]. Suppose we have k samples of independent observations. Let X_{ij} be the jth observation in the ith sample ($j \in \{1, \dots, n_i\}$ and $i \in \{1, \dots, k\}$). The ith sample is a sample of the distribution F^i. We consider the following hypotheses:

$$H_0 : F^1 = F^2 = \cdots = F^k$$
$$H_1 : F^r \neq F^s \text{ for some } r, s \in \{1, \dots, k\} \text{ with } r \neq s.$$

Let $F^i_{n_i}(x)$ denote the empirical distribution function of the ith sample. Let $H_N(x)$ be the empirical distribution of the combined (pooled) samples, hence $N = n_1 + n_2 + \cdots + n_k$. The k-sample Anderson-Darling test statistic is then defined as:

$$A^2_{kN} = \sum_{i=1}^{k} n_i \int_{B_N} \frac{\left(F^i_{n_i}(x) - H_N(x)\right)^2}{H_N(x)(1 - H_N(x))} \mathrm{d}H_N(x).$$

Here $B_N = \{x \in \mathbb{R} : H_N(x) < 1\}$. The statistic A^2_{kN} is then normalized to yield the following test statistic:

$$T_{kN} = \frac{A^2_{kN} - (k-1)}{\sqrt{\mathrm{Var}(A^2_{kN})}}.$$

For a computational formula of the test statistic, as well as the computation of p-values we refer to [26]. In all hypothesis tests, the confidence level $\alpha = 0.05$ is used. Consider the Laguerre diagrams computed with initial generator point **P.1**, **P.2** and **P.3**. For each of these diagrams, we have three samples of cross section areas, each sample corresponding to a distinct cross section. This test is used to determine whether these three samples of areas may be assumed to originate from the same distribution. If this may be assumed according to the test, for a particular set of initial generator points, then this is evidence that this choice of generators leads to a Laguerre diagram which fits our purposes. The results of the test are in Table 4.4. The reason for omitting **P.4** in this test is that the corresponding samples of cross sections are clearly not independent, there is in fact full dependence. For the other choices of initial points the cross sections are spaced far enough that we do not observe the same cell in multiple cross sections. Therefore, it does seem reasonable to assume independence in these cases.

Initial points	AD statistic	AD p-value
P.1	1.68	0.53
P.2	7.57	0.0015
P.3	2.05	0.38

Table 4.4: The results of the k-sample Anderson-Darling test, using the three cross section samples as computed for each Laguerre diagram. The Laguerre diagrams considered are the ones with the initial generator points **P.1**,**P.2** and **P.3** as shown in Figure 4.4.

Table 4.4 suggests that for **P.2**, the samples of the cross sections should not be assumed to come from the same distribution as $p < 0.05$. This is not the case for **P.1** and **P.3**, here we do not reject the null hypothesis. Therefore, the options **P.1** and **P.3** both seem to be reasonable choices as initial generator points.

Another interesting test, is whether the resulting Laguerre diagrams may be considered "the same". Given two Laguerre diagrams, each generated with different initial points, we compare the cross sections of these diagrams which were taken at the same height. Besides the k-sample Anderson-Darling test with $k = 2$ we use the two-sample Kolmogorov-Smirnov test. This test uses the following test statistic (using the same notation as for the Anderson-Darling test):

$$D_{n_1 n_2} = \max_{x \in \mathbb{R}} |F^1_{n_1}(x) - F^2_{n_2}(x)|. \tag{4.9}$$

The choice of initial points **P.1** is taken as the baseline. Hence, for the other three Laguerre diagrams we take the sample of areas in the cross section $z = c$ and compare this sample to the sample of areas in the cross section $z = c$ of the Laguerre diagram with initial points **P.1**. The results of this test are in Table 4.5.

As expected, comparing a cross section of the Laguerre diagram with initial points **P.4** to any of the cross sections of the diagram with initial points **P.1** leads to rejection of H_0. Hence, these distributions of cross section areas should not be assumed to originate from the same distribution. Other than that there is only one sample which lead to rejection of the null hypothesis. The sample of areas in section $z = l$ of Laguerre diagram with initial points **P.2** should not be

Cross section height	Initial points	KS statistic	KS p-value	AD statistic	AD p-value
$z = \frac{l}{3}$	**P.2**	0.06	0.66	0.86	0.44
	P.3	0.06	0.59	0.62	0.63
	P.4	0.82	0	480.44	3.9e-264
$z = \frac{2l}{3}$	**P.2**	0.05	0.89	0.30	0.94
	P.3	0.07	0.36	0.71	0.55
	P.4	0.83	0	535.97	1.2e-294
$z = l$	**P.2**	0.13	0.0036	6.16	0.00077
	P.3	0.07	0.32	1.32	0.23
	P.4	0.81	0	484.18	3.4e-266

Table 4.5: The results of testing whether cross section area distributions come from the same distribution when comparing cross sections of the Laguerre diagrams in Figure 4.4. We compare one of the choices of the initial generator points **P.2**, **P.3** or **P.4** against choosing **P.1**. Both the two-sample Kolmogorov-Smirnov (KS) test and the two-sample Anderson-Darling (AD) test are used.

assumed to come from the same distribution as the sample of areas in the section $z = l$ of the diagram with initial points **P.1**. This was also to be expected, this particular cross section appears very different from all other sections which are shown in Figure 4.5. From the tests which were performed it seems that for the cell volume distribution (4.8), both **P.1** and **P.3** are reasonable choices.

We now consider a more interesting volume distribution for the cells of the Laguerre diagrams. We sample $m_1, \ldots, m_n \overset{\text{iid}}{\sim} F_V$ with F_V being the lognormal distribution, with $\sigma = 0.8$, $\mu = -\frac{\sigma^2}{2}$. As we have seen that **P.4** does not result in a Laguerre diagram which is fit for our purposes we no longer consider this option. Otherwise, we proceed exactly as before. This means that the same sample of the m_i's is used to generate Laguerre diagrams in combination with the different choices of initial points **P.1**, **P.2** and **P.3**. Then, cross sections are taken at the same heights as before, and we perform the same hypothesis tests as before.

The resulting Laguerre diagrams are shown in Figure 4.7, and the cross sections can be seen in Figure 4.8. Judging from the images in these figures, most Laguerre diagrams appear quite similar to one another, the same can be said for the cross sections. The only cross section which looks slightly different is the section $z = l$ of the Laguerre diagram with initial points **P.2**. Table 4.6 shows similar results as in Table 4.3, it seems that the aforementioned cross section contains noticeably more (small) cells compared to the other sections. As before, for each diagram we test whether the three samples of cross section areas, each sample corresponding to a specific section, may be assumed to originate from the same distribution. The results of this test are shown in Table 4.7.

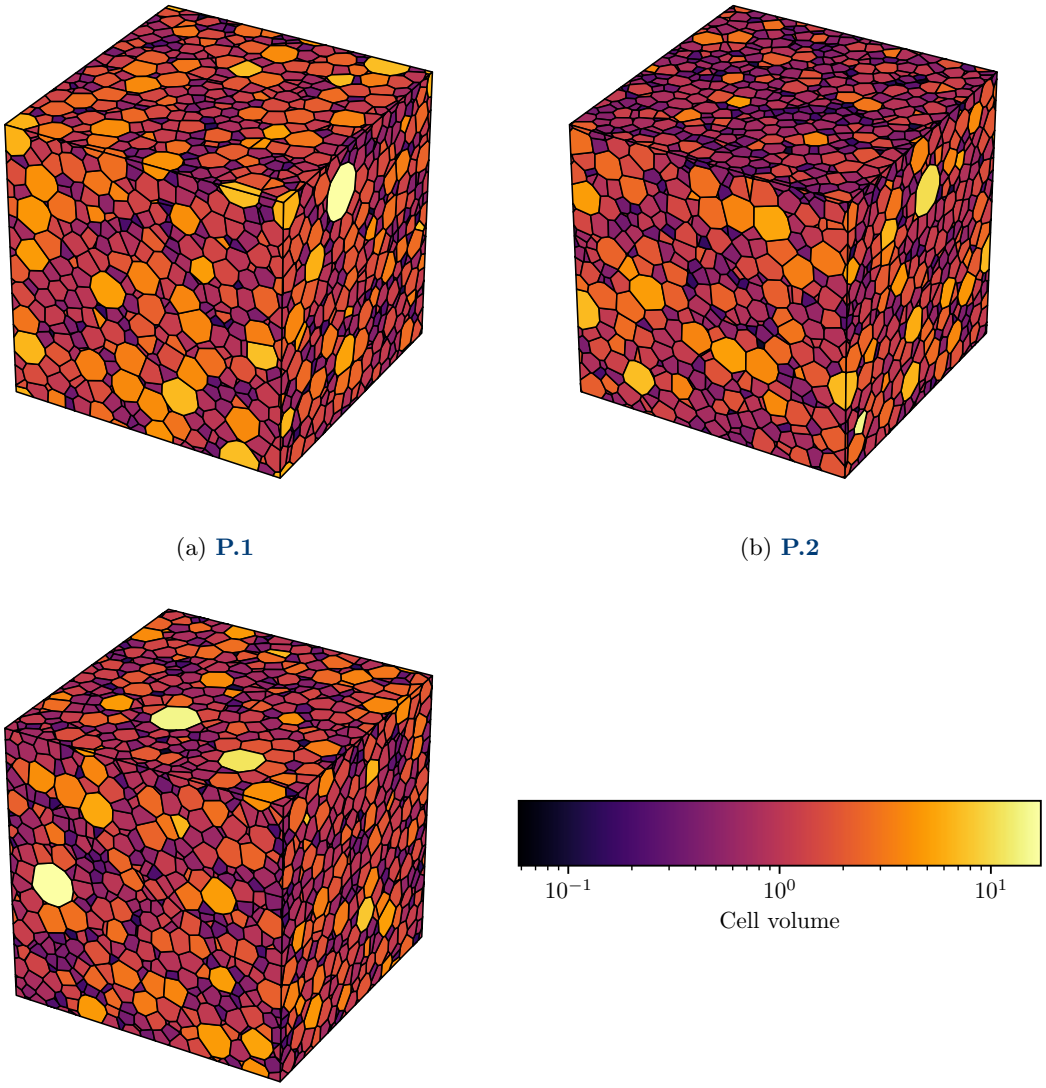
Initial points	$z = \frac{l}{3}$	$z = \frac{2l}{3}$	$z = l$
P.1	339	343	335
P.2	312	325	465
P.3	299	340	359

Table 4.6: Number of observed cells in each cross section when considering Laguerre diagrams shown in Figure 4.7. The number of observed cells is given for every combination of the options for the initial generator points (except for **P.4**) and cross section heights.

Initial points	AD statistic	AD p-value
P.1	0.90	0.92
P.2	23.24	4.4e-12
P.3	5.59	0.011

Table 4.7: The results of the k-sample Anderson-Darling test, using the three cross section samples as computed for each Laguerre diagram. The Laguerre diagrams considered are the ones with the initial generator points **P.1**,**P.2** and **P.3** as shown in Figure 4.7.

In Table 4.7 we observe that this lognormal cell volume distribution gives a slightly different result. In this case we see that the cross sections associated with **P.2** and **P.3** both result in rejecting the null hypothesis. In the other distribution we considered **P.3** did not lead to rejection. On the other hand, these results provide evidence that **P.1** is a reasonable choice.

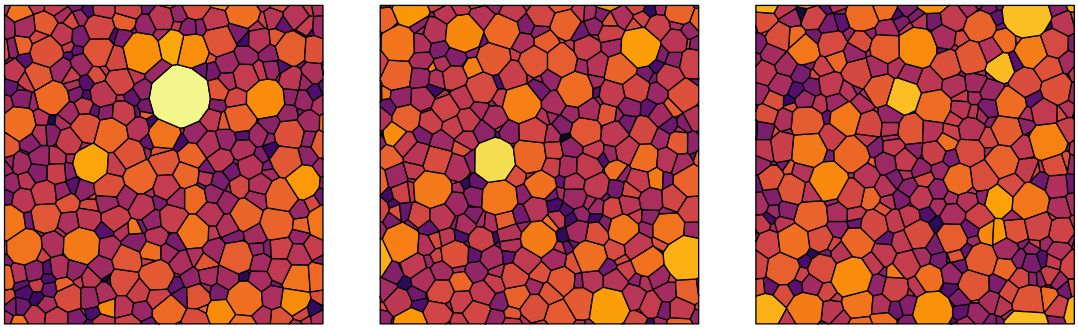


(a) **P.1**

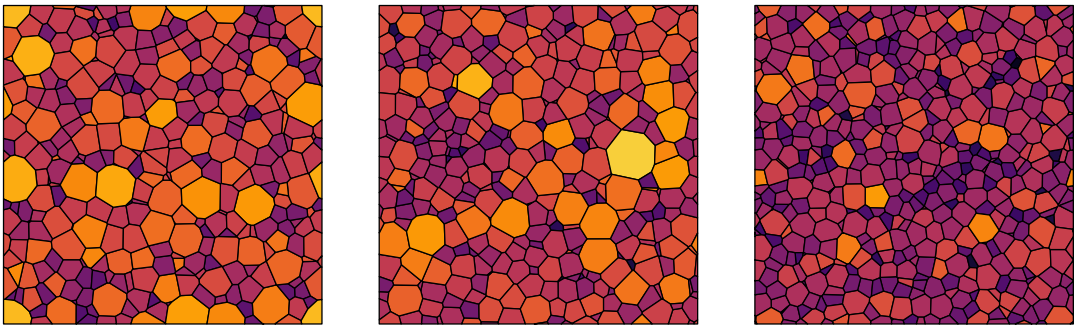
(b) **P.2**

(c) **P.3**

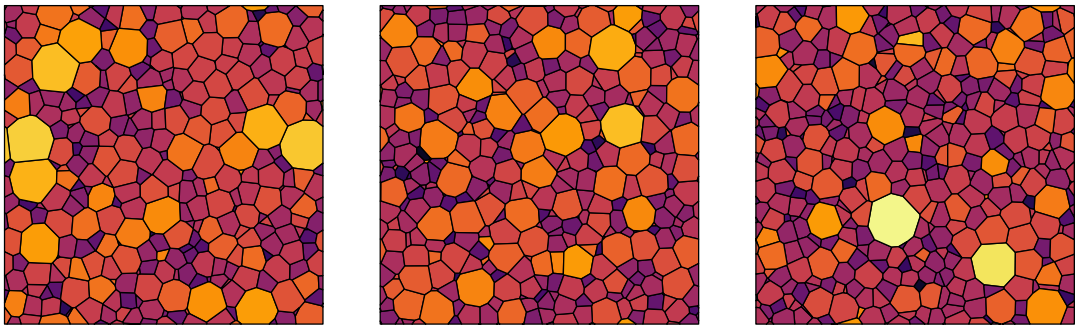
Figure 4.7: A visualization of Laguerre diagrams with a lognormal cell volume distribution ($\sigma = 0.8$, $\mu = -\frac{\sigma^2}{2}$), generated using Algorithm 1 for different choices of the initial generator points. Below each figure the chosen initial points are displayed, either **P.1**, **P.2** or **P.3**.



(a) **P.1**, $z = \frac{l}{3}$ (b) **P.1**, $z = \frac{2l}{3}$ (c) **P.1**, $z = l$



(d) **P.2**, $z = \frac{l}{3}$ (e) **P.2**, $z = \frac{2l}{3}$ (f) **P.2**, $z = l$



(g) **P.3**, $z = \frac{l}{3}$ (h) **P.3**, $z = \frac{2l}{3}$ (i) **P.3**, $z = l$

Figure 4.8: A visualization of cross sections taken of the Laguerre diagrams shown in Figure 4.7, taken at different z-levels. Below each figure the chosen initial points are displayed, either **P.1**, **P.2** or **P.3**. Below each figure we also provide the chosen height of the horizontal cross section: $z \in \{\frac{l}{3}, \frac{2l}{3}, l\}$.

Cross section height	Initial points	KS statistic	KS p-value	AD statistic	AD p-value
$z = \frac{l}{3}$	**P.2**	0.08	0.27	1.38	0.21
	P.3	0.09	0.15	2.19	0.072
$z = \frac{2l}{3}$	**P.2**	0.07	0.36	0.97	0.37
	P.3	0.05	0.83	0.53	0.72
$z = l$	**P.2**	0.16	0.00011	11.54	1.4e-06
	P.3	0.07	0.37	1.23	0.26

Table 4.8: The results of testing whether cross section area distributions come from the same distribution when comparing cross sections of the Laguerre diagrams in Figure 4.7. We compare one of the choices of the initial generator points **P.2** or **P.3** against choosing **P.1**. Both the two-sample Kolmogorov-Smirnov (KS) test and the two-sample Anderson-Darling (AD) test are used.

We repeat another test for this lognormal volume distribution. For two Laguerre diagrams, one using initial points **P.1** and the other with another set of initial points, we compare the cross sections of these diagrams which were taken at the same height. The two-sample Kolmogorov-Smirnov test and two-sample Anderson-Darling test are used again, the results are displayed in Table 4.8. While the values of the test statistics and p-values are not the same, given the confidence level $\alpha = 0.05$ results in the same conclusion as before. The only sample of cross sectional areas which is quite different from the others is $z = l$ of the Laguerre diagram with initial points **P.2**.

In this section we assessed a few possible choices of initial generator points, both visually and via hypothesis tests. The choice **P.1** seems to be appropriate for our purposes. The Laguerre diagrams computed with this choice for the initial points satisfy the conditions that the volumes of the cells are randomly distributed throughout the diagram. Specifically, for these diagrams the location of the cross section is not relevant. Different cross sections yielded samples of areas which may be assumed to originate from the same unknown distribution. [26] also suggests that the k-sample Anderson-Darling test may be used to argue that multiple samples may be pooled together to obtain a larger sample. In some later chapters this is used, particularly in cases where a large sample of cross section areas is desired. In those cases, to minimize the possibility of dependency between samples, we do not take multiple cross sections of the same Laguerre diagram. Instead, multiple Laguerre diagrams will be generated, taking only a single cross section of each diagram. This approach is made more precise in the final section of this chapter. The tests also suggest that **P.3** could be a reasonable choice. We do remark that **P.1** is a more practical choice, the gradient methods in Algorithm 2 and 3 tend to converge quicker for this choice of initial generator points in comparison to the other choices we consider.

4.3 Accelerating gradient methods via a permutation step

In this section we briefly discuss a method which accelerate gradient methods such as Algorithm 2 and 3 in some specific cases. First, the intuition behind the method is given, and then we precisely describe how the method works. Via simulations its strengths and weaknesses are highlighted.

The idea of the permutation step we propose is as follows. In practice, the location of a cell with a particular volume within the diagram is not that important. What is important, is

that the cell volumes in the Laguerre diagram follow a predetermined distribution. In Algorithm 1 we prescribe the volumes $m_1, \ldots, m_n$ which results in a Laguerre diagram such that cell L_i has volume m_i. Recall that we use a gradient method which iteratively updates the weights, eventually resulting in weights such that the cells have the desired volumes. The permutation step we propose rearranges the m_i's after each iteration of the gradient method, such that this rearrangement minimizes the Euclidean norm of the gradient. When implemented in a gradient method this means that the resulting Laguerre diagram still has cells with volumes $m_1, \ldots, m_n$, but cell L_i does not necessarily have volume m_i.

Consider the fixed set of generator points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Let $\{L_i\}_{i=1}^n$ be the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$. The function $f : \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$f(\boldsymbol{w}) := -\sum_{i=1}^{n}(m_i - |L_i(\boldsymbol{w})|)w_i - \sum_{i=1}^{n}\int_{L_i(\boldsymbol{w})} ||\boldsymbol{x} - \boldsymbol{x}_i||^2 \mathrm{d}\boldsymbol{x}. \tag{4.10}$$

As we have seen before, its gradient is given by:

$$\nabla f(\boldsymbol{w}) = (|L_1(\boldsymbol{w})| - m_1, |L_2(\boldsymbol{w})| - m_2, \ldots, |L_n(\boldsymbol{w})| - m_n).$$

Let $\mathcal{P}$ be the set of all permutations on $\{1, \ldots, n\}$. Each iteration of the gradient algorithm we permute the targeted volumes $m_1, \ldots, m_n$ such that the norm of the gradient is smaller then it would have been had we not done this:

$$\min_{\pi \in \mathcal{P}} \sqrt{\sum_{i=1}^{n} \left(m_{\pi(i)} - |L_i(\boldsymbol{w})|\right)^2} \leq \sqrt{\sum_{i=1}^{n} \left(m_i - |L_i(\boldsymbol{w})|\right)^2} = ||\nabla f(\boldsymbol{w})||.$$

When implementing this in the Barzilai-Borwein method (Algorithm 2) we obtain Algorithm 4. The reason that the computation of the permutation which minimizes the norm of the gradient is tractable is that it can be found via sorting. Specifically, given $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ consider:

$$\min_{\pi, \tau \in \mathcal{P}} \sum_{i=1}^{n} \left(x_{\pi(i)} - y_{\tau(i)}\right)^2. \tag{4.11}$$

It can easily be shown that this is minimized by permutations π and τ which sort the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. Meaning that: $x_{\pi(1)} \leq x_{\pi(2)} \leq \cdots \leq x_{\pi(n)}$ and $y_{\tau(1)} \leq y_{\tau(2)} \leq \cdots \leq y_{\tau(n)}$. When instead looking at:

$$\min_{\sigma \in \mathcal{P}} \sum_{i=1}^{n} \left(x_{\sigma(i)} - y_i\right)^2. \tag{4.12}$$

We may take permutations π and τ which minimize (4.11) and we obtain that (4.12) is minimized by the permutation σ which is defined as:

$$\sigma(\pi(i)) = \tau(i) \quad \text{for } i \in \{1, \ldots, n\}.$$

Algorithm 4 Barzilai-Borwein method with permutation step

Input: Initial weights: $\boldsymbol{w}_0 \in \mathbb{R}^n$, initial step size: $\alpha_0 > 0$, The function f as in (4.10).
Output: $\boldsymbol{w} \in \mathbb{R}^n$ such that $||\nabla f(\boldsymbol{w})||_\infty < \text{tol}_f$

residual $\leftarrow ||\nabla f(\boldsymbol{w}_0)||_\infty$
$j \leftarrow 0$
while $j < \text{maxit}_f$ **and** residual $> \text{tol}_f$ **do**
  $\boldsymbol{w}_{j+1} \leftarrow \boldsymbol{w}_j - \alpha_j \nabla f(\boldsymbol{w}_j)$
  $\pi \leftarrow \arg\min_{\pi \in \mathcal{P}} \sqrt{\sum_{i=1}^n \left(m_{\pi(i)} - |L_i(\boldsymbol{w}_{j+1})|\right)^2}$
  $(m_1, m_2, \ldots, m_n) \leftarrow (m_{\pi(1)}, m_{\pi(2)}, \ldots, m_{\pi(n)})$
  $\alpha_{j+1} \leftarrow \frac{||\boldsymbol{w}_{j+1} - \boldsymbol{w}_j||^2}{(\nabla f(\boldsymbol{w}_{j+1}) - \nabla f(\boldsymbol{w}_j))^T (\boldsymbol{w}_{j+1} - \boldsymbol{w}_j)}$
  residual $\leftarrow ||\nabla f(\boldsymbol{w}_{j+1})||_\infty$
  $j \leftarrow j + 1$
end while
return $\boldsymbol{w}_j$

Let us first consider a case in which the permutation step does indeed provide a speed-up. For the cell volume distribution F_V we take a lognormal volume distribution with $\sigma = 0.6$ and $\mu = -\frac{\sigma^2}{2}$. For the initial generator points we stick to the choice **P.1**. Given a sample $m_1, \ldots, m_n \overset{\text{iid}}{\sim} F_V$, we take the domain $\Omega = [0, l]^3$ with $l = \sqrt[3]{\sum_{i=1}^n m_i}$. The same sample of initial generator points and the same sample of volumes is used both with Algorithm 2 and with Algorithm 4. Essentially, we run Algorithm 1 for just one iteration, such that we only measure the performance of the gradient methods. The iteration counts and run-times for the Barzilai-Borwein method with and without the permutation step are displayed in Table 4.9 for different values of n.

	BB		BB with permutation	
n	Iterations	Time (s)	Iterations	Time (s)
1000	126	0.375	39	0.126
2000	229	1.45	18	0.201
4000	533	8.20	27	0.479
8000	736	29.0	16	0.734
16000	1978	214.0	21	2.52

Table 4.9: A comparison of the Barzilai-Borwein method (BB) with and without the permutation step (Algorithm 2 and 4 respectively). Laguerre diagrams with a lognormal cell volume distribution are generated, with $\sigma = 0.6$ and $\mu = -\frac{\sigma^2}{2}$ for different amounts of cells n.

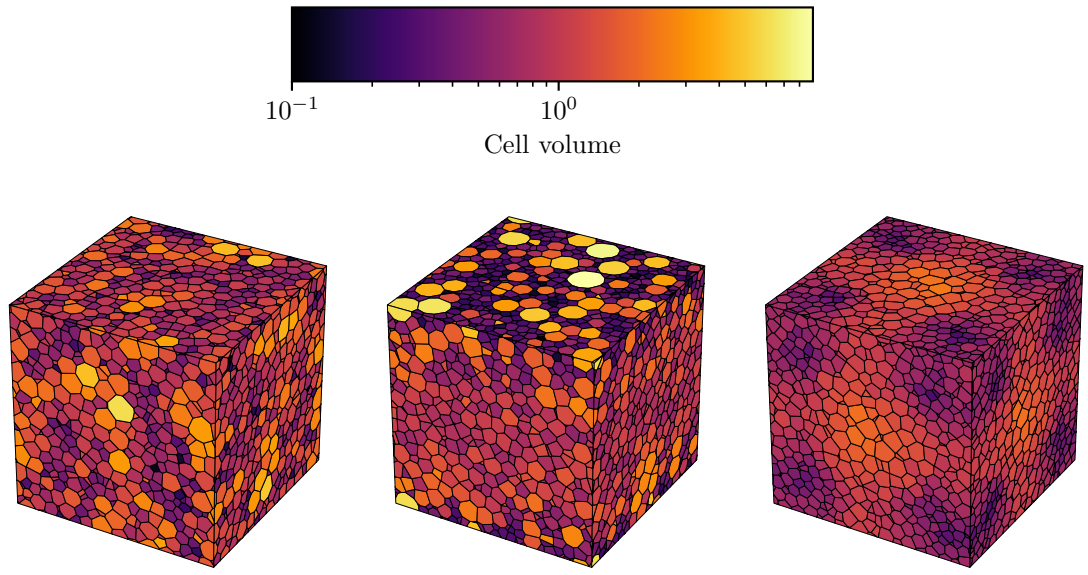
It can clearly be seen that the permutation step provides a significant reduction in both the amount of iterations as well as the run-time of the algorithm. We also observe that when using this permutation step the amount of iterations does not necessarily scale with n, which is the case when omitting the permutation step. It should be stressed that each number presented in Table 4.9 is the result of only a single run, they are not averages over multiple runs. Let us now consider a simulation setting which shows when the permutation step should not be used. Now the amount of cells is fixed as $n = 4000$. Different values of σ are considered for the lognormal volume distribution while keeping $\mu = -\frac{\sigma^2}{2}$. Other than that the simulation setting is the same as before. The results are presented in Table 4.10. We observe that for large values of σ the permutation step slows down the Barzilai-Borwein method a lot. It is clear that for large σ we

do not want to use the permutation step. For small values of σ the permutation step significantly speeds up the gradient method.

	BB		BB with permutation	
σ	Iterations	Time (s)	Iterations	Time (s)
0.3	299	4.55	7	0.144
0.6	559	8.68	17	0.311
0.9	968	15.3	125	2.08
1.2	572	9.29	2293	40.5
1.5	2931	54.5	20763	376.0

Table 4.10: A comparison of the Barzilai-Borwein method (BB) with and without the permutation step (Algorithm 2 and 4 respectively). Laguerre diagrams with a lognormal cell volume distribution are generated, with different values of σ and $\mu = -\frac{\sigma^2}{2}$ with $n = 4000$ cells.

We do not perform an extensive analysis of the influence of the initial generator points when Algorithm 1 is used in combination with Algorithm 4 as before. We do visualize the resulting Laguerre diagrams when repeating the simulation with a lognormal volume distribution in section 4.2.1, but now with the inclusion of the permutation step. Recall that for this simulation the same sample of volumes $m_1, \ldots, m_n$ is used in combination with the different choices for the initial generator points **P.1**, **P.2** and **P.3**. The resulting diagrams are shown in Figure 4.9. This highlights that the permutation step makes the gradient method much more sensitive for the choice of initial generator points. It could be useful when Laguerre diagrams with very specific



(a) **P.1** (b) **P.2** (c) **P.3**

Figure 4.9: A visualization of Laguerre diagrams with a lognormal cell volume distribution ($\sigma = 0.6$, $\mu = -\frac{\sigma^2}{2}$), generated using Algorithm 1 with gradient method Algorithm 4. Below each figure the chosen initial points are displayed, either **P.1**, **P.2** or **P.3**.

structure is desirable. Using **P.1** appears to result in a somewhat homogeneous distribution of the cell volumes. Of course, further tests are necessary in order to conclude this. In the case of **P.2** we especially observe that both cells with either a small volume or a large volume appear near the top and the bottom, while medium-sized cells appear in the center. The choice of initial points **P.3** seems to result in quite a curious looking Laguerre diagram indeed.

In this section we proposed a method for accelerating the gradient method which is used by Algorithm 1. Via some simulations we observed its strengths and its weaknesses. It seems particularly useful when one would like to generate a large Laguerre diagram with a low variance in the volume distribution, in the case of a lognormal volume distribution this means a small value of σ. As we cannot provide guarantees for its performance, or even its convergence for arbitrary volume distributions it will not be used for other simulations in this thesis. Further analysis of the method is out of the scope of this thesis. Nonetheless, the idea seems promising, perhaps other choices of permutations provide even better results. It could also be the case that it is more beneficial to not use the permutation step in every iteration of a gradient method, only under certain conditions.

4.4 Choosing methods for computing Laguerre diagrams and cross sections

As we approach the problem of estimating the cell volume distribution in a 3D Laguerre diagram from a cross section we have made clear that we need to precisely define how the Laguerre diagrams that we analyze are generated. In the previous sections we discussed how Algorithm 1 may be used to generate a specific class of Laguerre diagrams. As the choice of the initial generator points in this algorithm impacts the resulting Laguerre diagram we tested multiple options and concluded that sampling the initial points from a uniform distribution over the domain is a reasonable choice. In this section we provide a list of steps which states how we generate Laguerre diagrams in all further simulations in this thesis. As there are also multiple options for how a cross sections can be taken of a Laguerre diagram, we also describe how cross sections are taken in all further simulations.

4.4.1 Defining the class of Laguerre diagrams of interest

In section 4.1 we provided the parameters that are used in Algorithm 1, typically those parameters are indeed used, but in very specific cases we prefer using tighter tolerances for ε and ε_f. In those cases the chosen tolerances are of course mentioned. Let us now describe how Laguerre diagrams are generated in all simulations. Whenever we generate a Laguerre diagram we use periodic boundary conditions and we follow these three steps:

1. *Choosing volumes and defining the domain*: We choose a value for n, the number of cells or grains in the Laguerre diagram. When the true volume distribution is known, let $m_1, \dots, m_n$ denote the known volumes. However, we usually consider some probability distribution for the volume distribution and we denote the volume distribution function as F_V. Then, a sample is taken from this distribution: $m_1, \dots, m_n \overset{\text{iid}}{\sim} F_V$ for the volumes of the cells. We define $l = \sqrt[3]{\sum_{i=1}^n m_i}$ and take as a domain the cube with side length l: $\Omega = [0, l]^3 = [0, l] \times [0, l] \times [0, l]$. Then, by construction the volume of the domain is equal to the sum of the volumes: $|\Omega| = \sum_{i=1}^n m_i$.

2. *Choosing initial generator points*: As Algorithm 1 requires locations for the initial generator points we need to make a choice. We observed that sampling the initial generator points from the uniform distribution over the domain seems to be a good choice. Therefore, we sample: $\boldsymbol{x}_1^{(0)}, \boldsymbol{x}_2^{(0)}, \dots, \boldsymbol{x}_n^{(0)} \overset{\text{iid}}{\sim} \boldsymbol{U}([0, l]^3)$. As mentioned before, the initial weights for Algorithm 1 are all initialized with zero.

3. *Generating the Laguerre diagram*: All parameters that are required by Algorithm 1 are now defined, so we may use the algorithm to produce the Laguerre diagram $\{L_i\}_{i=1}^n$ with periodic boundary conditions. For the gradient method we use Algorithm 2, if it does not converge within the prescribed number of maximum iterations we use Algorithm 3. The volumes of the cells are equal to the provided volumes $m_1, \dots, m_n$ up to relative error $\varepsilon_f = 0.01$. How close the diagram is to a centroidal Laguerre diagram is determined by the tolerance $\varepsilon = 0.001$.

This class of Laguerre diagrams that are generated by these steps may be described as (approximately) centroidal Laguerre diagrams.

4.4.2 Defining the cross sectional area distribution

Now that we have a Laguerre diagram $\{L_i\}_{i=1}^n$ we give a precise description of the cross sectional area distribution of this diagram. Given the following plane:

$$P = \{(x, y, z) \in \mathbb{R}^3 : a(x - x_0) + b(y - y_0) + c(z - z_0) = 0\},$$

where $a, a_0, b, b_0, c, c_0 \in \mathbb{R}$. We may compute the intersection of each cell with the plane:

$$S_i := L_i \cap P \quad \text{for } i \in \{1, \ldots, n\}.$$

Let us now define the following index set:

$$I = \{i \in \{1, \ldots, n\} : S_i \neq \emptyset\}.$$

Then, we observe the following areas:

$$A_i = \text{area}(S_i) \quad \text{for } i \in I.$$

Therefore we consider the cross sectional area distribution to be the distribution of the A_i's conditional on $i \in I$. Hence the random variable $A_i | i \in I$ has as distribution the cross sectional area distribution. While we could consider arbitrary planes we focus on horizontal planes $z = c$ for some $c \in \mathbb{R}$. Whenever we take K cross sections of a Laguerre diagram with volume distribution F_V we use the following procedure:

For $k = 1, \ldots, K$:

1. Generate the Laguerre diagram $\{L_i\}_{i=1}^n$ with volume distribution F_V as described in section 4.4.1.
2. Sample $z_k \sim U(0, l)$.
3. Compute intersection of the cells with the plane $z = z_k$:
   $S_i = L_i \cap \{(x, y, z) \in \mathbb{R}^3 : z = z_k\}$ for $i \in \{1, \ldots, n\}$.
4. Define the index set: $I_k = \{i \in \{1, \ldots, n\} : S_i \neq \emptyset\}$.
5. Define the sample of areas of this cross section $\tilde{A}_k = \{A_i : i \in I_k\}$, where $A_i = \text{area}(S_i)$.

We then consider the sample of the cross section area distribution to be the result of combining all samples $\tilde{A}_k$ for $k \in \{1, \ldots, K\}$ into a single sample. It does not seem to be a problem to combine these samples, in section 4.2 we observed that Laguerre diagrams generated as described in 4.4.1 with the same volume distribution yield samples of sectional areas which may be assumed to come from the same (unknown) underlying distribution. The main reason for combining samples is that it is still quite computationally expensive to compute very large Laguerre diagrams. If we want to have a large sample of cross sectional areas we either need to compute a large Laguerre diagram, or we compute multiple smaller diagrams and we combine the samples of cross sectional areas as described.

4.5 Conclusion

In this chapter we discussed an algorithm for generating Laguerre diagrams with a cell volume distribution of choice. It is especially useful for our problem since we are looking for a relationship between the cell volume distribution and the cross sectional area distribution. Thanks to the algorithm the cell volume distribution can be fixed beforehand. We made choices for the parameters that are required by the algorithm, and in particular we investigated the influence of the choice of the initial generator points. We concluded the chapter by precisely describing how Laguerre diagrams are computed and how cross sections are taken in all further simulations. This means that for the direct and inverse problem we restrict ourselves to this particular class of Laguerre diagrams. We also proposed a permutation method that proved to speed up the gradient method used by Algorithm 1 for Laguerre diagrams with moderate variance in the volume distribution.

5

The role of the weights of a Laguerre diagram

Given that we have taken an elaborate look at the generator points of Laguerre diagrams it seems appropriate to consider the weights of a Laguerre diagram and investigate whether they have a simple interpretation. In section 4.2 we analyzed the influence of the choice of the initial generator points for Algorithm 1. Because this algorithm finds weights which maximize the function g (with g as in (4.4)) we do not have control over the weights. Via the initial choice of the weights, we checked if different choices for the initial weights have any impact on the final weights that are generated by Algorithm 1 but this turned out not to be the case. To be more specific, given a sample of volumes from a volume distribution and a chosen set of initial generator points the Laguerre diagram produced by Algorithm 1 did not seem to be affected by different choices of the chosen initial weights. As mentioned before we set all initial weights equal to zero, but another possible choice is:

$$w_i = \left(\frac{4m_i}{3\pi}\right)^{\frac{2}{3}},$$

as was suggested in [11]. The interpretation of this choice is that the weight w_i is initialized as the squared radius of a sphere of volume m_i. We may wonder whether the weights as produced by the algorithm have an interpretation which may be useful when considering the direct problem and the inverse problem, and that is what we investigate in this section.

First, we will simply investigate whether there is clear relationship between cell volumes, weights and sectional areas when fixing some parametric distribution for the cell volumes. Then, we consider interpreting the square root of the weights as the radii of spheres, as this is a frequently used approach when generating a Laguerre diagram via a packing of spheres. Consider for example the references [7, 6, 27, 17].

Remark. *Keep in mind that all Laguerre diagrams and cross sectional area distributions in the simulations in this chapter are computed as described in section 4.4.*

In this section we fix a parametric distribution for the cell volumes and then we investigate whether we observe any relationship between the distributions of the cell volumes, the weights and the sectional areas. To this end we consider three distributions:

1. A lognormal distribution with $\sigma = 0.6$, $\mu = -\frac{\sigma^2}{2}$.
2. An exponential distribution with $\lambda = 1$.
3. A point mass: all cells have volume 1.

Whenever we generate a Laguerre diagram with one of these cell volume distributions we generate a Laguerre diagram with $n = 1000$ cells. For the cross sectional area distribution we take $K = 10$ horizontal cross sections. In Figure 5.1 we observe the resulting volume distributions and weights distributions of a single run, while the sectional area distribution was obtained via 10 runs (as described in section 4.4.2).

When studying the distribution of the weights of a Laguerre diagram there is something very important we need to keep in mind, namely that the weights of a Laguerre diagram are not unique. Recall property **L.2** of Laguerre diagrams:

L.2 Given n distinct generator points: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \ldots, w_n \in \mathbb{R}$, the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_n, w_n)$ is the same as the Laguerre diagram generated by $(\boldsymbol{x}_1, w_1 + c), \ldots, (\boldsymbol{x}_n, w_n + c)$ for any $c \in \mathbb{R}$.

This property somewhat complicates fitting some parametric distribution to the weights. To standardize the weights of the generated Laguerre diagrams, we add the constant $c = -\min_i w_i$ to all weights such that all weights are non-negative. In Figure 5.1 we observe the result of fitting a three-parameter lognormal distribution to these standardized weights. We note that we cannot provide a physical reason why this distribution would provide a good fit. We cannot fit the ordinary lognormal distribution as this distribution is supported on $(0, \infty)$ and because of the standardization there is at least one weight equal to zero. We could try to find a better constant than c to add to all weights, but at that point we are effectively looking for the location parameter of a three-parameter lognormal distribution. The three-parameter lognormal distribution has an additional parameter γ which signifies the location shift. The usual lognormal distribution and the three-parameter variant satisfy the following property:

$$X \sim \text{Lognormal}(\mu, \sigma^2, \gamma) \iff X - \gamma \sim \text{Lognormal}(\mu, \sigma^2).$$

As can be observed in Figure 5.1, it seems that for the volume distributions we considered the distribution of the weights seems to be described reasonably well by a three-parameter lognormal distribution. The best current guess of why this provides a good fit is that this is a property of Algorithm 1. This distribution is then somehow caused by maximizing the function g and relocating the generator points to the centroids.

Another observation we can make is that the cross sectional area distribution is not in general the same as the distribution of the volumes or weights. When considering an exponential cell volume distribution it seems that we could approximate the sectional area distribution by an exponential distribution as well. When considering the sectional area distribution in the cases of a lognormal cell volume distribution, or the Laguerre diagram with cells of volume 1 we do not immediately recognize a known parametric distribution.

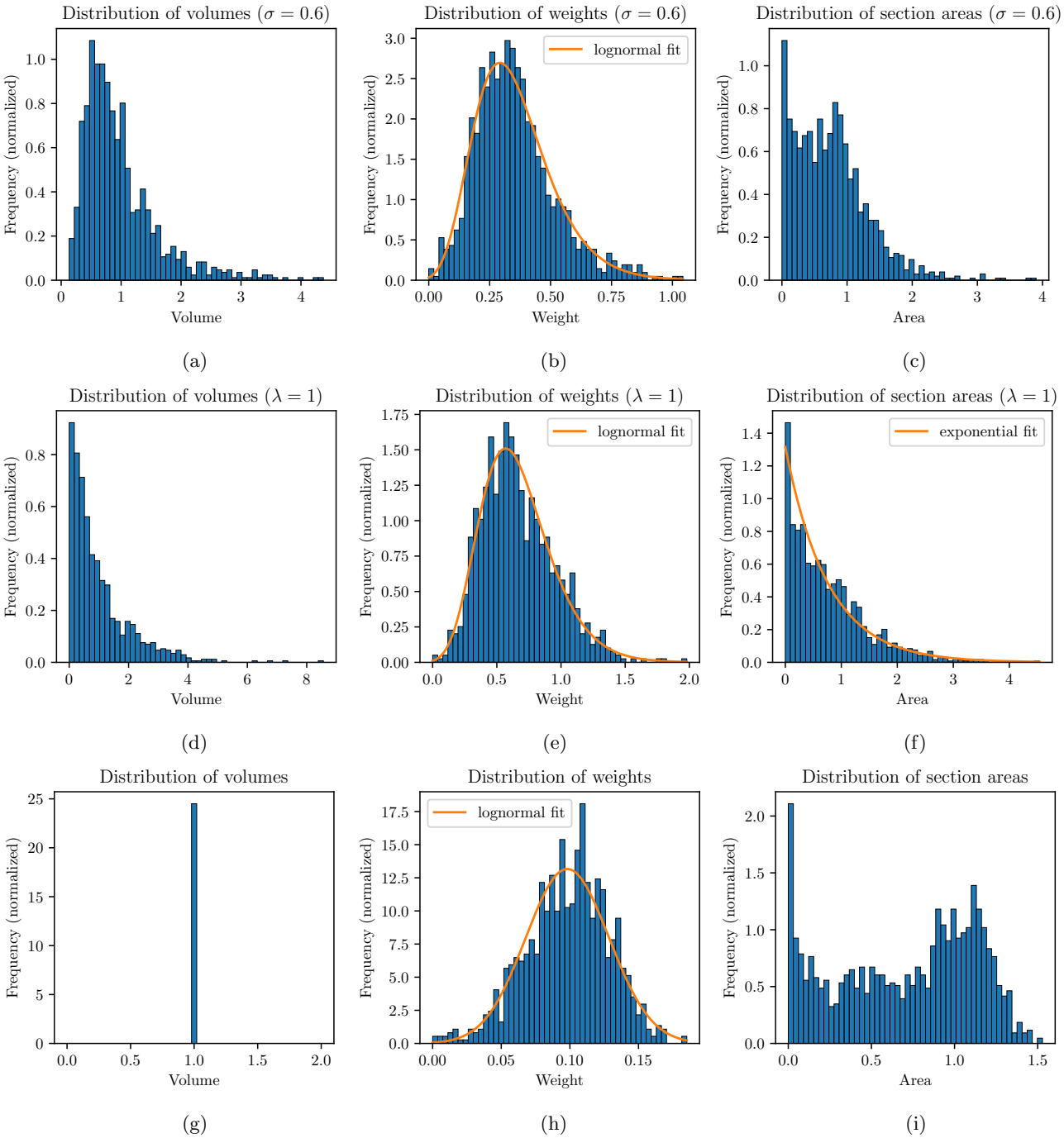


Figure 5.1: Distribution of the cell volumes, weights and cross sectional areas for Laguerre diagrams with different volume distributions. On the left: distribution of the cell volumes. In the middle: distribution of the weights. On the right: distribution of the cross sectional areas. (a), (b) and (c): Lognormal distribution ($\sigma = 0.6$, $\mu = -\frac{\sigma^2}{2}$). (d), (e) and (f): Exponential distribution ($\lambda = 1$). (g), (h) and (i): Cells with volume 1.

5.1 Interpreting the weights as sphere radii

Let us now give another parameterization of the Laguerre diagram which is also frequently used, as this helps us with one possible way to interpret the weights.

Definition 4 (Alternative parameterization of a Laguerre diagram)**.** Given a convex domain $\Omega \subset \mathbb{R}^d$, n distinct generator points: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding radii: $r_1, \ldots, r_n \geq 0$. The Laguerre-Voronoi diagram $\{L_i\}_{i=1}^n$ generated by $(\boldsymbol{x}_1, r_1), \ldots, (\boldsymbol{x}_n, r_n)$ is defined by:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x} - \boldsymbol{x}_i||_{\text{per}}^2 - r_i^2 \leq ||\boldsymbol{x} - \boldsymbol{x}_j||_{\text{per}}^2 - r_j^2 \ \ \forall j \in \{1, \ldots, n\}\}.$$

We can easily see that this is indeed another parameterization of a Laguerre diagram. After all, when using Definition 1 we can take a suitably large constant and add this constant to all weights to make them non-negative because of property **L.2**. In this section we therefore transform the computed weights of the Laguerre diagrams to the radii as in Definition 4 via the following transformation:

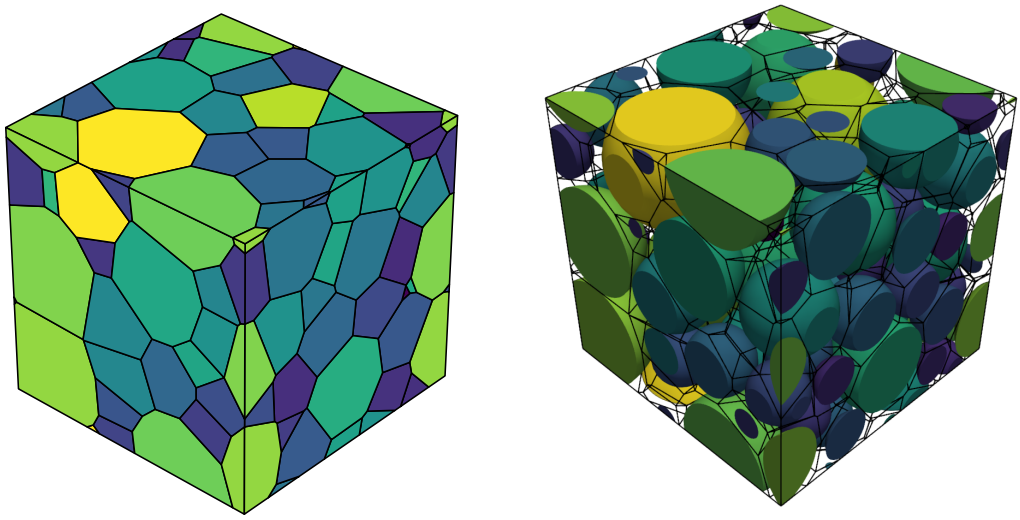
$$r_i(\alpha) = \sqrt{w_i - \left(\min_{j \in \{1,\ldots,n\}} w_j\right) + \alpha} \quad \text{for some fixed } \alpha \geq 0. \tag{5.1}$$

A reason to consider this alternative parameterization is that the r_i's may be considered as radii of spheres. We may generate a Laguerre diagram via a hard packing of spheres, this approach is for example taken in [7, 6, 27, 17]. Recall that a random close packing of spheres is obtained via an algorithm which efficiently positions non-overlapping spheres in some domain, filling as much space with spheres as possible. Suppose we have a Laguerre diagram $\{L_i\}_{i=1}^n$ as in Definition 4, which is generated by $(\boldsymbol{x}_1, r_1), \ldots, (\boldsymbol{x}_n, r_n)$. Consider balls of the form $B_i = \{\boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}_i - \boldsymbol{x}||_{\text{per}} < r_i\}$ with generator point $\boldsymbol{x}_i$ and radius r_i. In the case of a hard sphere packing we have that $B_i \cap B_j = \emptyset$ if $i \neq j$. What is interesting is that in case of a sphere packing, the ball B_i is contained within L_i for all $i \in \{1, \ldots, n\}$. This means that if the packing density of the sphere packing is high (the packing density is the volume fraction of the domain which is filled with spheres) then the volume of each sphere is a reasonable approximation of the corresponding cell volume.

A fairly natural question is to ask whether Algorithm 1 also gives rise to a hard sphere packing. Running the algorithm for many iterations seems to make cells more round, or at the very least makes cells look more like regular polyhedrons. We can verify this via a simulation experiment, we simply generate a Laguerre diagram using a lognormal distribution for the cell volumes and then we apply the transformation as in equation (5.1) with $\alpha = 0$. The reason for picking $\alpha = 0$ is that this yields the smallest possible radii that we can obtain via transformation (5.1), and therefore these are also the smallest radii that we can obtain directly via Algorithm 1. If the smallest possible radii we can obtain does not yield a sphere packing we also do not have a sphere packing for any $\alpha > 0$. We generate a Laguerre diagram as usual and then we verify whether we have a sphere packing by checking whether $||\boldsymbol{x}_i - \boldsymbol{x}_j||_{\text{per}} \geq r_i(0) + r_j(0)$ whenever $i \neq j$. Unfortunately, we never actually observed a sphere packing by following this procedure. In all cases there were overlapping spheres regardless of whether periodic boundary conditions were used or not. Trying to make the cells more round by using a smaller ε in Algorithm 1 also did not make a difference.

If one would want to use Algorithm 1 to produce a sphere packing this is still possible, but only indirectly. Finding the largest inscribed sphere in a cell (which is a polyhedron) is a linear

programming problem which is described in appendix B. We can therefore use Algorithm 1 to generate a sphere packing. An example of what the inscribed spheres inside cells of a Laguerre diagram look like is provided in Figure 5.2. In some sense the relationship between Laguerre diagrams and sphere packings always seems to be approximate. When generating a sphere packing with prescribed volumes for the spheres we may then generate a Laguerre diagram from this sphere packing using the centers and radii as generator points and r_i's respectively. Then, the volume of each sphere is an approximation of the volume of the corresponding cell, but it always underestimates the cell volume. When going in the other direction, we can generate a Laguerre diagram via Algorithm 1 with a prescribed volume distribution for the cells, and we obtain a sphere packing by inscribing a sphere in each cell. In that case each cell volume is an approximation of the volume of its inscribed sphere, but it always overestimates the sphere volume.



(a) (b)

Figure 5.2: (a): A Laguerre diagram with a lognormal cell volume distribution and periodic boundary conditions. (b): The largest inscribed (periodic) spheres in each cell of the Laguerre diagram as shown in (a).

We now investigate whether we find any relationship between the radius of the largest inscribed sphere within each cell and the corresponding r_i belonging to that cell. Let r_i^s be the radius of the largest sphere that is inscribed in cell L_i. As the r_i's of the Laguerre diagram are not unique we compare the r_i^s's to $r_i(\hat{\alpha})$ where $\hat{\alpha}$ is defined as:

$$\hat{\alpha} = \arg\min_{\alpha \geq 0} \frac{1}{n} \sum_{i=1}^{n} (r_i^s - r_i(\alpha))^2. \tag{5.2}$$

In words this means that we compare the radius of the inscribed spheres to the r_i's of the Laguerre diagram of the form (5.1), which minimize the mean squared error. We consider the same three volume distributions as in the previous section, and we compute: $r_i(\hat{\alpha})$ for $i \in \{1, \ldots, n\}$ which we call the fitted radii. The results are shown in Figure 5.3.

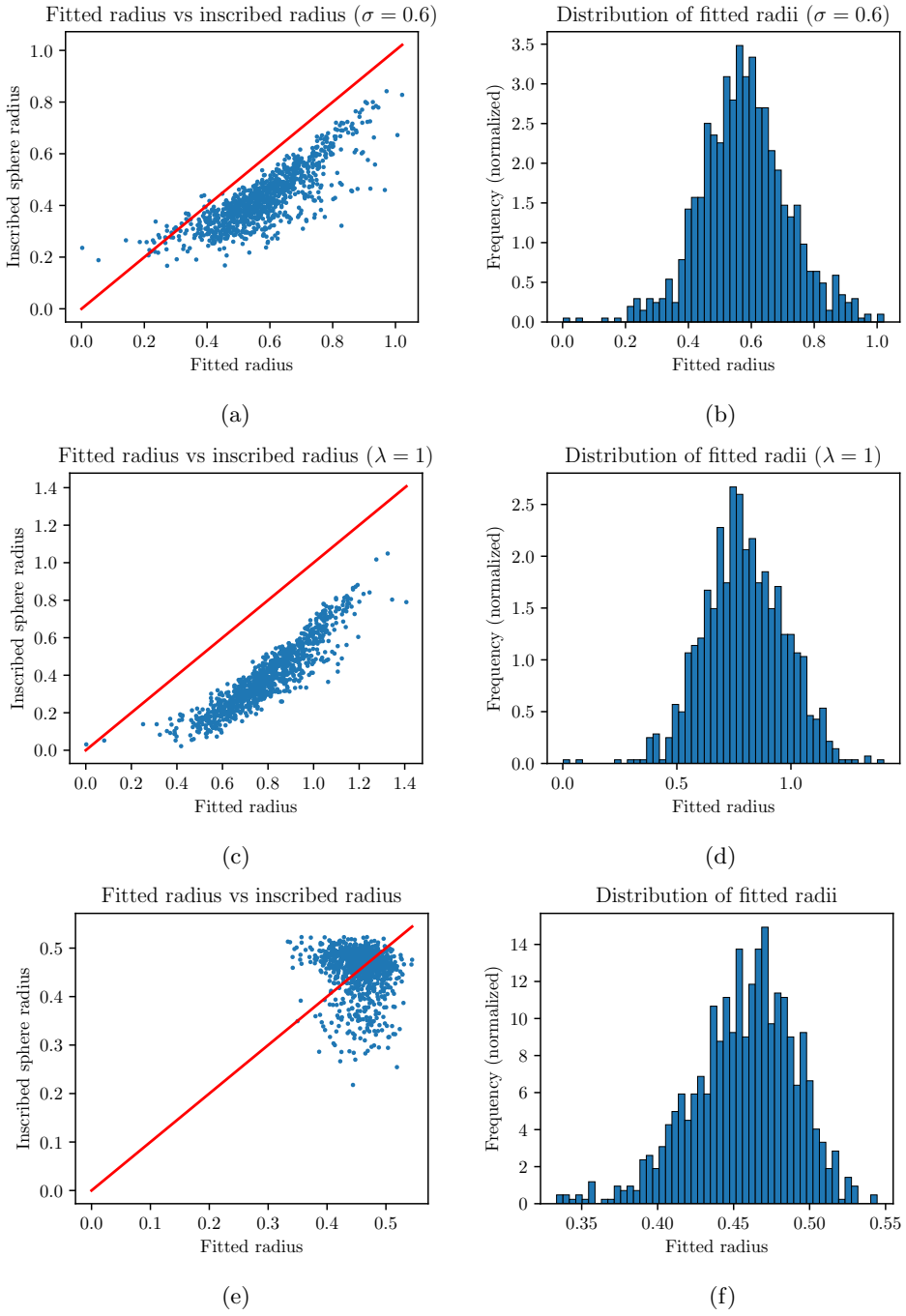


Figure 5.3: On the left: scatter plots of the fitted radii of the Laguerre diagram against the radii of the inscribed spheres, for Laguerre diagrams with various cell volume distributions. The red line indicates the line $y = x$. On the right: histograms which visualizes the distribution of the fitted radii. (a), (b): Lognormal distribution ($\sigma = 0.6$, $\mu = -\frac{\sigma^2}{2}$). (c), (d): Exponential distribution ($\lambda = 1$). (e), (f): All cells have volume 1.

As seen in Figure 5.3 it does not usually seem that reasonable to interpret the square root of the weights of a Laguerre diagram as the radius of its largest inscribed sphere. In the case of the lognormal distribution and the exponential distribution we observe that the radii of the largest inscribed spheres compare very poorly to the fitted radii. In those two cases we happened to observe that $\hat{\alpha} = 0$. A better fit could be possible if we allow $\alpha < 0$ but then we also need to allow $r_i(\alpha)^2 < 0$ and then we can no longer interpret the fitted radii as actual radii of spheres. In the case of a Laguerre diagram with cells of volume 1 we observe a better fit, but it still does not seem very convincing to interpret the square root of the weights as radii of the inscribed spheres. In some cases it does appear be reasonable to say that the relationship between the fitted radii and the radii of the inscribed spheres is approximately linear. In particular, in Figure 5.3 (a), (b) a straight line could be fitted to the data.

5.2 Conclusion

To summarize, using the simulations presented in this chapter we did not directly find a practical way to interpret the weights. These simulations could also simply be considered as some exploratory simulations. There were however some important observations which we need to keep in mind if we want to solve the direct problem and the inverse problem. Suppose we assume that the cell volumes of a Laguerre diagram are distributed according to a parametric family of distributions (e.g. the family of lognormal distributions), then we conclude that:

- The distribution of the weights does not necessarily belong to the same family of distributions.
- The distribution of the cross sectional areas does not necessarily belong to the same family of distributions.
- We may hypothesize that the distribution of the weights can be described quite well by a (three-parameter) lognormal distribution, regardless of the cell volume distribution. More simulations are necessary to show this, and there may be parametric distributions which provide an even better fit.
- The square root of the weights of a Laguerre diagram cannot in general be interpreted as the radius of its largest inscribed sphere.

In the case of generating a Laguerre diagram via a close packing of spheres there is a clear interpretation of the weights being the square root of the sphere radii. For the Laguerre diagrams we consider this interpretation cannot be used and does not add knowledge for our problem. As such, we conclude that for our problem we do not have to worry about the interpretation or the distribution of the weights.

6

2D sectional Laguerre diagrams

After having thoroughly presented the 3D counterpart of the problem, in this section the 2D sectional Laguerre diagrams are presented. As previously mentioned, a 2D section of a 3D Laguerre diagram is also a Laguerre diagram. In this chapter we investigate whether this sectional Laguerre diagram has properties that may help solve the inverse problem.

Let us first consider the following problem, given a 3D Laguerre diagram and a cross section plane we would like to know the weights and generator points of the resulting 2D diagram. As it turns out the generator points can be found by orthogonally projecting the points of the 3D diagram onto the cross section plane (for cells that intersect with the cross section plane). The resulting 2D diagram also has different weights which are related to the weights of the 3D Laguerre diagram. Lemma 1 provides a precise description of the 2D diagram. We could not find this particular result in the literature, but in [9, p. 131, 132] it is shown in a similar manner that the cross section of a Voronoi diagram is a Laguerre diagram.

Lemma 1. [1] *Given a convex domain $\Omega \subset \mathbb{R}^3$, n distinct generator points: $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \dots, w_n \in \mathbb{R}$. Let $\{L_i\}_{i=1}^n$ be the Laguerre diagram generated by $\{(\boldsymbol{x}_i, w_i)\}_{i=1}^n$ with domain Ω. Let $\boldsymbol{a}, \boldsymbol{x}_0 \in \mathbb{R}^3$, then we consider the cross section plane:*

$$P := \{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{a}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0\}.$$

Let I be the index set of cells which appear in the cross section plane P:

$$I := \{i \in \{1, \dots, n\} : P \cap L_i \neq \emptyset\}.$$

For $i \in I$ define:

$$\boldsymbol{x}_i^P := \boldsymbol{x}_i - \frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a} \tag{6.1}$$

$$w_i^P := w_i - \frac{\left(\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)\right)^2}{\boldsymbol{a}^T\boldsymbol{a}}. \tag{6.2}$$

Let $\{L_i^P\}_{i\in I}$ be the Laguerre diagram generated by $\left\{(\boldsymbol{x}_i^P, w_i^P)\right\}_{i\in I}$ with domain $\Omega \cap P$. The following holds:

$$L_i \cap P = L_i^P \quad \text{for all } i \in I.$$

[1] Laguerre diagrams in this lemma are non-periodic.

Proof. The proof can be found in appendix C. □

A technical remark is that Lemma 1 only holds for non-periodic Laguerre diagrams in its current form. As the weight w_i^P in some sense depends on the distance of the generator $\boldsymbol{x}_i$ to the plane P we have to keep in mind that the periodic distance does not always coincide with the Euclidean distance. Throughout this chapter we assume that we have a non-periodic Laguerre diagram $\{L_i\}_{i=1}^n$ which we intersect with a cross section plane P and we use the same notation as in Lemma 1. If we consider our inverse problem from the point of view of materials science, then we have some material of shape Ω which we cut with a known cross section plane P. Therefore let us assume that we know $\boldsymbol{a}$ and $\boldsymbol{x}_0$. If we also know $\boldsymbol{x}_i^P$, w_i^P for $i \in I$ then we could attempt to solve for $\boldsymbol{x}_i$ and w_i in equations (6.1) and (6.2). Unfortunately (6.1) cannot be solved for $\boldsymbol{x}_i$ since $\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i + t_i\boldsymbol{a}$ solves the equation for every $t_i \in \mathbb{R}$. As we only have one more equation, namely (6.2) we cannot solve for both t_i and w_i.

This means that if we know the 2D diagram which describes the cross section, then we only know on which line each of the original generator points of the 3D diagram is located. Another problem is that it does not seem reasonable to assume that the $\boldsymbol{x}_i^P$'s and w_i^P's are known. We assume the 3D Laguerre diagram to be opaque, hence intersecting it with a plane will only reveal the resulting cross section which shows us the polygonal shapes of the intersected cells. Now we may wonder whether we can determine the generator points and weights, $\boldsymbol{x}_i^P$, w_i^P for $i \in I$ from the observed polygons. This is in fact another inverse problem, given a tessellation of a 2D plane of polygons we want to find the generator points and weights of a Laguerre diagram which describes this tessellation. The direct problem would be to determine the 2D Laguerre diagram (i.e. the shapes of the polygonal cells) for given weights and generator points. This inverse problem is addressed in [28] for 2D Laguerre diagrams. They refer to this problem as the inversion of a Laguerre diagram. The authors show that this problem does not have a unique solution, it turns out that completely different sets of weights and generator points may

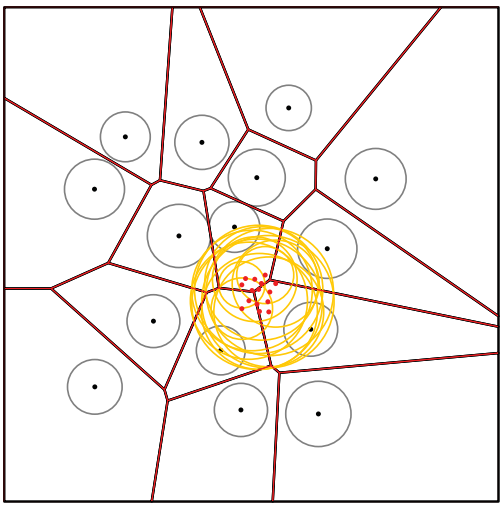
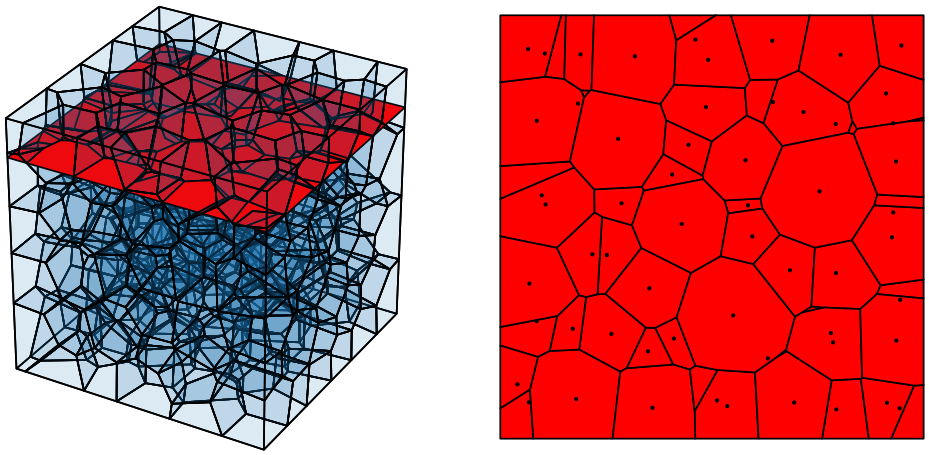


Figure 6.1: An example of a non-periodic 2D Laguerre diagram which can be represented by two different sets of generator points and weights. The two sets of generator points are the black dots and the red dots. The circles visualize the weights, the radii being the r_i's as in Definition 4. Image taken from [28].

still yield the same Laguerre diagram. One of the examples given in their paper is shown in Figure 6.1. In the same paper an algorithm is proposed which determines weights and generator points, providing as input the locations of vertices and edges for each polygon. As the solution is not unique the algorithm can be adapted to compute solutions satisfying some condition, we may for example require the solution to be a centroidal Laguerre diagram. Given that we consider (approximately) centroidal 3D Laguerre diagrams we could be tempted to think that the resulting sectional Laguerre diagram which we compute via Lemma 1 is also centroidal. In Figure 6.2 we have performed a simulation to assess this. We observe that the 2D diagram does seem approximately centroidal, but there are however various cells which do not contain their own generator point.



(a) (b)

Figure 6.2: (a): A 3D Laguerre diagram with a cross section plane. (b): A 2D Laguerre diagram which is equal to the cross section as in (a), computed via Lemma 1.

As a result of the non-uniqueness of the generator points and weights we can only expect to find estimates of $\boldsymbol{x}_i^P$ and w_i^P for $i \in I$, there is no direct approach to recovering the true values. This means that there is no obvious method to find a reasonable estimate of $\boldsymbol{x}_i$ and w_i for $i \in I$. We should also not forget that for cells L_i with $i \in \{1, \ldots, n\}$ but with $i \notin I$ we have no information at all. Suppose that we find a reasonable estimate of the 2D Laguerre diagram as described in Lemma 1. Given the information we have at present it seems that directly estimating the generator points and weights of all the cells using this estimated diagram is out of reach. We conclude this chapter by noting that the sectional Laguerre diagrams do not seem to have properties which may help in solving the direct and inverse problem.

7

Linking the volume distribution to the sectional area distribution

In this chapter a model is proposed which links the cell volume distribution of a regularized Laguerre diagram as generated by the procedure described in section 4.4.1 to its distribution of cross sectional cell areas which are computed as described in section 4.4.2. In the derivation of the model a scaling argument is used and an assumption about the shape of the cell is needed. Before stating this assumption we need to introduce the notion of similarity.

Definition 5 (similarity in $\mathbb{R}^3$)**.** [29, p. 177]. Let $A, B \subset \mathbb{R}^3$ be non-empty. We call A λ-*similar* to B if there exists a bijection $f : A \to B$ and a $\lambda > 0$ such that:

$$||f(\boldsymbol{x}) - f(\boldsymbol{y})|| = \lambda ||\boldsymbol{x} - \boldsymbol{y}||,$$

for all $\boldsymbol{x}, \boldsymbol{y} \in A$. The function f is also referred to as a *similarity*. λ is known as the *ratio of similitude*.

Let us provide some more intuition for this definition. Suppose we have some set $A \subset \mathbb{R}^3$ which is λ-similar to the set $B \subset \mathbb{R}^3$. This essentially means that A and B have the exact same shape. By uniformly scaling B, or more precisely, by scaling the distances between the points in B with λ, the resulting set will not only have the same shape but also the same size as A.

A similarity with $\lambda = 1$ is usually called an isometry instead of a similarity. It is well-known that translations, rotations, reflections and combinations thereof are isometries. It can be seen that the composition of similarities is again a similarity. Hence, operations that can (consecutively) be performed on a set while preserving similarity are: scaling, translations, rotations and reflections (these are also known as similarity transformations). A similarity is also an affine transformation. For more details on isometries and similarities in Euclidean space we refer to [30, p. 96-104].

The concept of similarity formalizes the following well-known facts:

- For any area, scaling the distance between all points with a factor λ scales the area by a factor λ^2.

- For any volume, scaling the distance between all points with a factor λ scales the volume by a factor λ^3.

This also means that given $A, B \subset \mathbb{R}^3$, with A λ-similar to B, then their volumes are related as follows: $|A| = \lambda^3 |B|$. Let us now introduce how similarity is used in the derivation of a model.

Suppose we have a regularized Laguerre diagram, generated as per the procedure described in section 4.4.1. Let us now consider a single cell $L^{(1)}$ which has volume 1: $|L^{(1)}| = 1$. Assume there exists another cell L_i with volume $|L_i|$ which is similar to $L^{(1)}$. This means that L_i is λ-similar to $L^{(1)}$, with $\lambda = \sqrt[3]{|L_i|}$. If we now take a cross section of cell $L^{(1)}$ and consider its area indicated by $A^{(1)}$ and if we take the same cross section of L_i with area A_i then we find that:

$$A_i = \lambda^2 A^{(1)} = \sqrt[3]{|L_i|}^2 A^{(1)} = |L_i|^{\frac{2}{3}} A^{(1)}. \tag{7.1}$$

While the above derivation holds when cells have the exact same shape, if we assume that all cells approximately have the same shape we hope this is still a reasonable starting point for a model. Hence, we now state the assumption that is used in our model:

Assumption 1 (shape of cells)**.** *Given a Laguerre diagram $\{L_i\}_{i=1}^n$ which is generated as per the procedure described in section 4.4.1. Assume that there exists a convex set $L^{(1)} \subset \mathbb{R}^3$ of volume 1 such that for all $i \in \{1, \ldots, n\}$ the cell L_i is approximately similar to $L^{(1)}$.*

Let us now immediately state the description of the model, which we discuss later.

Model 1. *Consider a regularized Laguerre diagram as generated by the procedure described in section 4.4.1.*

- *Let F_V, f_V be the cdf and pdf respectively, of the cell volume distribution, and let $V \sim F_V$.*
- *Let F_A, f_A be the cdf and pdf respectively, of the cross sectional area distribution, and let $A \sim F_A$.*
- *Let $F_{A^{(1)}}, f_{A^{(1)}}$ be the cdf and pdf respectively, of the cross sectional area distribution of a Laguerre diagram which only has cells of volume 1. Let $A^{(1)} \sim F_{A^{(1)}}$.*
- *Let $F_{V|I}, f_{V|I}$ be the cdf and pdf respectively, of the volume distribution of the cells that appear in cross sections. Specifically, we define $f_{V|I}$ in terms of f_V:*

$$f_{V|I}(v) = \frac{v^{\frac{1}{3}} f_V(v)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)}. \tag{7.2}$$

  Let $V_I \sim F_{V|I}$.

Then, we assume the relationship between the aforementioned random variables is as follows:

$$A \overset{d}{=} V_I^{\frac{2}{3}} A^{(1)}. \tag{7.3}$$

Where $\overset{d}{=}$ denotes equality in distribution, and where V_I and $A^{(1)}$ are independent.

This model description likely raises the following questions:

1. Why do we need the distribution $F_{V|I}$, and what is the justification of its definition? Given our earlier derivation of (7.1) one would naively think that we could take the following relationship instead of (7.3):
$$A \overset{d}{=} V^{\frac{2}{3}} A^{(1)}.$$
The reason we do not do this is that there is still something to account for. One can imagine that when taking a cross section of a Laguerre diagram that larger cells are more likely to appear in cross sections. The distribution $F_{V|I}$ represents the volume distribution of the cells that appear in cross sections, its definition (7.2) is motivated in the next section.

2. Why are V_I and $A^{(1)}$ assumed to be independent? Throughout this chapter we simply assume this. In chapter 8 some related stereological problems are discussed, namely the famous Wicksell's corpuscle problem and a generalization thereof. We then see that for these problems we can analogously define the random variables V_I and $A^{(1)}$ and show that they are independent in those settings.

3. What is the interpretation of the distribution $F_{A^{(1)}}$? One way to interpret this distribution function is the cross sectional area distribution function of the convex set $L^{(1)}$ which is mentioned in assumption 1. We could consider $L^{(1)}$ to be the average cell of a Laguerre diagram of volume 1.

Let us provide an overview of the structure of the remaining sections in this chapter. In the next section an estimator for f_A is derived and in this derivation we further address question 1 as stated above. After its derivation we take a look at the distribution $F_{A^{(1)}}$ in order to provide a more elaborate answer to question 3. Finally, some useful properties of the proposed estimator for f_A are discussed. Question 2 is addressed in chapter 8.

7.1 Deriving an estimator for the sectional area distribution

In this section model 1 is used to derive an estimator for the sectional area distribution. In model 1 the area distribution is described as the product of the two random variables $A^{(1)}$ and $V_I^{\frac{2}{3}}$ (with pdf's $f_{A^{(1)}}$ and $f_{V^{\frac{2}{3}}|I}$ respectively) which are assumed to be independent. Note that in general, for independent and non-negative random variables X and Y the probability density function of $Z = XY$ is given by [31]:
$$f_Z(z) = \int_0^\infty f_X(x) f_Y\left(\frac{z}{x}\right) \frac{1}{x} \mathrm{d}x. \tag{7.4}$$
Where f_X, f_Y, f_Z denote the probability density functions of X, Y and Z respectively. Let us for now assume that the random variable $A^{(1)}$ is bounded, therefore suppose that $f_{A^{(1)}}$ is supported on $(0, a_{\max}]$ for some $a_{\max} > 0$. Using the aforementioned formula we find the following expression for f_A:
$$f_A(a) = \int_0^{a_{\max}} f_{A^{(1)}}(x) f_{V^{\frac{2}{3}}|I}\left(\frac{a}{x}\right) \frac{1}{x} \mathrm{d}x. \tag{7.5}$$
The next step is to find an expression for $f_{V^{\frac{2}{3}}|I}$ and for this purpose we take some inspiration from Wicksell's corpuscle problem. Recall that Wicksell's problem was introduced in section 3.1. Therefore, assume we have a box that is filled with spheres which follow some volume distribution.

If we were to take a cross section of this box, we would observe that the volume distribution of the spheres that appear in the cross section is not the same as the sphere volume distribution. This is caused by the fact that it is more likely that the cross section plane passes through a large sphere than through a small sphere. Let us introduce some notation. Assume that we have a domain $\Omega = [0, l]^3$ for some $l > 0$ which contains non-overlapping spheres $B_1, \ldots, B_n$. Let $V_i := |B_i|$ denote the volume of sphere B_i, then we assume that $V_1, \ldots, V_n \overset{\text{iid}}{\sim} F_V$. Let I be the index set of spheres that appear in the random cross section plane P:

$$I = \{i \in \{1, \ldots, n\} : P \cap B_i \neq \emptyset\}.$$

We now show that the distributions of V_i and $V_i | i \in I$ are related, in terms of their respective densities f_V and $f_{V|I}$.

Assume that a sphere of radius R is placed at random inside Ω. If we perform a horizontal cross section (at a random height) on this box, the probability of observing the sphere is $\frac{2R}{l}$. This is simply the fraction of the height of box which is covered by the sphere, which is equal to the diameter of the sphere divided by the height of the box. If the volume of a sphere is given by V_i then its radius is given by:

$$V_i = \frac{4}{3}\pi R^3 \iff R = \left(\frac{3V_i}{4\pi}\right)^{\frac{1}{3}},$$

and the probability of observing the sphere is therefore given by $\left(\frac{3V_i}{4\pi}\right)^{\frac{1}{3}} \frac{2}{l}$. We may therefore state that:

$$\mathbb{P}\left(i \in I | V_i = v\right) = \left(\frac{3v}{4\pi}\right)^{\frac{1}{3}} \frac{2}{l}.$$

By applying Bayes' rule we find the following relationship:

$$f_{V|I}(v) = \frac{\mathbb{P}(i \in I | V_i = v) f_V(v)}{\int \mathbb{P}(i \in I | V_i = x) f_V(x) \mathrm{d}x} = \frac{\left(\frac{3v}{4\pi}\right)^{\frac{1}{3}} \frac{2}{l} f_V(v)}{\int \left(\frac{3x}{4\pi}\right)^{\frac{1}{3}} \frac{2}{l} f_V(x) \mathrm{d}x} = \frac{v^{\frac{1}{3}} f_V(v)}{\int x^{\frac{1}{3}} f_V(x) \mathrm{d}x} = \frac{v^{\frac{1}{3}} f_V(v)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)}. \quad (7.6)$$

Note that:

$$\mathbb{P}(i \in I) = \int \mathbb{P}(i \in I | V_i = x) f_V(x) \mathrm{d}x.$$

Recall that if X and Y are random variables with corresponding density functions f_X and f_Y, and if $Y = g(X)$ where g is a monotone function we may apply the well-known formula:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|. \quad (7.7)$$

Now we use (7.7) to compute the density function of $V_i^{\frac{2}{3}} | i \in I$ which we denote by $f_{V^{\frac{2}{3}}|I}$:

$$f_{V^{\frac{2}{3}}|I}(v) = f_{V|I}\left(v^{\frac{3}{2}}\right) \frac{3}{2}\sqrt{v} \overset{(7.6)}{=} \frac{\sqrt{v} f_V\left(v^{\frac{3}{2}}\right)}{\mathbb{E}(V^{\frac{1}{3}})} \frac{3}{2}\sqrt{v} = \frac{3}{2} \frac{v f_V\left(v^{\frac{3}{2}}\right)}{\mathbb{E}(V^{\frac{1}{3}})}. \quad (7.8)$$

Now we simply plug (7.8) into (7.5) and we obtain:

$$\begin{aligned} f_A(a) &= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} f_{A^{(1)}}(x) f_V\left(\left(\frac{a}{x}\right)^{\frac{3}{2}}\right) \frac{1}{x^2} \mathrm{d}x \quad (7.9) \\ &= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} f_V\left(\left(\frac{a}{x}\right)^{\frac{3}{2}}\right) \frac{1}{x^2} \mathrm{d}F_{A^{(1)}}(x). \end{aligned}$$

The drawback of the approach is that we assume $f_{A^{(1)}}$ and $F_{A^{(1)}}$ to be known while we do not have an expression for this distribution. Thanks to simulation, an accurate approximation can be provided. Let $a_1^{(1)}, \ldots, a_N^{(1)} \overset{\text{iid}}{\sim} F_{A^{(1)}}$ be a sample and let F^* denote its Empirical Cumulative Distribution Function (ECDF). Because F^* approximates $F_{A^{(1)}}$ we now define our estimator for the cross sectional area distribution as:

$$\begin{aligned}\hat{f}_A(a) &:= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} f_V\left(\left(\frac{a}{x}\right)^{\frac{3}{2}}\right) \frac{1}{x^2} \mathrm{d}F^*(x) \\ &= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} f_V\left(\left(\frac{a}{a_i^{(1)}}\right)^{\frac{3}{2}}\right) \frac{1}{\left(a_i^{(1)}\right)^2}. \end{aligned} \tag{7.10}$$

For some parametric distributions it turns out that $f_{V^{\frac{2}{3}}|I}(v)$ is easy to compute. It is important to state that the following expression for the estimator is equivalent, which is a direct consequence of the definition of $f_{V^{\frac{2}{3}}|I}$ as in (7.8):

$$\hat{f}_A(a) = \frac{1}{N} \sum_{i=1}^{N} f_{V^{\frac{2}{3}}|I}\left(\frac{a}{a_i^{(1)}}\right) \frac{1}{a_i^{(1)}}. \tag{7.11}$$

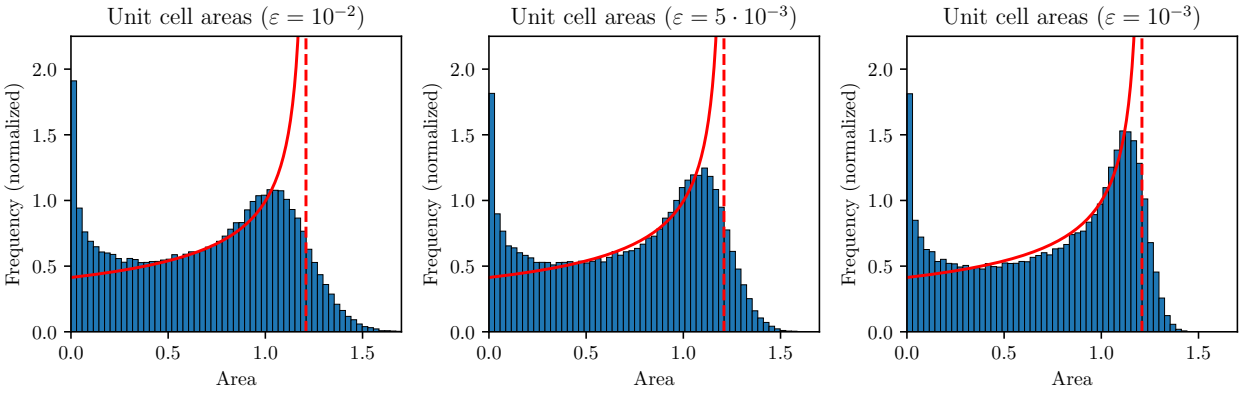
Because the sample $a_1^{(1)}, \ldots, a_N^{(1)}$ has such an important role in the estimator we briefly take a look at the distribution $F_{A^{(1)}}$ in the next section.

7.2 The sectional area distribution of unit cells

In this section the distribution of the sectional areas of cells of volume 1 (unit cells) is visualized and its interpretation is discussed. Recall that this distribution is denoted by $F_{A^{(1)}}$, and that we have the tools to take samples from this distribution. The reason to study this distribution is that it plays an important role in the estimator (7.10) derived in the previous section. We also consider some related distributions which help understand the shape of the distribution $F_{A^{(1)}}$. In particular, we consider distributions of cross sectional areas, obtained by performing a random cross section of a known geometrical shape.

As it turns out the distribution $F_{A^{(1)}}$ depends on the parameter ε as in Algorithm 1. Recall that we stated that we fixed $\varepsilon = 0.001$ for all simulations. For the sake of completeness the dependence of $F_A^{(1)}$ on ε is shown, to highlight that a different choice impacts the results. As before, using the procedures described in sections 4.4.1 and 4.4.2, the Laguerre diagrams and the cross section area distribution respectively are generated. We compute a Laguerre diagram with $n = 1000$ cells and then $K = 1000$ horizontal cross sections are taken (more precisely, we compute 1000 Laguerre diagrams and take a single cross section for each diagram). Because the argument revolves around the cells having volume 1 we take a tighter tolerance than usual: $\varepsilon_f = 10^{-10}$. This entire process is then repeated for different values of the tolerance ε, which controls the number of Lloyd centering steps of Algorithm 1 and therefore the regularity of the cells. The sample of cross sectional areas we obtain is visualized with a histogram in Figure 7.1. In this figure we also make a comparison with the sectional area distribution of a sphere of volume 1 (red line). In section 8.1.1 the sectional area distribution for spheres is derived analytically.

We can clearly see in these results, that for smaller ε the right tail of the cross section area distribution moves to the left. This can be explained because less regular cells (larger



(a) (b) (c)

Figure 7.1: Histograms of the simulated cross section area distribution of a Laguerre diagram with cells of volume 1 (unit cells) with different tolerances ε for Algorithm 1. The red line indicates the pdf of the sectional area distribution of a sphere of volume 1 along with its vertical asymptote which is indicated by the dashed line.

ε) may occasionally be "elongated", which may result in observing some slightly larger cross section areas. Meanwhile, more regular cells (smaller ε) results in cells which become a sort of approximation of a regular polyhedron and as a consequence we tend to observe slightly smaller cross section areas by comparison. It can also be seen that for a smaller value of ε the area distribution starts to look more similar to the sectional area distribution of a sphere, which again hints at the cells becoming more regular or rounder for smaller ε.

Remark. *As $\varepsilon \downarrow 0$ Algorithm 1 will produce a truly centroidal Laguerre diagram, hence we also expect the cross sectional area to converge to some distribution as $\varepsilon \downarrow 0$.*

Now the first five moments of the distribution $F_{A^{(1)}}$ are computed, these are needed when applying the method of moments to our estimator of the sectional area distribution. We consider the sample of $F_{A^{(1)}}$ as can be seen in Figure 7.1 (c). This sample contains $N = 137300$ areas. The estimates of the first five moments are in Table 7.1.

Moment	estimate
1	0.728332
2	0.694634
3	0.718036
4	0.773846
5	0.856534

Table 7.1: Estimates of the first five moments of the distribution $F_{A^{(1)}}$. These estimates are obtained by computing $\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^j$ for $j \in \{1, 2, \ldots, 5\}$.

Remark. *One way we could interpret $F_{A^{(1)}}$ is as the cross sectional area distribution of the average cell which has volume 1.*

Because cells of a Laguerre diagram are convex polyhedra it is interesting to look at the cross sectional area distribution of some known convex polyhedra. Previous studies [33, 32]

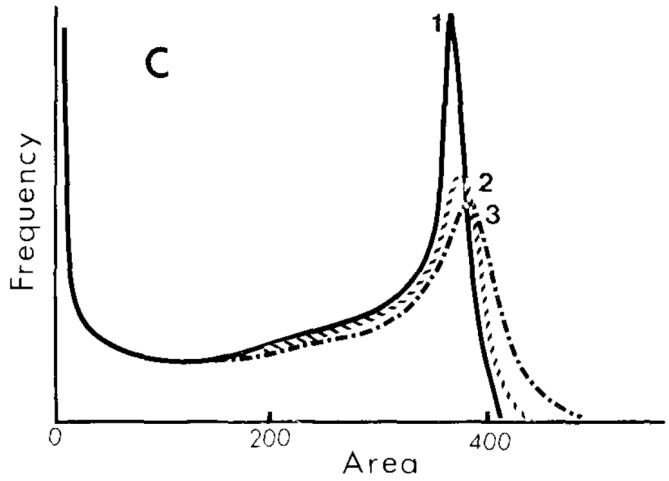


Figure 7.2: Cross sectional area distribution of a regular convex dodecahedron (1); lines (2) and (3) represent perturbed dodecahedra. Each of the twelve sides of a regular convex dodecahedron is a plane, namely a pentagon. The sides of the perturbed dodecahedra were perturbed by perturbing the angle of the normal vectors of these planes by some angle. For (2) the standard deviation for the perturbation was chosen to be 10 degrees, for (3) it was chosen to be 30 degrees. These results and this figure is taken from [32].

investigated the sectional area distribution of some (regular) convex polyhedra, and the shape of some of these distributions appear similar to the distribution $F_{A^{(1)}}$. In [33] the sectional area distribution of various convex regular polyhedra is considered such as the cube, the dodecahedron and the tetrakaidecahedron. In this study a particularly creative approach was taken, all results were obtained by measuring the sectional areas of metal wireframe models of the aforementioned shapes. In this paper there are various visualizations of sectional area distributions which are similar in appearance to $F_{A^{(1)}}$. In [32] results were instead obtained via computer simulations. Among the distributions of convex polyhedra investigated in this paper the one that resembles best $F_{A^{(1)}}$ are the dodecahedron and the truncated octahedron. The sectional area distribution of the dodecahedron is shown in Figure 7.2.

We have observed that the sectional area distribution of some convex regular polyhedra such as the regular convex dodecahedron appear very similar to $F_{A^{(1)}}$. Moreover, we have seen that for smaller values of ε (as in Algorithm 1) the cells do indeed become more round since the area distribution starts to look more similar to the sectional area distribution of a sphere.

7.3 Properties of the estimator for the sectional area distribution

Now that an estimator for the sectional area distribution $\hat{f}_A$ (7.10) is derived, its properties are discussed. For now, let us simply assume that the estimator $\hat{f}_A$ provides an accurate description of the sectional area distribution. It is important to study as the estimator $\hat{f}_A$ links the sectional area distribution to the volume distribution.

Assume that the cell volume distribution follows some known parametric distribution with parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ and density $f_V(x|\boldsymbol{\theta})$. In that case $\hat{f}_A$ is given by:

$$\hat{f}_A(a|\boldsymbol{\theta}) = \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} f_V\left(\left(\frac{a}{a_i^{(1)}} \right)^{\frac{3}{2}} \middle| \boldsymbol{\theta} \right) \frac{1}{\left(a_i^{(1)}\right)^2}. \tag{7.12}$$

Also note that:

$$\mathbb{E}(V^{\frac{1}{3}}) = \int_0^\infty x^{\frac{1}{3}} f_V(x|\boldsymbol{\theta}) \mathrm{d}x.$$

Hence, given a sample of observed areas we may employ methods such as the method of moments or maximum likelihood estimation to estimate $\boldsymbol{\theta}$. These methods allow directly estimating the cell volume distribution.

Let us first consider the approach via the method of moments. Suppose that $\hat{A} \sim \hat{f}_A$ and let $k \in \mathbb{N}$, the k-th moment of $\tilde{A}$ is given by:

$$\begin{aligned}
\mathbb{E}\left(\hat{A}^k\right) &= \int_0^\infty x^k \hat{f}_A(x) \mathrm{d}x \\
&= \int_0^\infty x^k \frac{3x}{2\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} f_V\left(\left(\frac{x}{a_i^{(1)}} \right)^{\frac{3}{2}} \right) \frac{1}{\left(a_i^{(1)}\right)^2} \mathrm{d}x \\
&= \frac{3}{2\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} \int_0^\infty x^{k+1} f_V\left(\left(\frac{x}{a_i^{(1)}} \right)^{\frac{3}{2}} \right) \frac{1}{\left(a_i^{(1)}\right)^2} \mathrm{d}x \\
&= \frac{1}{\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} \int_0^\infty \left(u^{\frac{2}{3}} a_i^{(1)}\right)^{k+\frac{1}{2}} f_V(u) \frac{1}{\sqrt{a_i^{(1)}}} \mathrm{d}u \qquad \text{substitute } u = \left(\frac{x}{a_i^{(1)}} \right)^{\frac{3}{2}} \\
&= \frac{1}{\mathbb{E}(V^{\frac{1}{3}})} \frac{1}{N} \sum_{i=1}^{N} \left(a_i^{(1)}\right)^k \int_0^\infty u^{\frac{2}{3}k+\frac{1}{3}} f_V(u) \mathrm{d}u \\
&= \frac{\mathbb{E}\left(V^{\frac{2}{3}k+\frac{1}{3}}\right)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)} \frac{1}{N} \sum_{i=1}^{N} \left(a_i^{(1)}\right)^k.
\end{aligned}$$

A downside of the estimator (7.10) is that it depends on data, without access to a sample of $F_{A^{(1)}}$ one cannot use it. However, if the parameters of the underlying volume distribution are estimated via the method of moments we only need to know some moments of $F_{A^{(1)}}$ instead of a sample. In this case we can simply rely on Table 7.1 with estimates of the first five moments which can easily be used in applications. Let $a_1, \ldots, a_m$ be a sample of areas, when using the method of moments we deal with equations of the form:

$$\frac{1}{m} \sum_{j=1}^{m} a_j^k = \mathbb{E}\left(\hat{A}^k\right) = \frac{\mathbb{E}\left(V^{\frac{2}{3}k+\frac{1}{3}}\right)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)} \frac{1}{N} \sum_{i=1}^{N} \left(a_i^{(1)}\right)^k \qquad \text{for } k = 1, 2, \ldots. \tag{7.13}$$

Once again, if the cell volume distribution is a parametric distribution with parameter $\boldsymbol{\theta}$ and

density $f_V(x|\boldsymbol{\theta})$, then we may define:

$$g_k(\boldsymbol{\theta}) := \frac{\mathbb{E}\left(V^{\frac{2}{3}k+\frac{1}{3}}\right)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)}.$$

Note that when V is lognormally distributed there is a simple closed-form expression for arbitrary moments of V as in (2.2), hence g_k is easy to compute. When V is distributed as a (generalized) gamma distribution then arbitrary moments can be expressed in terms of the gamma function as in (2.4). By the method of moments we may attempt to solve for $\boldsymbol{\theta}$ in the following (possibly nonlinear) system of equations:

$$\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^k}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^k} = g_k(\boldsymbol{\theta}) \quad \text{for } k \in \{1,\dots,p\}. \tag{7.14}$$

If it is not possible solve for $\boldsymbol{\theta}$ analytically we may resort to numerical methods. As for maximum likelihood estimation it is likely necessary to resort to use numerical methods for maximizing the (log)-likelihood. The method of moments estimate is a good option for the initial guess to the maximization algorithm that is used.

7.4 Conclusion

In this chapter we derived an estimator for the pdf of the sectional area distribution. The estimator provides a clear relationship between the distribution of sectional areas and the distribution of the cell volumes. A downside of the estimator is that it depends on data, namely on a sample of the distribution $F_{A^{(1)}}$. When assuming a specific parametric distribution for the cell volume distribution we showed how to use the estimator for estimating the parameters of the cell volume distribution. In the case of estimating these parameters via the method of moments we note that a sample of $F_{A^{(1)}}$ is no longer needed, it suffices to know estimates of the first few moments of this distribution which are provided in Table 7.1. In the derivation of the estimator some assumptions have been made which still require justification, this is addressed in the next chapter.

8

Applying the estimator to related stereological problems

In the previous chapter an estimator for the sectional area distribution was derived, under some assumptions. In this chapter we consider two related stereological problems. In these settings we show that the estimator can be defined analogously and we show that these aforementioned assumptions hold. We begin with studying its relation to Wicksell's corpuscle problem: in this case the pdf of the sectional area distribution can be derived analytically. We show that when applied to Wicksell's corpuscle problem the estimator of the sectional area distribution corresponds to the sectional area pdf. Then, a generalization of Wicksell's corpuscle problem is discussed, which considers similar convex particles instead of spheres. In this case there is still an analytical expression for pdf of the sectional area distribution. We also show that our estimator for the sectional distribution corresponds to this sectional area pdf. The stereological problem of similar convex particles also provides an explanation for assumption 1.

Recall that the estimator of for the sectional area distribution (7.10) depends on a sample of $F_{A^{(1)}}$. Throughout this chapter we consider the estimator in the form (7.9) where we assume $f_{A^{(1)}}$ to be known. This means that f_A is given by:

$$f_A(a) = \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} f_{A^{(1)}}(x) f_V\left(\left(\frac{a}{x}\right)^{\frac{3}{2}}\right) \frac{1}{x^2} \mathrm{d}x.$$

This is equivalent to:

$$f_A(a) = \int_0^{a_{\max}} f_{A^{(1)}}(x) f_{V^{\frac{2}{3}}|I}\left(\frac{a}{x}\right) \frac{1}{x} \mathrm{d}x,$$

with $f_{V^{\frac{2}{3}}|I}$ defined as in (7.8). As a result, if $A \sim f_A$ then A can be decomposed as the product of two independent random variables. So, let A be a random variable distributed as the sectional area distribution in Wicksell's corpuscle problem (or in the generalization we consider). If we want to show that A can be decomposed as the product of the two independent random variables as in model 1 it is sufficient to show that f_A is indeed the density of A.

8.1 Wicksell's corpuscle problem

It is interesting to compare cutting a Laguerre diagram with a plane to cutting a box filled with spheres with a plane, also known as Wicksell's corpuscle problem [15]. Recall that this problem was introduced in section 3.1. The main purpose of this section is to show that the sectional area distribution in this problem setting corresponds with our estimator (7.9). In the derivation of our estimator very little use is made of properties of Laguerre diagrams, this section also highlights that the estimator may be applied to shapes other than cells of a Laguerre diagram such as spheres.

If we consider Wicksell's original statement of the problem, as stated in section 3.1 we see that Wicksell considers the problem in terms of the diameters of spheres. We consider it in terms of the radii of the spheres. Let f be the probability density function associated with the radius of the 3D spheres. Let ϕ be the probability density function of the radii of the apparent 2D circles (the circles that are observed in the plane section). Recall equation (3.1):

$$\phi(x) = \frac{x}{\mathbb{E}(R)} \int_x^\infty \frac{f(r)}{\sqrt{r^2 - x^2}} \mathrm{d}r.$$

Here, $\mathbb{E}(R)$ is the average radius of the 3D spheres:

$$\mathbb{E}(R) = \int_0^\infty r f(r) \mathrm{d}r.$$

Let us now express (3.1) in terms of the observed areas instead of the observed radii. As the apparent radius r of a circle has density ϕ, then the observed area is given by $A = \pi r^2$ with density $\tilde{f}_A$:

$$\tilde{f}_A(a) = \phi\left(\sqrt{\frac{a}{\pi}}\right) \frac{1}{2\sqrt{a\pi}} = \frac{\sqrt{\frac{a}{\pi}}}{\mathbb{E}(R)} \left(\int_{\sqrt{\frac{a}{\pi}}}^\infty \frac{f(r)}{\sqrt{r^2 - \frac{a}{\pi}}} \mathrm{d}r \right) \frac{1}{2\sqrt{a\pi}} = \frac{1}{2\pi\mathbb{E}(R)} \int_{\sqrt{\frac{a}{\pi}}}^\infty \frac{f(r)}{\sqrt{r^2 - \frac{a}{\pi}}} \mathrm{d}r, \tag{8.1}$$

which we obtain via (7.7). Via a change of variables we express $f(r)$ in (8.1) in terms of f_V. This density f_V is defined as the probability density function which describes the volume distribution of the spheres. Hence, we substitute:

$$r = \left(\frac{3v}{4\pi}\right)^{\frac{1}{3}} \iff v = \frac{4}{3}\pi r^3.$$

Once again, we use the change of variables formula (7.7) and obtain:

$$f(r) = f_V\left(\frac{4}{3}\pi r^3\right) 4\pi r^2.$$

Let us also define the following constant:

$$a_{\max} := \frac{\pi}{\left(\frac{4}{3}\pi\right)^{\frac{2}{3}}} = \sqrt[3]{\frac{9\pi}{16}}. \tag{8.2}$$

This constant $a_{\max}$ is the largest possible cross section area of a sphere of volume 1. Then, it can be shown that:

$$\tilde{f}_A(a) = \frac{3}{4a_{\max}\mathbb{E}(V^{\frac{1}{3}})} \int_{\frac{a}{a_{\max}}}^\infty \frac{f_V\left(x^{\frac{3}{2}}\right)}{\sqrt{1 - \frac{a}{x a_{\max}}}} \mathrm{d}x. \tag{8.3}$$

See appendix E for more details on the computation). The original equation of Wicksell contained the average radius, after the change of variables (8.3) the sectional area pdf instead depends on the expected value of a random variable $V \sim f_V$. The reason (3.1) is rewritten such that it is expressed in terms of probability densities of the observed sectional area distribution and the sphere volume distribution is that it allows for easier comparison to model 1 and the estimator (7.9). As in model 1, a similar notation is used:

- Let f_A be the probability density function of the observed areas of the spheres (cross sectional areas), and let $A \sim f_A$.
- Let $f_{V|I}$ be the probability density function of the volumes of the spheres that appear in the cross section (the observed volume distribution), and let $V_I \sim f_{V|I}$.
- Let f_V be the probability density function of the volumes of the spheres.
- Let $f_{A^{(1)}}$ be the probability density function of the area resulting of taking a random cross section of a sphere of volume 1, and let $A^{(1)} \sim f_{A^{(1)}}$.

In the following sections we show that:

$$A \overset{d}{=} V_I^{\frac{2}{3}} A^{(1)}, \tag{8.4}$$

where $\overset{d}{=}$ denotes equality in distribution, and where V_I and $A^{(1)}$ are independent. Recall that the relationship between the densities f_V and $f_{V|I}$ was defined in (7.6) as:

$$f_{V|I}(v) = \frac{v^{\frac{1}{3}} f_V(v)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)},$$

where $V \sim f_V$.

8.1.1 The sectional area distribution of a sphere of volume 1

Assume that we have a domain $\Omega = [0, l]^3$ for some $l > 0$ which contains non-overlapping spheres $B_1, \ldots, B_n$. Let $V_i := |B_i|$ denote the volume of sphere B_i.

As with the Laguerre diagrams a random horizontal cross section is taken at some height $z_k \sim U(0, l)$. We compute the intersection of each sphere with the plane:

$$S_i := B_i \cap \{(x, y, z) \in \mathbb{R}^3 : z = z_k\} \quad \text{for } i \in \{1, \ldots, n\}.$$

We define the following index set:

$$I = \{i \in \{1, \ldots, n\} : S_i \neq \emptyset\}. \tag{8.5}$$

Then, the following areas are observed:

$$A_i = \text{area}(S_i) \quad \text{for } i \in I.$$

Trivially, we have that:

$$A_i = \underbrace{\left(\frac{A_i}{V_i^{\frac{2}{3}}}\right)}_{:=A^{(1)}} V_i^{\frac{2}{3}} \quad \text{for } i \in I.$$

Remark. *While we define $A^{(1)}$ as the fraction $\frac{A_i}{V_i^{\frac{2}{3}}}$ we show that it is equivalent to define it as the area of a random cross section of a sphere of volume 1.*

Assume we have a sphere of radius R, and in a cross section we observe a circle of radius r. The volume, cross section area and $A^{(1)}$ are respectively given by:

$$V = \frac{4}{3}\pi R^3, \qquad A = \pi r^2, \qquad A^{(1)} = \frac{A}{V^{\frac{2}{3}}} = \frac{\pi r^2}{\left(\frac{4}{3}\pi R^3\right)^{\frac{2}{3}}} = a_{\max}\frac{r^2}{R^2}, \tag{8.6}$$

with $a_{\max}$ as in (8.2). We now use that the probability of observing a circle of radius r with $r_1 \leq r \leq r_2$. When cutting a sphere with radius R at an arbitrary height this probability is given by:

$$\mathbb{P}(r_1 \leq r \leq r_2) = \frac{1}{R}\left(\sqrt{R^2 - r_1^2} - \sqrt{R^2 - r_2^2}\right).$$

This probability is used to determine the distribution function $F_{A^{(1)}}(a)$. Clearly, the expression in (8.6) shows that $0 \leq A^{(1)} \leq a_{\max}$. Consider $0 \leq a \leq a_{\max}$, the distribution function $F_{A^{(1)}}(a)$ for a sphere with radius R does not depend on R (as a result it also does not depend on the volume of the sphere):

$$F_{A^{(1)}}(a) = \mathbb{P}\left(a_{\max}\frac{r^2}{R^2} \leq a\right) = \mathbb{P}\left(r \leq \sqrt{\frac{a}{a_{\max}}}R\right) = 1 - \frac{1}{R}\sqrt{R^2 - R^2\frac{a}{a_{\max}}} = 1 - \sqrt{1 - \frac{a}{a_{\max}}}.$$

Let us now show that this is also the sectional area distribution of a sphere of volume 1. Let R_1 be the radius of a sphere of volume 1, then:

$$\frac{4}{3}\pi R_1^3 = 1 \iff R_1 = \left(\frac{3}{4\pi}\right)^{\frac{1}{3}}.$$

It follows that the largest cross section area of a sphere of volume 1 is given by $a_{\max}$:

$$\pi R_1^2 = \pi\left(\frac{3}{4\pi}\right)^{\frac{2}{3}} = \sqrt[3]{\frac{9\pi}{16}} = a_{\max}.$$

With $a_{\max}$ as in (8.2). Assume that in a cross section of a sphere with radius R_1 a circle of radius $r \leq R_1$ is observed. The sectional area distribution of this sphere is given by:

$$F_{A^{(1)}}(a) = \mathbb{P}\left(\pi r^2 \leq a\right) = \mathbb{P}\left(r \leq \sqrt{\frac{a}{\pi}}\right) = 1 - \frac{1}{R_1}\sqrt{R_1^2 - \frac{a}{\pi}} = 1 - \sqrt{1 - \frac{a}{\pi R_1^2}} = 1 - \sqrt{1 - \frac{a}{a_{\max}}}.$$

As announced, the same distribution function is obtained. Taking the derivative of this distribution function yields the density $f_{A^{(1)}}$:

$$f_{A^{(1)}}(a) = \frac{\mathrm{d}}{\mathrm{d}a}F_{A^{(1)}}(a) = \frac{1}{2a_{\max}\sqrt{1 - \frac{a}{a_{\max}}}} \quad \text{for } 0 \leq a < a_{\max}. \tag{8.7}$$

8.1.2 Computing the probability density function of the sectional area distribution

In previous section we have shown that indeed

$$A \overset{d}{=} V_I^{\frac{2}{3}} A^{(1)}.$$

Because it was shown that $A^{(1)}$ does not depend on the volume of a sphere this implies that $A^{(1)}$ and $V_I^{\frac{2}{3}}$ are independent. Moreover, the probability density functions of $A^{(1)}$ was derived in (8.7). Let us now plug the density of $f_{A^{(1)}}$ (8.7) into the estimator (7.9):

$$\begin{aligned}
f_A(a) &= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} f_{A^{(1)}}(y) f_V\left(\left(\frac{a}{y}\right)^{\frac{3}{2}}\right) \frac{1}{y^2} \mathrm{d}y \\
&= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})} \int_0^{a_{\max}} \frac{1}{2a_{\max}\sqrt{1-\frac{y}{a_{\max}}}} f_V\left(\left(\frac{a}{y}\right)^{\frac{3}{2}}\right) \frac{1}{y^2} \mathrm{d}y \\
&= \frac{3}{4a_{\max}\mathbb{E}(V^{\frac{1}{3}})} \int_{\frac{a}{a_{\max}}}^{\infty} \frac{f_V\left(x^{\frac{3}{2}}\right)}{\sqrt{1-\frac{a}{xa_{\max}}}} \mathrm{d}x \qquad \text{substitute } x = \frac{a}{y} \\
&= \tilde{f}_A(a).
\end{aligned}$$

We may now conclude that indeed $f_A \equiv \tilde{f}_A$ and therefore when applied to Wicksell's corpuscle problem the estimator (7.9) corresponds to the analytical solution.

8.2 The corpuscle problem for similar convex particles

We have shown that the estimator for the sectional area distribution corresponds to the integral relationship that describes Wicksell's corpuscle problem (3.1). However, with this knowledge it may still seem a bit ambitious to then simply apply the estimator to Laguerre diagrams. In this section we show that the estimator also corresponds to the sectional area distribution pdf in a generalization of Wicksell's problem. Say we take some convex particle of our choice. Assume that a domain is filled with instances of this particular convex particle, in random sizes and orientations. More precisely, we assume all the convex particles in the domain to be similar (recall Definition 5). Such a convex particle is also referred to as a convex body, a convex body is a compact convex set with a non-empty interior.

The analytical derivation of the pdf associated with the sectional area distribution in this problem can be found in Santaló's works on integral geometry [34, p. 286, 287]. Let us introduce the necessary notation and the pdf as described in [34].

Suppose we have some convex body Q which contains a certain number of convex and non overlapping particles distributed at random. Assume that all particles are similar to some convex body K. Let λ be the ratio of similitude. We denote K_λ to be the convex body which is similar to K with ratio λ. Note that this means that $K_1 = K$. Let $H(\lambda)$ be the pdf which describes the distribution of the ratios of similarity of the convex particles. We intersect Q with a random plane E and let $h(\sigma)$ be the pdf which describes the distribution of the observed sectional areas. Finally, let $\phi(\sigma)$ be the pdf of the area σ of $E \cap K$ and denote by σ_m the maximal value of σ (this means that ϕ is supported on $(0, \sigma_m]$). Then, Santaló [34] derives the following equation:

$$\int_{\sqrt{\frac{\sigma}{\sigma_m}}}^{\infty} \phi\left(\frac{\sigma}{\lambda^2}\right) \frac{H(\lambda)}{\lambda} \mathrm{d}\lambda = \alpha h(\sigma), \tag{8.8}$$

where α is a constant related to the normalization constant of the density h. Now we need to know the relation between $H(\lambda)$ and the volume distribution in order to show that this expression

is equivalent to our area estimator. This is fairly straightforward. Note that the volumes of K_λ and K are related as follows:

$$|K_\lambda| = \lambda^3 |K| .$$

Now we simply assume that K has volume 1. Hence, by setting $V = \lambda^3 = |K_\lambda|$ we get:

$$f_V(\lambda^3) = \frac{H(\lambda)}{3\lambda^2} \iff H(\lambda) = 3\lambda^2 f_V(\lambda^3).$$

As before, f_V indicates the pdf of the volume distribution of the particles. Plugging this into (8.8) yields:

$$\begin{aligned} h(\sigma) &= \frac{1}{\alpha} \int_{\sqrt{\frac{\sigma}{\sigma_m}}}^{\infty} \phi\left(\frac{\sigma}{\lambda^2}\right) \frac{3\lambda^2 f_V(\lambda^3)}{\lambda} \mathrm{d}\lambda \\ &= \frac{1}{\alpha} \frac{3\sigma}{2} \int_0^{\sigma_m} \phi(x)\, f_V\left(\left(\frac{\sigma}{x}\right)^{\frac{3}{2}}\right) \frac{1}{x^2} \mathrm{d}x \qquad \text{substitute } x = \frac{\sigma}{\lambda^2}. \end{aligned} \tag{8.9}$$

Now we rewrite (8.9) in terms of the notation that was introduced earlier. This means that $h \equiv f_A$ and $\phi \equiv f_{A^{(1)}}$. Then we observe that the pdfs in (8.9) and (7.9) are equal up to the multiplication of a constant, as the constant α has not yet been computed. Because both are proper density functions they must therefore be equal and we have:

$$\frac{1}{\alpha} = \frac{1}{\mathbb{E}\left(V^{\frac{1}{3}}\right)} \iff \alpha = \mathbb{E}\left(V^{\frac{1}{3}}\right).$$

When assumption 1 holds exactly, meaning that all cells in the Laguerre diagrams we consider are similar, our estimator corresponds to the analytical solution (8.9). This justifies assumption 1.

Remark. *In the previous chapter we state that $F_{A^{(1)}}$ may be interpreted as the sectional area distribution of the average cell which has volume 1. We could say that we interpret K as this average cell, and cells of the Laguerre diagram are assumed to be approximately similar to this average cell.*

8.2.1 Solution of the corpuscle problem for similar convex particles

Recall that for Wicksell's corpuscle problem, the analytical solution was presented in chapter 3. In Wicksell's problem f is the unknown the probability density function associated with the radii of the 3D spheres. ϕ is the known probability density function of the radii of the apparent 2D circles. In chapter 3 we remarked that f can be expressed in terms of ϕ, which solves the problem. A solution of Santaló's integral equation, (8.8) has only recently been discovered in [35]. By a solution of this equation we refer to solving it for $H(\lambda)$. In this paper this solution is derived by first assuming it has a particular form, the solution is eventually derived and then the authors provide some technical conditions for its existence and uniqueness. Due to the way we derived our estimator we can provide a direct way of arriving at this solution. This is a result of the decomposition of the sectional area distribution into the product of two independent random variables. Let us now derive this result.

Let $\Lambda \sim H$ and define:

$$H_I(\lambda) = \frac{\lambda H(\lambda)}{\mathbb{E}(\Lambda)} \quad \text{for } \lambda > 0.$$

Clearly, $H_I(\lambda)$ is a proper probability density function. Let $\Lambda_I \sim H_I$, $A \sim h$ and $B \sim \phi$. We now show that the following relationship holds:

$$A \overset{d}{=} B\Lambda_I^2.$$

with B and Λ_I^2 independent, and $\overset{d}{=}$ denoting equality in distribution. The density of Λ_I^2 denoted by $f_{\Lambda_I^2}$ is given by:

$$f_{\Lambda_I^2}(\lambda) = H_I(\sqrt{\lambda})\frac{1}{2\sqrt{\lambda}} = \frac{\sqrt{\lambda}H(\sqrt{\lambda})}{\mathbb{E}(\Lambda)}\frac{1}{2\sqrt{\lambda}} = \frac{1}{2}\frac{H(\sqrt{\lambda})}{\mathbb{E}(\Lambda)}. \tag{8.10}$$

Via (7.4) we compute the density function of the random variable $A := B\Lambda_I^2$ which we denote by $\tilde{h}$, assuming B and Λ_I^2 are independent:

$$\begin{aligned}
\tilde{h}(\sigma) &= \int_0^{\sigma_m} \phi(x) f_{\Lambda_I^2}\left(\frac{\sigma}{x}\right)\frac{1}{x}\mathrm{d}x \\
&= \frac{1}{\mathbb{E}(\Lambda)}\int_0^{\sigma_m} \phi(x)\frac{1}{2x}H\left(\left(\frac{\sigma}{x}\right)^{\frac{1}{2}}\right)\mathrm{d}x \\
&= \frac{1}{\mathbb{E}(\Lambda)}\int_{\sqrt{\frac{\sigma}{\sigma_m}}}^{\infty} \phi\left(\frac{\sigma}{\lambda^2}\right)\frac{H(\lambda)}{\lambda}\mathrm{d}\lambda \qquad \text{substitute } \lambda = \left(\frac{\sigma}{x}\right)^{\frac{1}{2}}.
\end{aligned} \tag{8.11}$$

As before, the functions $\tilde{h}$ as in (8.11) and h as in (8.8) are equal up to the multiplication of a constant. As both are proper density functions we obtain that $\tilde{h} \equiv h$ and:

$$\alpha = \mathbb{E}(\Lambda).$$

The solution of the problem can be expressed in terms of (inverse) Mellin transforms [35]. There are various technical conditions for the existence of (inverse) Mellin transforms, in the remaining derivation we simply assume that these transforms exist. For an overview of the Mellin transform we refer to chapter 12 of [36]. For technicalities related to the Mellin transform in this problem we refer to [35]. Let us now consider the definition of the Mellin transform.

Definition 6 (Mellin transform)**.** [36]. The Mellin transform is defined for a function $f : (0,\infty) \to \mathbb{R}$ as:

$$\mathcal{M}(f(x))(s) = f^*(s) = \int_0^{\infty} x^{s-1} f(x)\mathrm{d}x,$$

whenever this integral converges absolutely. In this setting we consider $s \in \mathbb{C}$, meaning that it is a complex integral.

These integrals do not typically converge for all $s \in \mathbb{C}$. In the case that f is a pdf on $(0,\infty)$ the integral converges at least for $s \in \mathbb{C}$ such that $\mathrm{Re}(s) = 1$. $\mathrm{Re}(s)$ indicates the real part of s. There exists also an inverse Mellin transform $\mathcal{M}^{-1}$ with an explicit formula which under some technical assumptions satisfies: $\mathcal{M}^{-1}(f^*(s))(x) = f(x)$. There is a clear relationship between Mellin transforms and expected values of random variables. Given a density f and a random variable $X \sim f$ we have $f^*(s) = \mathbb{E}(X^{s-1})$. If we consider the Mellin transform of the density associated with the product of two independent random variables then this is computed as the product of the Mellin transforms of their densities. This is similar as to how the expectation of the product of two independent random variables is the product of their expectations, and is referred to as the Mellin convolution theorem. Applying this theorem yields:

$$h^*(s) = \phi^*(s) f^*_{\Lambda_I^2}(s). \tag{8.12}$$

The Mellin transform of $f_{\Lambda_I^2}$ (8.10) may be written as:

$$f^*_{\Lambda_I^2}(s) = \int_0^\infty x^{s-1} \frac{1}{2} \frac{H(\sqrt{x})}{\mathbb{E}(\Lambda)} \mathrm{d}x \overset{(\lambda=\sqrt{x})}{=} \int_0^\infty \lambda^{2s-1} \frac{H(\lambda)}{\mathbb{E}(\Lambda)} \mathrm{d}\lambda = \frac{\mathcal{M}\left(H(\lambda)\right)(2s)}{\mathbb{E}(\Lambda)} = \frac{H^*(2s)}{\mathbb{E}(\Lambda)}.$$

Plugging this back into (8.12) yields:

$$h^*(s) = \phi^*(s) \frac{H^*(2s)}{\mathbb{E}(\Lambda)} \iff H^*(s) = \mathbb{E}(\Lambda) \frac{h^*(\frac{s}{2})}{\phi^*(\frac{s}{2})}.$$

Note that the above equivalence holds when assuming $\phi^*(\frac{s}{2}) \neq 0$. Taking the inverse Mellin transform on both sides yields the solution as found in [35]:

$$H(\lambda) = \mathcal{M}^{-1}\left(H^*(s)\right)(\lambda) = \mathcal{M}^{-1}\left(\mathbb{E}(\Lambda) \frac{h^*(\frac{s}{2})}{\phi^*(\frac{s}{2})}\right)(\lambda) = \mathbb{E}(\Lambda) \mathcal{M}^{-1}\left(\frac{h^*(\frac{s}{2})}{\phi^*(\frac{s}{2})}\right)(\lambda). \tag{8.13}$$

A property of the inverse Mellin transform is that constants may be pulled out, hence we may pull $\mathbb{E}(\Lambda)$ out of the inverse transform.

It is interesting to note that this problem may also be interpreted as a multiplicative deconvolution problem. In a more general setting the problem may be described as follows, let $Y \sim f_Y$, $X \sim f_X$ and $\epsilon \sim f_\epsilon$ then we consider:

$$Y = X\epsilon,$$

With X and ϵ independent. The densities f_ϵ and f_Y are assumed to be known and we need to solve for the density f_X. Effectively the variable of interest is distorted by multiplicative noise, and only this noisy data is observed. In [37] it is shown that a sufficient condition for the determination of f_X to be an identifiable problem is that at least one of the variables X or ϵ to be non-negative. In practice we have data, samples of f_Y, and we may attempt to estimate f_X. Estimators for f_X have been proposed in for example [37], in [38] another estimator is considered in the case that both X and ϵ are non-negative. These estimators use a non-parametric estimator of the Mellin transform of f_Y and they assume that the Mellin transform of f_ϵ is known. One reason we do not attempt to use these is that in our case even f_ϵ or in the notation we tend to use: $f_{A^{(1)}} \equiv \phi$ is not a known function. The estimators in the aforementioned paper do assume this density to be known and further analysis is required to determine their performance when only an approximation of this density is provided. Finally, we need to stress that in our case (8.8) is merely an approximation to the problem we consider. For now it seems more sensible to investigate whether we can use our estimator in the case of parametric volume distributions, which is much more tractable. Nonetheless, if the proposed estimator perform well enough in the parametric case, it may provide a good starting point for estimating volume distributions in a non-parametric setting.

8.3 Conclusion

In this chapter we considered two stereological problems, Wicksell's corpuscle problem and one of its possible generalizations. In both problems an analytical expression for the sectional area distribution is available and it corresponds to the estimator for the sectional area distribution proposed in this thesis. This provides support on the use of the estimator for Laguerre diagrams. We should keep in mind that using this estimator means that we assume cells in a Laguerre diagram to be approximately similar (assumption 1).

9

A lognormal cell volume distribution

An estimator for the cross sectional area distribution was derived in chapter 7 and in chapter 8 we provided theoretical support for its use. Throughout this chapter we assume that that the volume distribution of the cells follows a lognormal distribution. As mentioned before, in materials science the lognormal distribution is a well-known parametric distribution used for approximating the grain size distribution both in 3D and in 2D. This motivates us to consider this particular instance. We obtain the estimator for the cross sectional area distribution (7.10) in the lognormal case, and examine two possible approaches for estimating the parameters μ and σ of the cell volume distribution. Finally, we perform simulations to verify the reasonability of this approach for estimating the parameters of the underlying lognormal distribution from a sample of observed areas.

Recall the estimator in its general form:

$$\begin{aligned}\hat{f}_A(a) &= \frac{3a}{2\mathbb{E}(V^{\frac{1}{3}})}\frac{1}{N}\sum_{i=1}^{N} f_V\left(\left(\frac{a}{a_i^{(1)}}\right)^{\frac{3}{2}}\right)\frac{1}{\left(a_i^{(1)}\right)^2} \\ &= \frac{1}{N}\sum_{i=1}^{N} f_{V^{\frac{2}{3}}|I}\left(\frac{a}{a_i^{(1)}}\right)\frac{1}{a_i^{(1)}}.\end{aligned}$$

with $f_{V^{\frac{2}{3}}|I}$ defined as in (7.8). As it turns out, when f_V is the density of a lognormal distribution then $f_{V^{\frac{2}{3}}|I}$ also corresponds to a lognormal distribution but with different parameters. More details on the derivation are presented in appendix F. For a lognormal volume distribution with parameters μ and σ^2 we show that $f_{V^{\frac{2}{3}}|I}$ is the pdf of a lognormal distribution with parameters $\bar{\mu} = \frac{2}{3}\mu + \frac{2}{9}\sigma^2$ and $\bar{\sigma}^2 = \frac{4}{9}\sigma^2$. As a result the estimator for the area distribution (7.10) becomes a mixture of lognormal distributions and its density function and distribution function are given

by:

$$\hat{f}_A(a|\mu,\sigma) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{a\sqrt{2\pi\frac{4}{9}\sigma^2}} \exp\left(-\frac{\left(\log(a) - \frac{2}{3}\mu - \frac{2}{9}\sigma^2 - \log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right) \tag{9.1}$$

$$\hat{F}_A(a|\mu,\sigma) = \frac{1}{N}\sum_{i=1}^{N} \Phi\left(\frac{\log(a) - \frac{2}{3}\mu - \frac{2}{9}\sigma^2 - \log\left(a_i^{(1)}\right)}{\frac{2}{3}\sigma}\right). \tag{9.2}$$

Here, we let Φ denote the distribution function of a standard normal distribution. Given that we observe areas $a_1, \ldots, a_m$, we would like to fit the proposed estimator of the sectional area distribution. This can be done by estimating the parameters μ and σ. We can then assess whether the distribution as in (9.1) and (9.2) provides a good fit and whether we retrieve the original μ and σ of the volume distribution. We derive two estimators: the first one based on the method of moments, the second based on numerically computing maximum likelihood estimates of μ and σ.

9.1 Parameter estimation

We may obtain estimators for μ and σ via the method of moments, by (7.14). This implies solving the following system of equations:

$$\frac{1}{m}\sum_{j=1}^{m} a_j = \frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)\exp\left(\frac{2}{3}\mu + \frac{4}{9}\sigma^2\right) \tag{9.3}$$

$$\frac{1}{m}\sum_{j=1}^{m} a_j^2 = \frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2\exp\left(\frac{4}{3}\mu + \frac{4}{3}\sigma^2\right). \tag{9.4}$$

Solving for μ and σ yields the following moment estimators:

$$\begin{aligned}
\hat{\mu}_{\mathrm{MM}} &= \frac{3}{2}\left(3\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j}{\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}}\right) - \log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2}\right)\right) \\
&= \frac{3}{2}\left(\log\left(\frac{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^3}{\frac{1}{m}\sum_{j=1}^{m} a_j^2}\right) - \log\left(\frac{\left(\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}\right)^3}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2}\right)\right) \\
\hat{\sigma} &= \frac{3}{2}\sqrt{\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2}\right) - 2\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j}{\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}}\right)} \\
&= \frac{3}{2}\sqrt{\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^2}\right) + \log\left(\frac{\left(\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2}\right)}.
\end{aligned} \tag{9.5}$$

While $\hat{\mu}_{\text{MM}}$ is always well-defined, this is not in general true for $\hat{\sigma}$. Due to Jensen's inequality we have:

$$\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^2} \geq 1 \quad \text{and} \quad \frac{\left(\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2} \leq 1.$$

This means that the term within the square root may be negative. If the estimator tends to provide reasonable estimate of σ for most values of σ, then we expect to only run into trouble when the true value of σ is close to zero. To ensure that the estimator is always well-defined, we use a corrected method of moments estimator for σ:

$$\hat{\sigma}_{\text{MM}} = \frac{3}{2}\sqrt{\left|\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^2}\right) + \log\left(\frac{\left(\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2}\right)\right|}. \tag{9.6}$$

It is useful to remark that $\frac{1}{N}\sum_{i=1}^{N} a_i^{(1)}$ and $\frac{1}{N}\sum_{i=1}^{N}\left(a_i^{(1)}\right)^2$ were computed before (see Table 7.1). Hence, these values may be plugged into (9.5) and (9.6) to obtain simple formulas for the estimators:

$$\hat{\mu}_{\text{MM}} \approx \frac{3}{2}\left(\log\left(\frac{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^3}{\frac{1}{m}\sum_{j=1}^{m} a_j^2}\right) + 0.58662\right) \tag{9.7}$$

$$\hat{\sigma}_{\text{MM}} \approx \frac{3}{2}\sqrt{\left|\log\left(\frac{\frac{1}{m}\sum_{j=1}^{m} a_j^2}{\left(\frac{1}{m}\sum_{j=1}^{m} a_j\right)^2}\right) - 0.26963\right|}. \tag{9.8}$$

Now that method of moment estimators for μ and σ are derived, we consider the maximum likelihood method. The log-likelihood is given by:

$$\begin{aligned} l(\mu,\sigma) &= \sum_{j=1}^{m} \log\left(\hat{f}_A(a_j|\mu,\sigma)\right) \\ &= -\left(\sum_{j=1}^{m}\log(a_j)\right) - \frac{m}{2}\log\left(2\pi\frac{4}{9}\sigma^2\right) + \\ &\quad + \sum_{j=1}^{m}\log\left(\frac{1}{N}\sum_{i=1}^{N}\exp\left(-\frac{\left(\log(a_j) - \frac{2}{3}\mu - \frac{2}{9}\sigma^2 - \log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)\right). \end{aligned} \tag{9.9}$$

The maximum likelihood estimator is then defined as:

$$(\hat{\mu}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}) = \arg\max_{\mu,\sigma} l(\mu,\sigma). \tag{9.10}$$

Let us define:

$$b_{ij} := \log(a_j) - \frac{2}{3}\mu - \frac{2}{9}\sigma^2 - \log\left(a_i^{(1)}\right) \quad \text{and} \quad \beta := \frac{1}{\frac{8}{9}\sigma^2}.$$

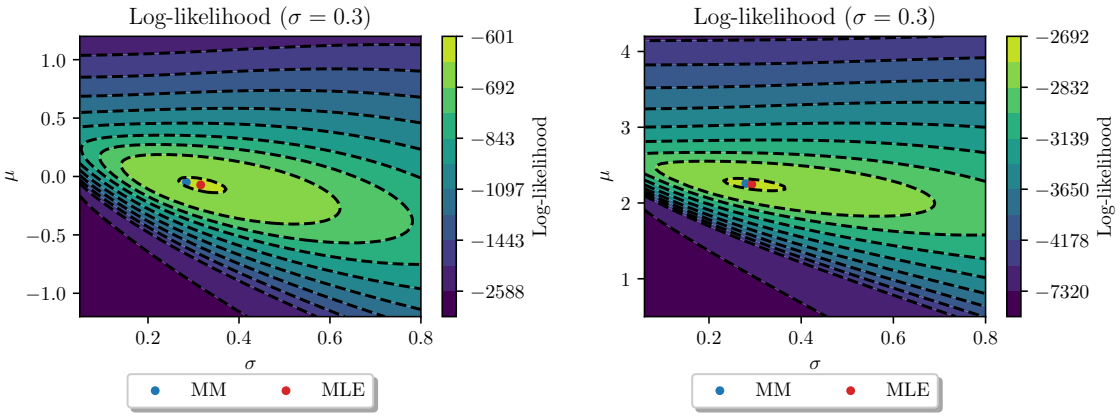
Via straightforward computation we find the partial derivatives of the log-likelihood. For more details of the computation see appendix D.

$$\frac{\partial}{\partial\mu}l(\mu,\sigma) = \sum_{j=1}^{m}\frac{\sum_{i=1}^{N}\exp\left(-b_{ij}^2\beta\right)\frac{4}{3}b_{ij}\beta}{\sum_{i=1}^{N}\exp\left(-b_{ij}^2\beta\right)} \tag{9.11}$$

$$\frac{\partial}{\partial\sigma}l(\mu,\sigma) = -m\beta\frac{8}{9}\sigma + \sum_{j=1}^{m}\frac{\sum_{i=1}^{N}\exp\left(-b_{ij}^2\beta\right)\left(\frac{4}{9}\sigma^2 b_{ij} + b_{ij}^2\right)\frac{16}{9}\sigma\beta^2}{\sum_{i=1}^{N}\exp\left(-b_{ij}^2\beta\right)}. \tag{9.12}$$

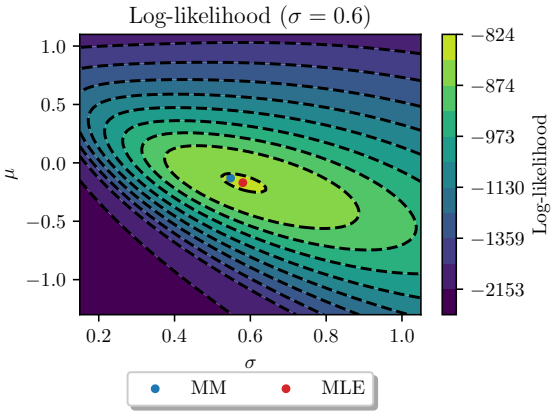
We first run preliminary simulations. We generate a Laguerre diagram with $n = 1000$ cells with a lognormal volume distribution, then we take $K = 1000$ horizontal cross sections (hence we compute 1000 Laguerre diagrams and take a single cross section of each diagram). The samples of cross section areas are then combined into a single large sample for each lognormal distribution considered. In simulations in the next section we do not combine samples: we estimate the parameters from just one cross section. We take $\sigma \in \{0.3, 0.6, 0.9\}$ and either $\mu = -\frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 1$ or $\mu = \log(10) - \frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 10$. Then μ and σ are estimated from the observed sectional areas both via the method of moments (9.5), (9.6) and via maximum likelihood estimation (9.10). The maximum likelihood estimates of μ and σ are obtained by numerically maximizing the log-likelihood (9.9) using the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS), via the implementation that is provided in the open-source Python package SciPy. As initial value, the estimates of μ and σ obtained with the method of moments are used. Because it is not obvious that the log-likelihood is a concave function contour plots of the log-likelihood are shown in Figure 9.1. It can clearly be seen that in these cases there is indeed a unique maximum. For further simulations we will assume that the log-likelihood has a unique maximum

In Table 9.1 the results of these simulations are shown. In Table 9.1 we observe the actual values of μ and σ, which describe the lognormal cell volume distribution as well as the result of estimating the parameters from the cross sectional areas. It is important to stress that this table only contains point estimates. Each pair of estimates of μ and σ is estimated from a large sample of areas, which is obtained by combining 1000 cross sections. In Figure 9.2 we show the result of plugging the true values of μ and σ as well as the estimates (as presented in Table 9.1) into the estimator of the sectional area distribution (9.1). From a graphical inspection, we can conlude that the estimator $\hat{f}_A$ describes the observed distributions quite accurately. The three densities, namely $\hat{f}_A$ with (μ, σ), $(\hat{\mu}_{\mathrm{MM}}, \hat{\sigma}_{\mathrm{MM}})$ and $(\hat{\mu}_{\mathrm{MLE}}, \hat{\sigma}_{\mathrm{MLE}})$ are quite close in all the considered instances.

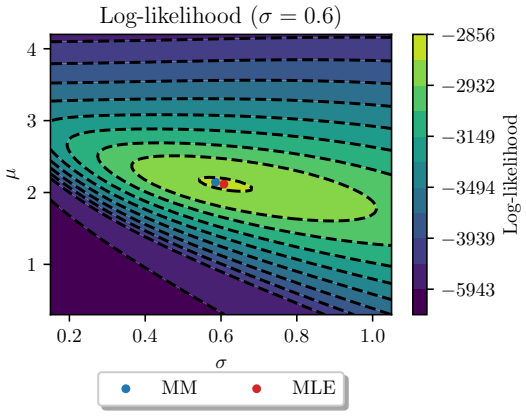


(a) $\mathbb{E}(V) = 1$

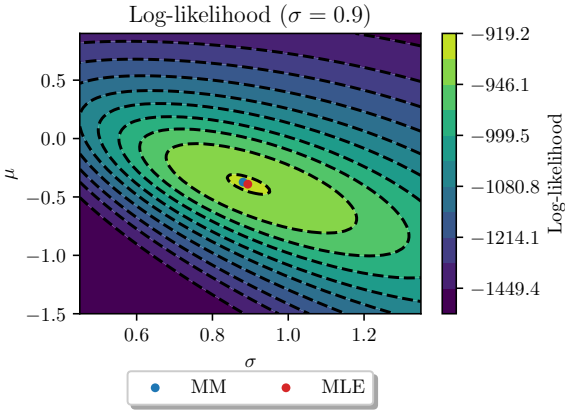
(b) $\mathbb{E}(V) = 10$



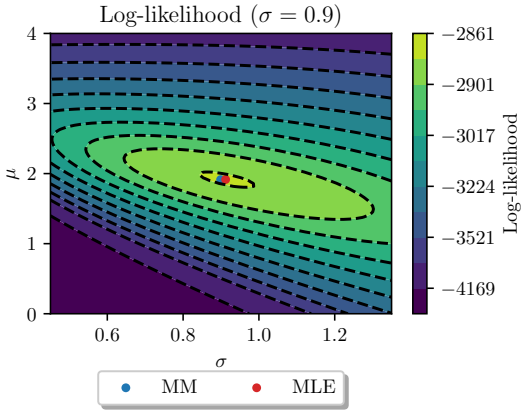
(c) $\mathbb{E}(V) = 1$



(d) $\mathbb{E}(V) = 10$

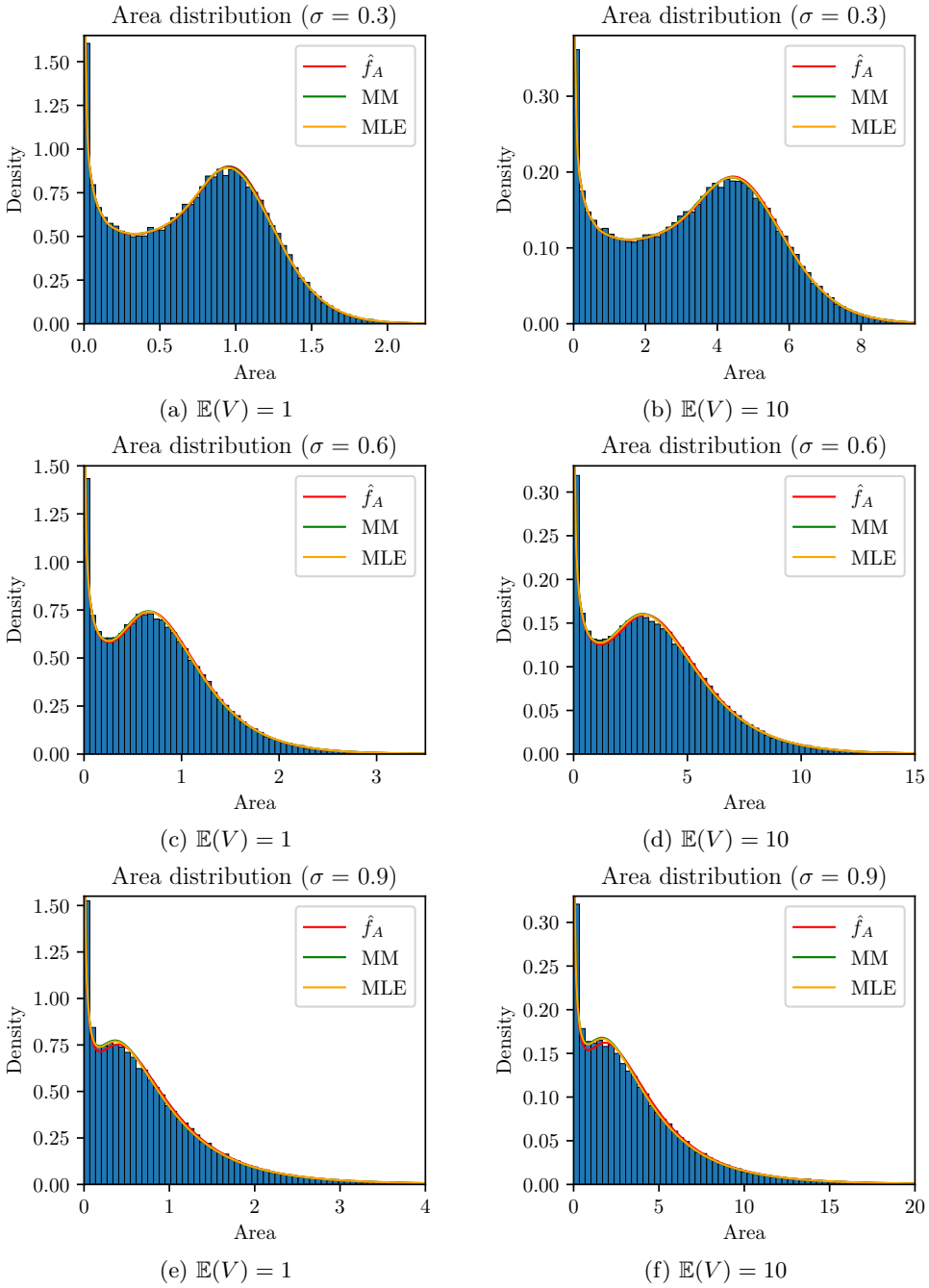


(e) $\mathbb{E}(V) = 1$



(f) $\mathbb{E}(V) = 10$

Figure 9.1: Contour plots of the log-likelihood for samples of cross sectional areas from Laguerre diagrams with a lognormal cell volume distribution. We consider $\sigma \in \{0.3, 0.6, 0.9\}$ and either $\mu = -\frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 1$ or $\mu = \log(10) - \frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 10$ (where $\mathbb{E}(V)$ is the expected cell volume).



(a) $\mathbb{E}(V) = 1$

(b) $\mathbb{E}(V) = 10$

(c) $\mathbb{E}(V) = 1$

(d) $\mathbb{E}(V) = 10$

(e) $\mathbb{E}(V) = 1$

(f) $\mathbb{E}(V) = 10$

Figure 9.2: The cross sectional area distribution when considering Laguerre diagrams with a lognormal cell volume distribution. We consider $\sigma \in \{0.3, 0.6, 0.9\}$ and either $\mu = -\frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 1$ or $\mu = \log(10) - \frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 10$ (where $\mathbb{E}(V)$ is the expected cell volume). We plot $\hat{f}_A$ with the true values of μ and σ, along with $\hat{f}_A$ where μ, σ are estimated via the Method of Moments (MM), or Maximum Likelihood Estimation (MLE).

σ	$\hat{\sigma}_{\text{MM}}$	$\hat{\sigma}_{\text{MLE}}$	μ	$\hat{\mu}_{\text{MM}}$	$\hat{\mu}_{\text{MLE}}$
0.3	0.305	0.309	−0.0450	−0.0485	−0.0504
0.3	0.307	0.305	2.26	2.25	2.25
0.6	0.614	0.611	−0.180	−0.200	−0.191
0.6	0.617	0.615	2.12	2.10	2.10
0.9	0.929	0.930	−0.405	−0.458	−0.449
0.9	0.933	0.934	1.90	1.83	1.84

Table 9.1: Simulation results. μ and σ are the true values chosen for the volume distribution for the Laguerre diagram; $\hat{\mu}_{\text{MM}}$ and $\hat{\sigma}_{\text{MM}}$ are the method of moments estimates computed from the observed cross sectional areas (9.5), (9.6). $\hat{\mu}_{\text{MLE}}$ and $\hat{\sigma}_{\text{MLE}}$ are the maximum likelihood estimates computed from the observed cross sectional areas (9.10).

9.2 Simulation study

In this section a more extensive approach to simulations is considered. We are particularly interested in the performance of the estimators when estimating the parameters μ and σ from a single cross section. By repeating simulations multiple times we investigate how close the parameter estimates are to the actual parameter values.

In this section we use the following simulation procedure. First, the number of cells n along with parameters μ and σ of the lognormal volume distribution are chosen. Then, we:

1. Generate a Laguerre diagram with n cells, and a lognormal cell volume distribution with parameters μ and σ as per the procedure in section 4.4.1.
2. Take a single cross section as per the procedure in section 4.4.2. This yields a sample $a_1, \ldots, a_m$ of cross sectional areas.
3. Compute the estimates $\hat{\mu}_{\text{MM}}, \hat{\sigma}_{\text{MM}}$ as in (9.5), (9.6) and $\hat{\mu}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}$ as in (9.10) given the sample $a_1, \ldots, a_m$.
4. Steps 1,2 and 3 are repeated K times such that we have estimates:
   $(\hat{\mu}_{\text{MM},1}, \hat{\sigma}_{\text{MM},1}), (\hat{\mu}_{\text{MM},2}, \hat{\sigma}_{\text{MM},2}), \ldots, (\hat{\mu}_{\text{MM},K}, \hat{\sigma}_{\text{MM},K})$, and:
   $(\hat{\mu}_{\text{MLE},1}, \hat{\sigma}_{\text{MLE},1}), (\hat{\mu}_{\text{MLE},2}, \hat{\sigma}_{\text{MLE},2}), \ldots, (\hat{\mu}_{\text{MLE},K}, \hat{\sigma}_{\text{MLE},K})$.

First, we consider $n = 4000$ and $n = 16000$ cells with various combinations of μ and σ. We repeat steps 1,2 and 3 for $K = 100$ times. Following the outlined steps yields a sample of estimates of μ and σ. For these samples we compute the average of $\hat{\mu}_{\text{MM},1}, \hat{\mu}_{\text{MM},2}, \ldots, \hat{\mu}_{\text{MM},K}$ and the average of $\hat{\sigma}_{\text{MM},1}, \hat{\sigma}_{\text{MM},2}, \ldots, \hat{\sigma}_{\text{MM},K}$; the same is done for the MLE estimates. For these averages a 95% confidence interval is computed via bootstrapping. Specifically the Basic bootstrap, also known as the bootstrap pivotal confidence interval [39] via $B = 10000$ bootstrapped samples. Next, point estimates of the standard deviation are computed for each sample of parameter estimates. Finally, we also compute the average number of observed areas in the samples of cross sectional areas. The results for MM are given in Table 9.2 and the results for MLE are in Table 9.3.

Let us first discuss the results in Table 9.2. It seems reasonable to say that on average the estimates tend to be quite accurate, since the average estimates are close to the actual values.

What is peculiar is that in some cases the average estimates are better for $n = 4000$ cells compared to $n = 16000$ cells. We do not expect a smaller sample of sectional areas to lead to a better estimate. What is expected is that the sample standard deviation of the estimated parameters is smaller in every case, when comparing $n = 16000$ to $n = 4000$. Something which is of concern is that the confidence intervals of the average estimates does not always contain the true value of the parameter. This suggests that the estimator may be biased.

$n = 4000$ cells

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.291	[0.283, 0.306]	0.073	−0.0450	−0.0456	[−0.0515, −0.0355]	0.052	343
0.3	0.290	[0.282, 0.303]	0.066	2.26	2.26	[2.26 , 2.27]	0.054	342
0.6	0.606	[0.600, 0.617]	0.053	−0.180	−0.194	[−0.201 , −0.182]	0.061	335
0.6	0.612	[0.606, 0.622]	0.052	2.12	2.10	[2.09 , 2.11]	0.062	335
0.9	0.907	[0.899, 0.921]	0.071	−0.405	−0.422	[−0.433 , −0.405]	0.091	318
0.9	0.905	[0.893, 0.924]	0.10	1.90	1.87	[1.86 , 1.90]	0.13	320
1.2	1.19	[1.17 , 1.21]	0.13	−0.720	−0.734	[−0.756 , −0.695]	0.19	299
1.2	1.21	[1.19 , 1.23]	0.12	1.58	1.54	[1.51 , 1.58]	0.20	300
1.5	1.48	[1.46 , 1.51]	0.16	−1.13	−1.13	[−1.16 , −1.06]	0.32	274
1.5	1.44	[1.42 , 1.47]	0.17	1.18	1.25	[1.21 , 1.30]	0.29	279

$n = 16000$ cells

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.280	[0.275, 0.288]	0.042	−0.0450	−0.0272	[−0.0308, −0.0210]	0.031	857
0.3	0.282	[0.278, 0.291]	0.042	2.26	2.27	[2.27 , 2.28]	0.031	857
0.6	0.603	[0.600, 0.610]	0.033	−0.180	−0.178	[−0.182 , −0.169]	0.042	835
0.6	0.603	[0.600, 0.610]	0.033	2.12	2.12	[2.12 , 2.13]	0.042	835
0.9	0.908	[0.902, 0.917]	0.050	−0.405	−0.419	[−0.426 , −0.407]	0.061	799
0.9	0.909	[0.903, 0.920]	0.055	1.90	1.88	[1.87 , 1.89]	0.064	803
1.2	1.23	[1.22 , 1.24]	0.087	−0.720	−0.786	[−0.801 , −0.759]	0.13	752
1.2	1.21	[1.20 , 1.23]	0.089	1.58	1.54	[1.53 , 1.57]	0.13	752
1.5	1.51	[1.49 , 1.53]	0.14	−1.13	−1.19	[−1.22 , −1.14]	0.25	695
1.5	1.51	[1.49 , 1.53]	0.13	1.18	1.11	[1.09 , 1.16]	0.24	697

Table 9.2: Simulation results for the estimated parameters μ and σ of the lognormal cell volume distribution of a Laguerre diagram. Results are shown for diagrams with 4000 and 16000 cells, parameters are estimated via MM: (9.5), (9.6). For each μ, σ pair 100 Laguerre diagrams are computed, each time estimating μ and σ from a single cross section of each diagram. sd: standard deviation of the sample of estimated parameters. N_A: mean number of observed cells.

$n = 4000$ cells

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.291	[0.287, 0.297]	0.031	−0.0450	−0.0386	[−0.0423, −0.0324]	0.032	343
0.3	0.291	[0.287, 0.297]	0.034	2.26	2.27	[2.26 , 2.27]	0.037	342
0.6	0.603	[0.597, 0.612]	0.047	−0.180	−0.183	[−0.189 , −0.173]	0.053	335
0.6	0.609	[0.604, 0.618]	0.045	2.12	2.11	[2.10 , 2.12]	0.063	335
0.9	0.921	[0.913, 0.934]	0.068	−0.405	−0.428	[−0.438 , −0.411]	0.087	318
0.9	0.916	[0.908, 0.931]	0.073	1.90	1.87	[1.86 , 1.88]	0.094	320
1.2	1.23	[1.22 , 1.25]	0.096	−0.720	−0.784	[−0.801 , −0.755]	0.15	299
1.2	1.24	[1.23 , 1.26]	0.094	1.58	1.50	[1.48 , 1.53]	0.15	300
1.5	1.59	[1.58 , 1.62]	0.11	−1.13	−1.32	[−1.34 , −1.28]	0.19	274
1.5	1.57	[1.56 , 1.60]	0.12	1.18	1.02	[0.997 , 1.06]	0.21	279

$n = 16000$ cells

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.296	[0.293, 0.300]	0.020	−0.0450	−0.0373	[−0.0398, −0.0330]	0.022	857
0.3	0.296	[0.293, 0.299]	0.020	2.26	2.27	[2.26 , 2.27]	0.022	857
0.6	0.601	[0.598, 0.607]	0.028	−0.180	−0.172	[−0.177 , −0.164]	0.040	835
0.6	0.603	[0.599, 0.609]	0.031	2.12	2.13	[2.12 , 2.14]	0.043	835
0.9	0.910	[0.905, 0.917]	0.039	−0.405	−0.410	[−0.415 , −0.400]	0.049	799
0.9	0.906	[0.901, 0.913]	0.039	1.90	1.89	[1.88 , 1.90]	0.053	803
1.2	1.25	[1.25 , 1.26]	0.057	−0.720	−0.812	[−0.822 , −0.796]	0.085	752
1.2	1.24	[1.24 , 1.25]	0.057	1.58	1.51	[1.50 , 1.53]	0.083	752
1.5	1.60	[1.59 , 1.62]	0.079	−1.13	−1.34	[−1.36 , −1.31]	0.13	695
1.5	1.58	[1.57 , 1.59]	0.073	1.18	1.01	[0.991 , 1.03]	0.12	697

Table 9.3: Simulation results for the estimated parameters μ and σ of the lognormal cell volume distribution of a Laguerre diagram. Results are shown for diagrams with 4000 and 16000 cells, parameters are estimated via MLE: (9.10). For each μ, σ pair 100 Laguerre diagrams are computed, each time estimating μ and σ from a single cross section of each diagram. sd: standard deviation of the sample of estimated parameters. N_A: mean number of observed cells.

Comparing these numbers to the results in Table 9.3 we observe that some things have changed slightly. For one, it seems that in the case of MLE the average estimate is slightly worse for large values of σ compared to MM. Compared to MM, the standard deviation of the sample of parameters for MLE is smaller for each combination of μ and σ. Once again, in multiple cases the confidence intervals do not contain the true value of the parameter, indicating that there is a bias. Luckily, on average MLE still yields quite accurate estimates.

Because we consider $K = 100$ to be a relatively small sample of parameter estimates we now repeat the simulations with $K = 1000$ repetitions. This does mean that we have to use fewer cells in the Laguerre diagrams, in order to obtain the results in a reasonable amount of time. Therefore, we now consider diagrams with $n = 1000$ cells. The results of these simulations are

in Table 9.4. These results are mostly in line with the previous simulations. On average the estimates are again quite accurate, but the confidence intervals do not always contain the actual parameter value.

MM

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.283	[0.279, 0.289]	0.11	−0.0450	−0.0419	[−0.0451, −0.0364]	0.085	136
0.3	0.288	[0.285, 0.295]	0.10	2.26	2.26	[2.26 , 2.27]	0.085	136
0.6	0.600	[0.596, 0.605]	0.088	−0.180	−0.185	[−0.189 , −0.179]	0.10	132
0.6	0.602	[0.599, 0.607]	0.087	2.12	2.11	[2.11 , 2.12]	0.11	132
0.9	0.903	[0.899, 0.911]	0.12	−0.405	−0.431	[−0.437 , −0.420]	0.17	127
0.9	0.895	[0.891, 0.902]	0.11	1.90	1.88	[1.88 , 1.89]	0.16	127
1.2	1.17	[1.17 , 1.18]	0.17	−0.720	−0.723	[−0.732 , −0.706]	0.26	120
1.2	1.18	[1.17 , 1.19]	0.17	1.58	1.56	[1.55 , 1.58]	0.27	120
1.5	1.41	[1.41 , 1.43]	0.20	−1.13	−1.04	[−1.05 , −1.02]	0.34	111
1.5	1.41	[1.41 , 1.43]	0.21	1.18	1.26	[1.24 , 1.28]	0.35	111

MLE

σ				μ				
actual value	average estimate	95% CI	sd	actual value	average estimate	95% CI	sd	N_A
0.3	0.293	[0.291, 0.296]	0.050	−0.0450	−0.0442	[−0.0463, −0.0405]	0.058	136
0.3	0.290	[0.288, 0.294]	0.055	2.26	2.26	[2.26 , 2.27]	0.059	136
0.6	0.599	[0.596, 0.604]	0.079	−0.180	−0.178	[−0.181 , −0.171]	0.10	132
0.6	0.599	[0.596, 0.604]	0.080	2.12	2.12	[2.12 , 2.13]	0.10	132
0.9	0.919	[0.914, 0.926]	0.12	−0.405	−0.436	[−0.442 , −0.426]	0.16	127
0.9	0.913	[0.909, 0.920]	0.11	1.90	1.88	[1.87 , 1.89]	0.16	127
1.2	1.25	[1.24 , 1.25]	0.16	−0.720	−0.812	[−0.821 , −0.798]	0.23	120
1.2	1.24	[1.23 , 1.25]	0.15	1.58	1.49	[1.48 , 1.50]	0.23	120
1.5	1.57	[1.56 , 1.58]	0.19	−1.13	−1.31	[−1.32 , −1.29]	0.32	111
1.5	1.58	[1.57 , 1.59]	0.19	1.18	0.974	[0.962 , 0.994]	0.32	111

Table 9.4: Simulation results for the estimated parameters μ and σ of the lognormal cell volume distribution of a Laguerre diagram. Results are shown for diagrams with 1000 cells, parameters are estimated via MM: (9.5), (9.6) and MLE: (9.10). For each μ, σ pair 1000 Laguerre diagrams are computed, each time estimating μ and σ from a single cross section of each diagram. sd: standard deviation of the sample of estimated parameters. N_A: mean number of observed cells.

For better understanding the behaviour of the bias we make a visualization of the estimated bias. Given parameter estimates $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K$ of a known μ the bias may be estimated as:

$$\widehat{\text{Bias}}(\hat{\mu}, \mu) = \frac{1}{K} \sum_{i=1}^{K} \hat{\mu}_i - \mu.$$

The estimated bias is analogously computed for σ. Similar to how confidence intervals were computed for the average estimate of a parameter a confidence interval may be computed for the estimated bias. The estimated bias along with a 95% confidence interval is computed for the parameter estimates which were obtained for Laguerre diagrams with $n = 1000$ and $n = 16000$ cells. The lognormal distributions we consider are $\sigma \in \{0.3, 0.6, 0.9, 1.2, 1.5\}$ and $\mu = \log(10) - \frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 10$. The results are shown in Figure 9.3. In this figure we observe that the MM estimator appears more dependent on the size of the Laguerre diagram (n) compared to

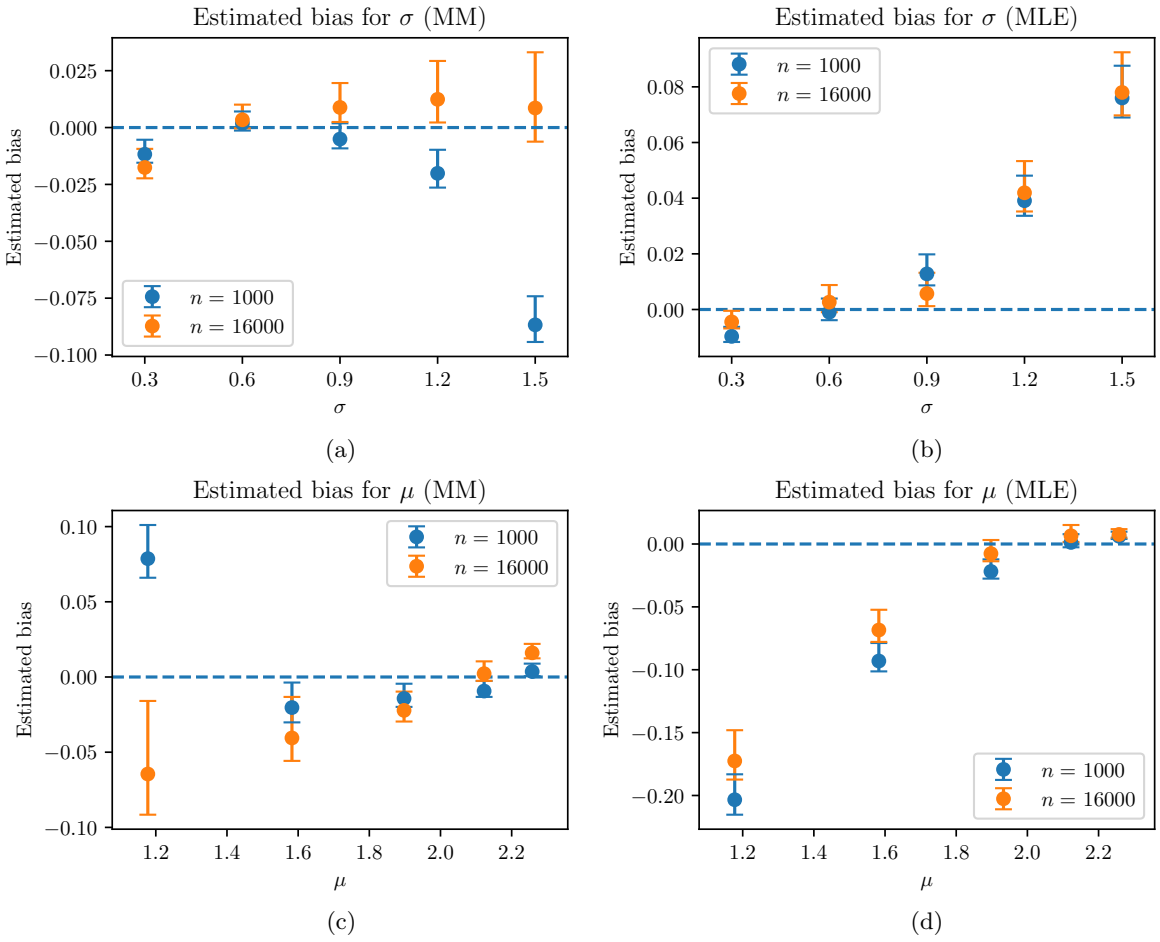


Figure 9.3: Estimated bias with 95% confidence intervals for some of the parameter estimates which are summarized in Tables 9.2, 9.3 and 9.4. We only consider Laguerre diagrams with $n = 1000$ and $n = 16000$ cells. The lognormal distributions we consider are $\sigma \in \{0.3, 0.6, 0.9, 1.2, 1.5\}$ and $\mu = \log(10) - \frac{\sigma^2}{2}$ such that $\mathbb{E}(V) = 10$.

MLE and therefore is more dependent on the cross sectional areas sample size. Depending on the actual value of μ or σ in some cases MLE yields a smaller bias and in other cases MM results in a smaller bias. For large values of σ it appears that the method of moments estimator is preferred when a large sample is used ($n = 16000$). It may not immediately be clear why the bias for estimates of μ is smaller when μ is large. This can be explained by the fact that a large μ corresponds to a small σ, since $\mu = \log(10) - \frac{\sigma^2}{2}$. In most cases the estimated bias does not behave linearly. It may be possible to use the estimated bias to find bias-corrected estimators, however, in the next section we highlight how the bias may be the result of model misspecification. We also show how these issues may be overcome.

9.3 Addressing bias and model misspecification

In the simulations in the previous section we observed that when using the estimator (9.1) for estimating the parameters μ and σ via MM or MLE works quite well. The resulting estimates are on average close to the true values of μ and σ. There does appear to be a bias in the estimates. While it may be possible to correct the estimates by estimating the bias and proposing correction coefficients some other approaches are highlighted. It is important to stress that the estimator for the sectional area distribution (7.10) approximately describes the sectional area distribution. One of the conditions for the estimator to precisely describe the sectional area distribution is that assumption 1 is satisfied. Hence, if this assumption is violated we cannot expect accurate results. The bias can then also be explained as model misspecification. Up till this point we have not quantified what it means for two sets or cells to be approximately similar. We now discuss two possible approaches of quantifying this similarity.

One possible approach in the case of Laguerre diagrams is to investigate whether cells are equally centroidal. How centroidal a cell L_i is depends on how close its generator point $\boldsymbol{x}_i$ is to its centroid $\boldsymbol{c}_i$. It appears that in hindsight, the residual which was used in Algorithm 1 seems inappropriate. Recall that this residual is used to determine the number of required iterations in the algorithm, it terminates when the Laguerre diagram is considered sufficiently centroidal. During simulations it was noticeable that depending on the chosen cell volume distribution and the number of cells the algorithm used vastly different amounts of iterations. On one end a Laguerre diagram with $n = 1000$ cells, each cell having volume 1 could be computed in 10 iterations. On the other end, a Laguerre diagram with $n = 16000$ cells which follow a lognormal volume distribution with $\sigma = 1.5$ could require well over 100 iterations. This could explain that the estimates of μ and σ were in some cases worse for diagrams with more cells.

Let us now show that the algorithm does not necessarily generate Laguerre diagrams with cells which are equally centroidal. Given a Laguerre diagram $\{L_i\}_{i=1}^n$ generated by $(\boldsymbol{x}_1, w_1), \ldots, (\boldsymbol{x}_1, w_1)$ on the domain $\Omega = [0, 1]^3$. It was computed by Algorithm 1, hence $m_i = |L_i|$ for all $i \in \{1, \ldots, n\}$. Each cell L_i has centroid $\boldsymbol{c}_i$. Now suppose we uniformly scale all distances in this diagram with some $\lambda > 0$. We obtain a Laguerre diagram $\{L_i^\lambda\}_{i=1}^n$ generated by $(\boldsymbol{x}_1^\lambda, w_1^\lambda), \ldots, (\boldsymbol{x}_1^\lambda, w_1^\lambda)$ on the domain $\Omega^\lambda = [0, \lambda]^3$. This could be computed by Algorithm 1 with m_i^λ's such that $m_i^\lambda = |L_i^\lambda|$ for all $i \in \{1, \ldots, n\}$. Each cell L_i^λ has centroid $\boldsymbol{c}_i^\lambda$.

Because of the uniform scaling L_i^λ is λ-similar to L_i. Using the properties of similarities we

obtain:

$$
\begin{aligned}
|L_i|\lambda^3 &= |L_i^\lambda| \\
m_i\lambda^3 &= m_i^\lambda \\
||\boldsymbol{x}_i - \boldsymbol{c}_i||\lambda &= ||\boldsymbol{x}_i^\lambda - \boldsymbol{c}_i^\lambda||.
\end{aligned}
$$

It seems sensible to consider both diagrams to be equally centroidal, after all, the distance of each generator point to the corresponding centroid is unchanged relative to the size of each cell. It turns out that the residual which was used does not consider these diagrams equally centroidal. Letting R denote the residual associated with $\{L_i\}_{i=1}^n$ and R_λ the residual associated with $\{L_i^\lambda\}_{i=1}^n$ we find:

$$
\begin{aligned}
R_\lambda &= \sqrt{\frac{\sum_{i=1}^n m_i^\lambda ||\boldsymbol{x}_i^\lambda - \boldsymbol{c}_i^\lambda||^2}{|\Omega^\lambda|\lambda}} \\
&= \sqrt{\frac{\sum_{i=1}^n m_i\lambda^3\lambda^2 ||\boldsymbol{x}_i - \boldsymbol{c}_i||^2}{\lambda^3|\Omega|\lambda}} \\
&= \sqrt{\lambda}\sqrt{\frac{\sum_{i=1}^n m_i ||\boldsymbol{x}_i - \boldsymbol{c}_i||^2}{|\Omega|}} \\
&= \sqrt{\lambda}R.
\end{aligned}
$$

This means for example that for $\lambda > 1$ that $R_\lambda > R$. A consequence of this is that Algorithm 1 cannot produce $\{L_i^\lambda\}_{i=1}^n$, instead it uses comparatively more iterations, resulting in cells that appear more round. It is possible to define a residual which do not suffer the from the problem illustrated above. Via a similar computation as above it can be shown that the following choices do not have this problem:

$$
\frac{1}{n}\sum_{i=1}^n \frac{||\boldsymbol{x}_i - \boldsymbol{c}_i||}{m_i^{\frac{1}{3}}} \quad \text{and} \quad \max_{i\in\{1,\ldots,n\}} \frac{||\boldsymbol{x}_i - \boldsymbol{c}_i||}{m_i^{\frac{1}{3}}}.
$$

We do stress that further simulations are necessary to investigate whether this is indeed a better choice for generating Laguerre diagrams. This potential issue with the residual highlights that better parameter estimates cannot only be obtained by adjusting the estimators it can also be achieved by considering a slightly different class of Laguerre diagrams. Because the proposed estimators work well on average we are looking for subtle changes that can be made which yield better results. Using a different stopping criterion in Algorithm 1 is a step that can be taken for subtly changing the way Laguerre diagrams are computed.

Another approach to quantifying similarity is to consider the distribution of the number cell faces. After all, sets are similar when they have the exact same shape. For a polyhedron one way of characterizing its shape is the number of faces. We know that Laguerre diagrams generated by Algorithm 1 do not have cells which all have the same number of faces. When assuming cells to be approximately similar this could for example mean that the distribution of the number of cell faces does not depend on the cell volume distribution. Let us therefore consider a simple simulation setting, comparing the distribution of the number of cell faces in two Laguerre diagrams with different cell volume distributions. We use $n = 16000$ cells and a lognormal cell volume distribution with $\sigma = 0.3$ and $\sigma = 1.5$. In both cases we use $\mu = -\frac{\sigma^2}{2}$. The results are in Figure 9.4. Clearly these distributions are very different. The average number

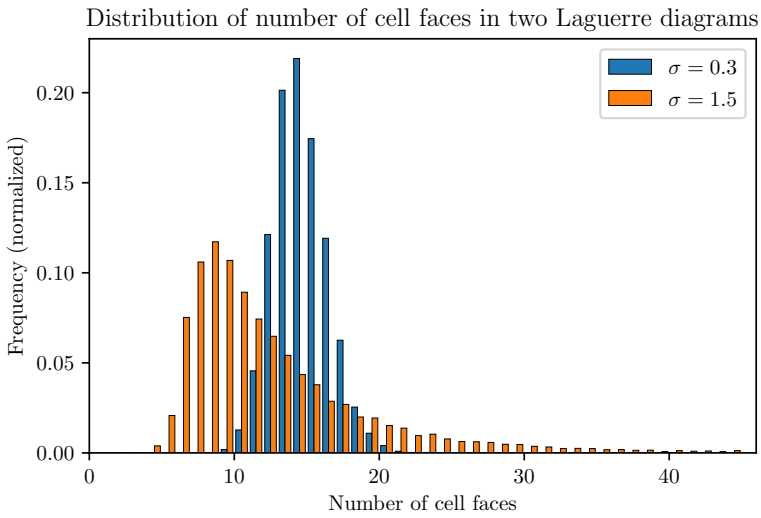


Figure 9.4: Distribution of the number of cell faces for two Laguerre diagrams with a lognormal cell volume distribution. The diagrams have $n = 16000$ cells. The lognormal distributions have parameters $\sigma = 0.3$ and $\sigma = 1.5$, and in both cases $\mu = -\frac{\sigma^2}{2}$.

of faces in the two Laguerre diagrams are still reasonable close to each other, for $\sigma = 0.3$ the average is 14.2 and for $\sigma = 1.5$ the average is 12.6.

Hence, the shape of the cells is affected by the choice of the cell volume distribution. Because the estimator does work well on average this raises the question in what sense the cells may still be considered approximately similar. At present we cannot provide a satisfactory answer to this question, this point will be further addressed in the discussion.

The only aspect of the estimator which is related to the shape of the cells is the distribution $F_{A^{(1)}}$. Note that we fixed this distribution beforehand. Especially in the case of the method of moments estimator it may be worthwhile to instead initially consider this distribution unknown. Recall that the method of moments estimators are obtained via equations (9.3), (9.4). This estimator only depends on estimates of $\mathbb{E}(A^{(1)})$ and $\mathbb{E}((A^{(1)})^2)$. Let $\alpha_1 = \mathbb{E}(A^{(1)})$ and $\alpha_2 = \mathbb{E}((A^{(1)})^2)$. During simulations the volume distribution is chosen and is therefore known. The sectional area distribution is observed and this allows for computing $\frac{1}{m}\sum_{j=1}^{m} a_j$ and $\frac{1}{m}\sum_{j=1}^{m} a_j^2$ from a sample of areas denoted by $a_1, \ldots, a_m$. By performing many simulations with different lognormal volume distributions it is possible to estimate which values of α_1, α_2 yields the most accurate method of moments estimator for all the considered volume distributions. In the worst-case scenario this procedure yields the same method of moments estimator as derived before, but it may result in an estimator which is more accurate. For maximum likelihood estimation such an approach is much more complicated. Estimating which distribution for $F_{A^{(1)}}$ performs the best over a range of volume distribution when using maximum likelihood estimation appears more difficult then estimating two constants.

9.4 Conclusion

In this chapter we considered the general estimator (7.10) of the cross sectional area distribution and derived its form in the case of a lognormal cell volume distribution. We also showed how this estimator may be used to estimate the parameters μ and σ of the lognormal cell volume distribution from a sample of cross sectional areas. Two approaches are presented, the method of moments and maximum likelihood estimation. In simulations we observed that both approaches work well on average, the average estimates of μ and σ are close to the actual values. The proposed estimators are biased, we provided some suggestions of how this bias may be addressed in future research.

10

Estimating the 3D grain volume distribution from real 2D EBSD data

Now that we have a way to estimate the cell volume distribution of a specific class of Laguerre diagrams from cross section areas we would like to know whether we may apply this method on real-world data. After all, while the proposed methods work quite well for the Laguerre diagrams we investigated, we would like to know whether we may apply these methods to real metals. Specifically, given measurements of the observed areas of grains in a metal we attempt to estimate μ and σ of the grain volume distribution of the metal microstructure which is assumed to follow a lognormal distribution. We remark that this chapter presents a single case study, as we consider data corresponding to a single sample of metal.

In this chapter we apply the derived estimators to real-world data. The data we use is known as Electron Backscatter Diffraction (EBSD) data. This is a technique based on Scanning Electron Microscope (SEM). This technique is used to extract various information from the surface of the material. EBSD data are obtained by directing a beam of electrons onto the material and by collecting the diffracted signal. For more information on EBSD we refer to [40]. EBSD is for example used for phase identification in materials and grain boundary characterisation. For our purposes it is especially useful that EBSD allows for the determination of the areas of the observed grains in a (2D) cross section. There also exist a 3D counterpart for EBSD known as 3D EBSD. By performing many parallel 2D sections of a material it is possible to acquire 3D information of the material, such as the grain volumes. This means that whenever 3D EBSD data is available of a material there is also 2D EBSD data because the 3D data is obtained via 2D EBSD data. It is important to stress that in practice 3D EBSD data are not always available. It is both expensive and time-consuming to obtain 3D EBSD data. This highlights that it is still relevant to estimate the grain volume distribution from 2D EBSD data. We would like to be able to use the proposed methodology to estimate the grain volume distribution when 3D EBSD data are not available.

Let us now discuss the specifics of the data. The data and relevant information about these data are kindly provided by Carola Celada-Casero, principal researcher at Tata steel. The data are relative to a single-phase steel called Interstitial Free (IF). This steel has a ferritic microstructure, which is soft and ductile. In terms of applications of such a steel its mechanical properties make it useful for forming applications, such as outer car panels. The 2D EBSD data

are obtained via a single section and they contain information relative to 1691 observed grains. For materials there is a commonly used frame of reference to indicate the direction of the cross section, also known as the sample reference frame. In the case of Laguerre diagrams we could simply say that we consider cross sections parallel to the xy-plane. In the sample reference frame the three axes are known as the rolling direction (RD), transverse direction (TD) and normal direction (ND). The 2D EBSD data originates from the surface of the steel which in this frame of reference is parallel to the ND-TD plane. It contains information of 1691 observed grains. From the 2D data the only variable we use are the areas of the observed grains. Let $a_1, \ldots, a_m$ denote this sample of areas from the 2D EBSD. In this specific example, we also have access to 3D EBSD data, which contains information of 9211 observed grains. From the 3D data we only take the volumes of these grains and we compare this data to the volume distribution that we estimate from the areas in the 2D EBSD data. Let $v_1, \ldots, v_n$ denote the sample of volumes from the 3D EBSD.

As before, we also assume that grain volume distribution in the material is approximately lognormally distributed. We use the estimator of the area distribution $\hat{f}_A(a|\mu, \sigma)$ (9.1) and we employ the Method of Moments estimators (MM) (9.5), (9.6) and the Maximum Likelihood Estimator (MLE) (9.10) to estimate the μ and σ which describe the (lognormal) grain volume distribution. Similarly to the simulations performed in chapter 9 the MLE is obtained by numerically maximizing the log-likelihood, and as the initial values for μ and σ we use the MM estimates. In this chapter we perform the following steps:

1. We compute the MM estimates: $\hat{\mu}_{\text{MM}}$, $\hat{\sigma}_{\text{MM}}$ and the MLE: $\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}_{\text{MLE}}$ from the 2D EBSD data. These are estimates of the parameters μ and σ of the grain volume distribution which we assume to be lognormal.

2. We generate a Laguerre diagram using the procedure described in section 4.4.1 with a lognormal cell volume distribution with parameters $\hat{\mu}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}$. Then we take a single cross section using the procedure described in section 4.4.2 and compare the obtained area distribution to the area distribution in the 2D EBSD data.

3. We also generate a Laguerre diagram using the procedure described in section 4.4.1 with the volume distribution obtained from the 3D EBSD data. This means that cells in this diagram have volumes $v_1, \ldots, v_n$. Then, the cross sectional area distribution of this diagram is also compared to the area distribution in the 2D EBSD data.

4. Finally, we take the lognormal distribution with the parameters $\hat{\mu}_{\text{MM}}, \hat{\sigma}_{\text{MM}}$ and the lognormal distribution with parameters $\hat{\mu}_{\text{MLE}}, \hat{\sigma}_{\text{MLE}}$ and we compare these to the volume distribution that is present in the 3D EBSD data.

Let us start with the first step, we fit $\hat{f}_A$ to the areas from the 2D EBSD data. The result of this fitting procedure can be seen in Figure 10.1. We see that the resulting distributions fit the data quite well, with the MLE providing a slightly better fit than the MM estimate. To verify that the numerical procedure of maximizing the log-likelihood terminated in a global maximum we make a contour plot of the log-likelihood shown in Figure 10.2. This picture suggests that we do indeed find a global maximum. Note that the contour levels in the figure (see colorbar) follow a highly non-linear scale to make the location of the maximum more clearly visible. The estimates of μ and σ are shown in Table 10.1.

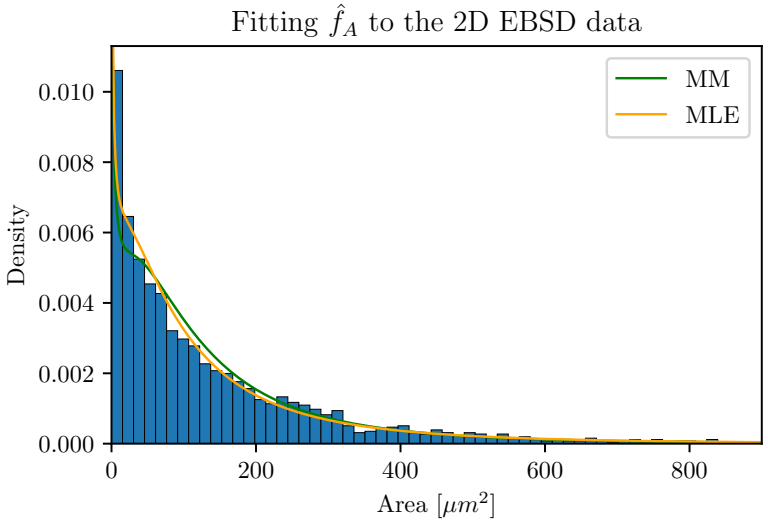


Figure 10.1: The result of fitting the distribution with density $\hat{f}_A(a|\mu, \sigma)$ (9.1) to the 2D EBSD data. The parameters μ and σ are estimated via the Method of Moments (MM) and via Maximum Likelihood Estimation (MLE).

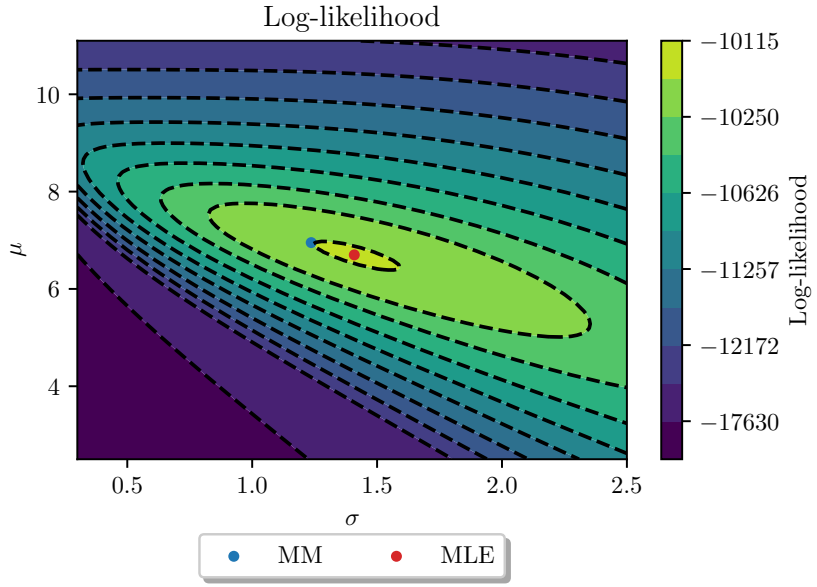


Figure 10.2: A contour plot of the log-likelihood for the sample of areas obtained from the 2D EBSD data set. The blue dot indicates the Method of Moments estimate (MM) of μ and σ which is used as the initial guess for numerically maximizing the log-likelihood. The red dot indicates the Maximum Likelihood Estimate (MLE).

Method	$\hat{\mu}$	$\hat{\sigma}$
MM	6.952	1.236
MLE	6.700	1.409
3D EBSD fit	6.871	1.666

Table 10.1: Estimated parameters of the grain volume distribution, which is assumed to be lognormal. The parameters in the first two rows are obtained by fitting $\hat{f}_A$ to the 2D EBSD data. These parameters are estimated via MM and MLE. The third row contains the estimates of μ and σ obtained by fitting a lognormal distribution to the 3D EBSD data.

The next step is to generate a Laguerre diagram with a lognormal distribution with parameters $\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}_{\text{MLE}}$. One aspect which is not covered in detail in this thesis is the estimation of the number of cells n from a cross section. A seemingly reasonable approach is to estimate n by an estimate of the volume of the Laguerre diagram, divided by an estimate of the average cell volume. As we considered only cube-shaped Laguerre diagrams and horizontal cross sections we may take the following approach:

$$\hat{n} = \frac{\left(\sum_{j=1}^{m} a_j\right)^{\frac{3}{2}}}{\exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right)}. \tag{10.1}$$

If we assume the data to come from a horizontal cross section of a cube-shaped Laguerre diagram then the sum of the observed areas is equal to the area of one of the sides of the cube. Hence $\left(\sum_{j=1}^{m} a_j\right)^{\frac{3}{2}}$ is equal to the volume of the cube. Then we divide the volume of the cube by the (estimated) average grain volume to get an estimate of the number of cells. This approach yields $\hat{n} = 55762$ (MM) and $\hat{n} = 57132$ (MLE). As both estimates are fairly close it does not matter too much which one is used in this case. We generate a Laguerre diagram with $n = 55762$ cells, with a lognormal cell volume distribution with parameters $\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}_{\text{MLE}}$. Then, we take a single cross section which happens to contain 1673 cells, which is quite close to the 1691 grains in the 2D EBSD data. Because we have seen in chapter 9 that $\hat{f}_A$ provides a good description of what the area distribution looks like for a given lognormal volume distribution we expect the resulting area distribution to be similar to the 2D EBSD data. The results are visualized in Figure 10.3. The sample of cross sectional areas of this Laguerre diagram with lognormal cell volume distribution ($\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}_{\text{MLE}}$) is labelled with "simulated lognormal" in Figure 10.3. We observe that these two distributions appear quite similar.

Figure 10.3 contains another histogram. We also generate a Laguerre diagram using the volumes in the 3D EBSD data set as the cell volume distribution. This means that cells in this diagram have volumes $v_1, \ldots, v_n$. Recall that the 3D EBSD data set contains information of 9211 observed grains. As a result this Laguerre diagram is much smaller compared to the previously discussed diagram with the estimated lognormal distribution. A consequence is that cross sections of this Laguerre diagram yield much smaller samples of sectional areas. Therefore, we take three cross sections of the same diagram, which are sufficiently spaced apart. Given the domain of the diagram $\Omega = [0, l]^3$, we take sections at $z = \frac{l}{3}$, $z = \frac{2l}{3}$ and $z = l$. This yields samples of sectional areas of sizes 530, 494 and 507. The combined sample is shown in Figure 10.3 and is labelled with "simulated 3D EBSD". This distribution also appears reasonably close to the 2D EBSD data.

Finally, we take a look at the estimated grain volume distributions. In Figure 10.4 we show the distribution of the volumes that is taken from the 3D EBSD data set along with the pdfs of

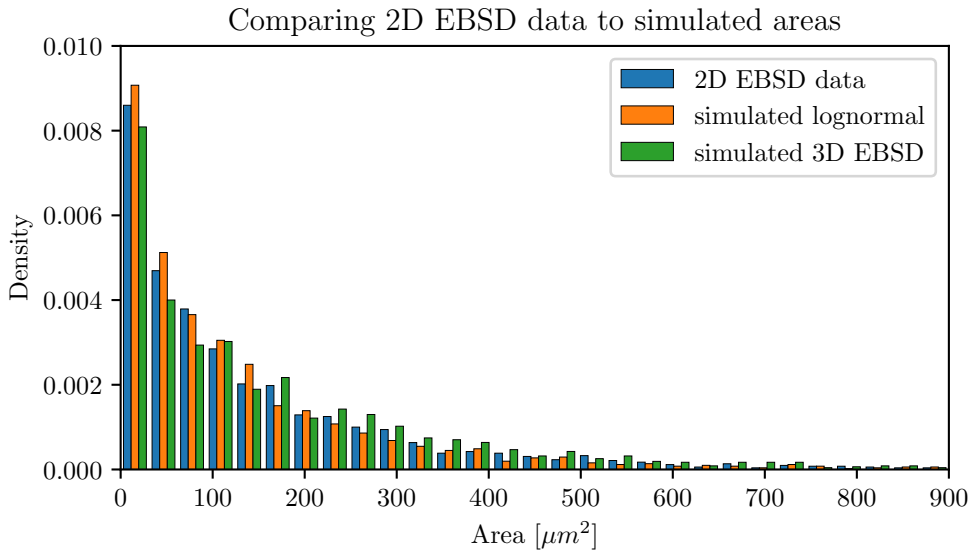


Figure 10.3: A comparison between the area distribution in the 2D EBSD data to the sectional area distributions of two Laguerre diagrams. Simulated lognormal: the sectional area distribution of a Laguerre diagram with a lognormal cell volume distribution with parameters $\hat{\mu}_{\text{MLE}}$, $\hat{\sigma}_{\text{MLE}}$. Simulated 3D EBSD: the sectional area distribution of a Laguerre diagram with the 3D EBSD data as volume distribution.

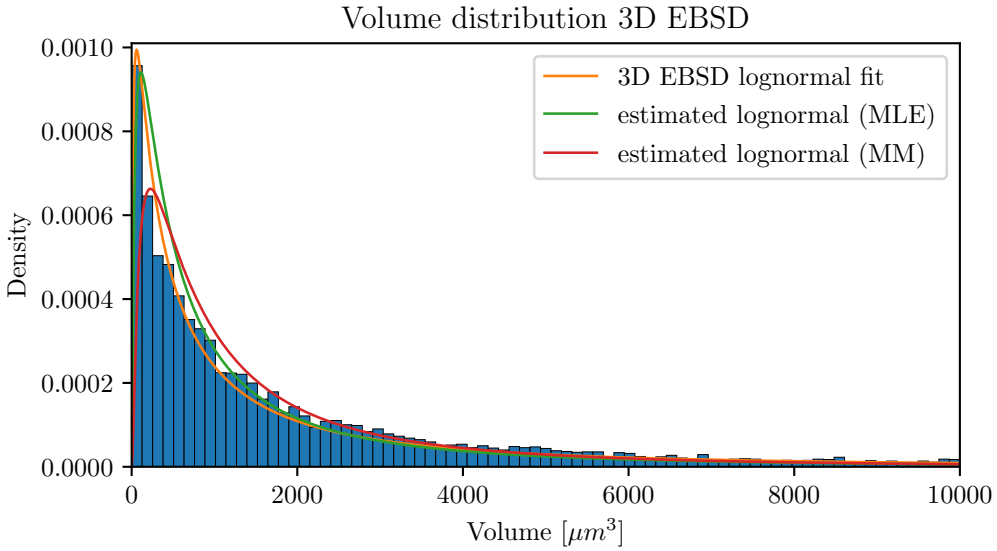


Figure 10.4: A comparison of the volume distribution in the 3D EBSD data to three lognormal distributions. These distributions visualize the lognormal distributions with parameters as in Table 10.1.

three lognormal distributions. The figure contains the lognormal distribution which was fitted directly to the 3D EBSD data via MLE. The figure also contains the lognormal distributions which were estimated from the 2D EBSD data. Recall that the parameters of these lognormal distributions are shown in Table 10.1.

An important question is whether the estimates are good enough to be used in practice. We also want to use the methodology when 3D EBSD data is not available. We cannot expect to estimate a lognormal distribution from the 2D EBSD data which provides a better fit to the 3D EBSD data than the lognormal distribution which is fitted directly to the 3D EBSD data. In that sense the estimates using the proposed method are reasonably close, especially the parameters obtained by MLE. An important note is that the 3D EBSD data is not truly lognormally distributed. There is a simple observation which shows that care should be taken. Say we are interested in the average grain volume in the material. Since we have estimates of μ and σ we may compute an estimate of the average volume via the formula: $\exp(\mu + \frac{\sigma^2}{2})$. We can compare these estimates to the empirical average of the 3D EBSD data: $\frac{1}{n}\sum_{i=1}^{n} v_i$ where $v_1, \ldots, v_n$ are the volumes taken from the 3D EBSD data. The results are in Table 10.2.

Method	Average volume estimate $[\mu m^3]$
MM	2244.2
MLE	2190.4
3D EBSD fit	3860.2
3D EBSD empirical average	2793.1

Table 10.2: Estimates of the average grain volume. The first three rows contain estimates of the average grain volume by using the estimated parameters of the lognormal distribution as in Table 10.1. The average is obtained by computing $\exp(\hat{\mu} + \frac{\hat{\sigma}^2}{2})$. The fourth row contains the empirical average of the volumes in the 3D EBSD data set.

We clearly see that if we take the lognormal distribution that is fitted to the 3D EBSD data as the true volume distribution, we would severely overestimate the average grain volume. This shows that if we would like to obtain a more accurate estimate of the volume distribution we either need to use another parametric distribution which provides a better fit, or we need a way to estimate the volume distribution in a nonparametric fashion.

11

Discussion

With this chapter we conclude the thesis. We begin with a summary of our findings and mention the main results. Then, suggestions are made for directions which may be taken in future research.

11.1 Summary

Motivated by the modelling of metals microstructures the Laguerre diagram was introduced. After providing the necessary definitions of this diagram and discussing its properties we described the problem which we set out to solve in this thesis. We consider the inverse problem of the determination of the cell volume distribution in a 3D Laguerre diagram from the distribution of areas in a 2D cross section of such a diagram. We also described the related direct problem of finding the distribution of cross sectional areas for a Laguerre diagram with a known cell volume distribution. Via some examples it was shown that in these problems we cannot consider arbitrary Laguerre diagrams.

In chapter 4 Algorithm 1 was discussed for computing Laguerre diagrams. It is particularly attractive because it allows for computing a Laguerre diagram with a chosen cell volume distribution. The direct and inverse problem consider the relationship between two distributions, namely the cell volume distribution and the cross sectional area distribution. Therefore, this algorithm allows for controlling one side of the equation. After describing the algorithm, the parameters required by the algorithm were considered. In particular the influence of the initial generator points, because these have a large impact on the resulting Laguerre diagram. The eventual choice for the initial generators was motivated using visualizations and hypothesis tests. We also proposed a permutation method that proved to speed up the gradient method used by Algorithm 1 for Laguerre diagrams with moderate variance in the volume distribution. The chapter was concluded with a precise description of how Laguerre diagrams are computed and how cross sections are taken in all further simulations. This means that for our problem we restricted ourselves to this specific class of Laguerre diagrams.

In chapter 5 we investigated the role of the weights of a Laguerre diagram. In the case of generating a Laguerre diagram via a close packing of spheres there is a clear interpretation of the weights being the square root of the sphere radii. For the Laguerre diagrams we consider this

interpretation cannot be used and does not add knowledge for our problem. We did determine another relationship with sphere packings, since we explained how Algorithm 1 may be used to generate a sphere packing. In chapter 6 we briefly investigated the properties of 2D sectional Laguerre diagrams which was concluded by noting that these diagrams do not seem to have properties which may help in solving the inverse problem.

In chapter 7 we derived an estimator for the pdf of the sectional area distribution. The estimator provides a clear relationship between the distribution of sectional areas and the distribution of the cell volumes. When assuming a specific parametric distribution for the cell volume distribution we showed how to use the estimator for estimating the parameters of this cell volume distribution. The use of the estimator was motivated in the following chapter, chapter 8. We considered two stereological problems, Wicksell's corpuscle problem and one of its possible generalizations. In both problems an analytical expression for the sectional area distribution is available and it corresponds to the proposed estimator for the sectional area distribution. The estimator does come with an assumption, by using it we assume that cells in a Laguerre diagram are approximately similar (assumption 1). Given that the problem is motivated by a materials science application we considered Laguerre-Voronoi diagrams with a lognormal cell volume distribution, which is commonly used in this field. We described the form of the estimator for the sectional area distribution in the case of a lognormal cell volume distribution. We also showed how this estimator may be used to estimate the parameters μ and σ of the lognormal cell volume distribution from a sample of cross sectional areas. Two approaches are presented, the method of moments and maximum likelihood estimation. In simulations we observed that both approaches work well on average, the average estimates of μ and σ are close to the actual values. For small values of σ the maximum likelihood estimator is preferred. For large values of σ the method of moments estimator is preferred, provided that the sample of cross sectional areas is sufficiently large. The proposed estimators for μ and σ are however biased, we provided some suggestions for how this bias may be addressed in future research.

In the final chapter, chapter 10, the proposed estimator for the sectional area distribution is applied to real-world data. The grain volume distribution in a steel microstructure is estimated from 2D image data. Because the grain volume distribution was known in the form of 3D EBSD data the estimated volume distribution could be compared to the actual volume distribution. A lognormal distribution was assumed for the grain volume distribution in this steel microstructure, which was reasonable, but there was some level of misspecification. Keeping this in mind, the volume distributions estimated from the 2D image data provided a decent description of the actual volume distribution.

Whether the inverse problem which we set out to solve may be considered solved depends on the required accuracy of the solution. Of course, we only extensively considered the case of lognormal volume distributions, hence it cannot be considered solved for arbitrary volume distributions. In the lognormal case the proposed estimators for μ and σ describing the volume distribution work quite well, but there is a bias and therefore there is still some room for improvement. Depending on the accuracy required in an application of these methods, the provided simulation results may be used to judge whether these estimators are sufficiently accurate.

11.2 Future research

Let us start with a direction for future research which we already mentioned. The proposed estimators for the parameters μ and σ of the lognormal cell volume distribution are biased, in future research the suggestions which were made for addressing this bias in section 9.3 could be investigated. We noted that the bias may be caused by the fact that assumption 1 is violated. The assumption that all cells approximately have the same shape is somewhat inconvenient given that we know that the cells do in fact have different shapes. The provided suggestions for addressing bias cannot entirely overcome this difficulty, but they may still improve the accuracy of the estimators for μ and σ. In future research the development of a method could be pursued which does not require assumption 1. It is likely that some assumption is still necessary but ideally this is a much weaker assumption. There is another assumption which was made for the estimator of the sectional area distribution, namely that cells of a Laguerre diagram are convex. This assumption was only mentioned when we considered a generalization of Wicksell's problem. We did not further discuss this assumption because it is always satisfied for cells of Laguerre diagrams. There are however materials microstructures with non-convex grains, it is important to note that the proposed methods may not work well when considering such microstructures.

Because models other than Laguerre diagrams may be considered for the modelling of materials microstructures with convex grains it may also be investigated how well the proposed method works for some other models. Especially in the case of a model for which assumption 1 is in fact a reasonable assumption the estimator of the sectional area distribution could be useful. Another suggestion for future research is the derivation of a method which estimates the cell volume distribution of a Laguerre diagram in a non-parametric setting. In chapter 8 we mentioned some relevant literature which may provide a starting point when pursuing this research direction.

Bibliography

[1] E. O. Hall, "The deformation and ageing of mild steel: III discussion of results," *Proceedings of the Physical Society. Section B*, vol. 64, 9 1951.

[2] N. J. Petch, "The cleavage strength of polycrystals," *J. Iron Steel Inst*, vol. 174, pp. 25 – 28, 1953.

[3] W. Li, M. Vittorietti, G. Jongbloed, and J. Sietsma, "The combined influence of grain size distribution and dislocation density on hardness of interstitial free steel," *Journal of Materials Science & Technology*, vol. 45, pp. 35–43, 5 2020.

[4] S. Bargmann, B. Klusemann, J. Markmann, J. E. Schnabel, K. Schneider, C. Soyarslan, and J. Wilmers, "Generation of 3d representative volume elements for heterogeneous materials: A review," *Progress in Materials Science*, vol. 96, pp. 322–384, 7 2018.

[5] J. Ohser and F. Mücklich, *Statistical Analysis of Microstructures in Materials Science*. John Wiley & Sons, Inc., 2000.

[6] C. Lautensack, "Fitting three-dimensional laguerre tessellations to foam structures," *Journal of Applied Statistics*, vol. 35, 9 2008.

[7] Z. Fan, Y. Wu, X. Zhao, and Y. Lu, "Simulation of polycrystalline structure with voronoi diagram in laguerre geometry based on random closed packing of spheres," *Computational Materials Science*, vol. 29, pp. 301–308, 3 2004.

[8] R. Quey and L. Renversade, "Optimal polyhedral description of 3d polycrystals: Method and application to statistical and synchrotron x-ray diffraction data," *Computer Methods in Applied Mechanics and Engineering*, vol. 330, pp. 308–333, 3 2018.

[9] A. Okabe, B. Boots, K. Sugihara, S. N. Chiu, and D. G. Kendall, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Inc., 7 2000.

[10] F. Aurenhammer, R. Klein, and D.-T. Lee, *Voronoi Diagrams and Delaunay Triangulations*. World Scientific, 8 2013.

[11] D. P. Bourne, P. J. J. Kok, S. M. Roper, and W. D. T. Spanjer, "Laguerre tessellations and polycrystalline microstructures: a fast algorithm for generating grains of given volumes," *Philosophical Magazine*, vol. 100, 11 2020.

[12] M. Vittorietti, P. J. Kok, J. Sietsma, and G. Jongbloed, "Accurate representation of the distributions of the 3d poisson-voronoi typical cell geometrical features," *Computational Materials Science*, vol. 166, 8 2019.

[13] A. Baddeley and E. B. V. Jensen, *Stereology for Statisticians*. Chapman and Hall/CRC, 2004.

[14] G. Jongbloed, *Three Statistical Inverse Problems: Estimators-algorithms-asymptotics*. Unpublished Ph.D. dissertation, Delft University of Technology, 1995.

[15] S. D. Wicksell, "The corpuscle problem. a mathematical study of a biometric problem," *Biometrika*, vol. 17, 1925.

[16] U. Hahn and U. Lorz, "On the precision of some stereological estimators for the model parameter of the spatial poisson-voronoi tessellation," *Acta Stereologica*, vol. 13, 1994.

[17] D. Depriester and R. Kubler, "Radical voronoï tessellation from random pack of polydisperse spheres: Prediction of the cells' size distribution," *Computer-Aided Design*, vol. 107, 2 2019.

[18] S. A. Saltikov, "The determination of the size distribution of particles in an opaque material from a measurement of the size distribution of their sections," in *Stereology* (H. Elias, ed.), pp. 163–173, Springer Berlin Heidelberg, 1967.

[19] Y. Gulbin, "On estimation and hypothesis testing of the grain size distribution by the saltykov method," *Image Analysis & Stereology*, vol. 27, 5 2008.

[20] M. A. Lopez-Sanchez and S. Llana-Fúnez, "An extension of the saltykov method to quantify 3d grain size distributions in mylonites," *Journal of Structural Geology*, vol. 93, 12 2016.

[21] F. Aurenhammer, F. Hoffmann, and B. Aronov, "Minkowski-type theorems and least-squares clustering," *Algorithmica*, vol. 20, 1 1998.

[22] Q. Mérigot, "A multiscale approach to optimal transport," *Computer Graphics Forum*, vol. 30, 8 2011.

[23] J. Kuhn, M. Schneider, P. Sonnweber-Ribic, and T. Böhlke, "Fast methods for computing centroidal laguerre tessellations for prescribed volume fractions with applications to microstructure generation of polycrystalline materials," *Computer Methods in Applied Mechanics and Engineering*, vol. 369, 9 2020.

[24] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, vol. 8, 1988.

[25] Y. Malitsky and K. Mishchenko, "Adaptive gradient descent without descent," *arXiv preprint arXiv:1910.09529*, 2019.

[26] F. W. Scholz and M. A. Stephens, "K-sample anderson–darling tests," *Journal of the American Statistical Association*, vol. 82, 9 1987.

[27] C. Redenbach, "Microstructure models for cellular materials," *Computational Materials Science*, vol. 44, pp. 1397–1407, 2 2009.

[28] Q. Duan, D. P. Kroese, T. Brereton, A. Spettl, and V. Schmidt, "Inverting laguerre tessellations," *The Computer Journal*, vol. 57, 9 2014.

[29] K. Borsuk and W. Szmielew, *Foundations of Geometry: Euclidean, Bolyai-Lobachevskian, and Projective Geometry*. Dover Publications, Inc., 2018.

[30] H. S. M. Coxeter, *Introduction to Geometry*. John Wiley & Sons, Inc., 1969.

[31] E. V. Huntington, "Frequency distribution of product and quotient," *The Annals of Mathematical Statistics*, vol. 10, 6 1939.

[32] J. L. Paul, "Distribution curves of sectional areas through some families of convex particles," *Journal of Microscopy*, vol. 122, 5 1981.

[33] F. C. Hull and W. J. Houk, "Statistical grain structure studies: Plane distribution curves of regular polyhedrons," *JOM*, vol. 5, 4 1953.

[34] L. A. Santaló and M. Kac, *Integral Geometry and Geometric Probability*. Cambridge University Press, 10 2004.

[35] J. Kisel'ák and G. Baluchová, "On particle-size distribution of convex similar bodies in $\mathbb{R}^3$," *Journal of Mathematical Imaging and Vision*, vol. 63, 1 2021.

[36] A. D. Poularikas, ed., *Transforms and Applications Handbook*. CRC Press, 9 2018.

[37] D. Belomestny and A. Goldenshluger, "Nonparametric density estimation from observations with multiplicative measurement errors," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 56, 2 2020.

[38] S. B. Miguel, F. Comte, and J. Johannes, "Spectral cut-off regularisation for density estimation under multiplicative measurement errors," *Electronic Journal of Statistics*, vol. 15, 1 2021.

[39] L. Wasserman, *All of Nonparametric Statistics*. Berlin, Heidelberg: Springer-Verlag, 2006.

[40] V. Randle, "Electron backscatter diffraction: Strategies for reliable data acquisition and processing," *Materials Characterization*, vol. 60, 9 2009.

[41] C. H. Rycroft, "Voro++: A three-dimensional voronoi cell library in c++," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 19, 12 2009.

[42] J. Jordan, "pyvoro." https://github.com/joe-jordan/pyvoro. (accessed: 11.01.2021).

[43] R. M. Murray, Z. Li, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC Press, 2017.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 3 2004.

A

Appendix: Implementation details

In this section some details of the implementation in Python are discussed. The code is available as a Python package on Github: https://github.com/thomasvdj/vorostereology. For code examples see the Github page. Laguerre diagrams are computed by interfacing with the C++ library Voro++ [41], this library also provides functions to compute cell-based statistics such as volumes, centroids or areas of cell faces. In order to be able to use Voro++ in Python it was interfaced using Cython, and as a starting point the codebase of the Python library pyvoro [42] was used to achieve this. The features that were required but were not present in pyvoro were added to the existing codebase and these are discussed in this section.

A.1 Computing Laguerre diagrams

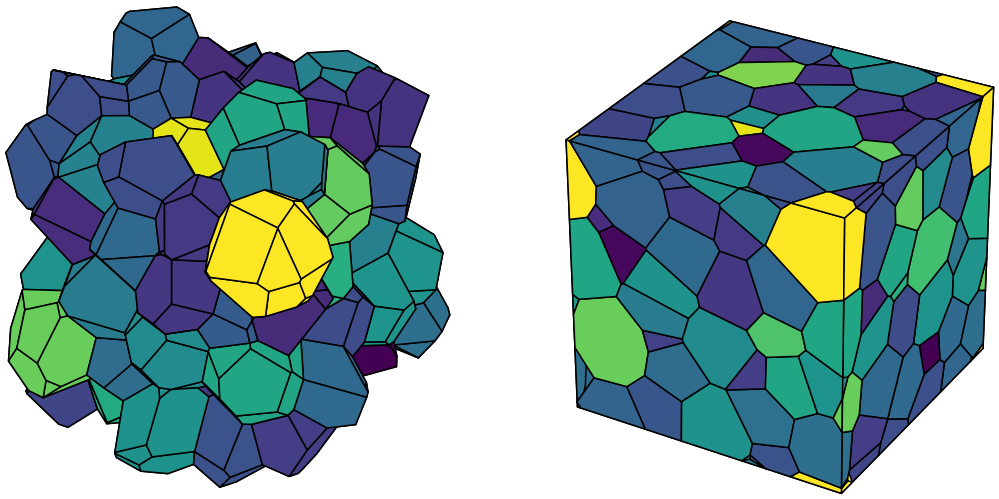
The computation of Laguerre diagrams is mostly handled by Voro++ but in the periodic case some additional work is required. This is due to the fact that the implementation of Voro++ does not correspond to the mathematical definition a periodic Laguerre diagram. Instead, it computes each cell in the frame of reference of its generator point as follows:

$$\tilde{L}_i = \{\boldsymbol{x} \in \mathbb{R}^3 : ||\boldsymbol{x}-\boldsymbol{x}_i||^2_{\text{per}} - w_i \leq ||\boldsymbol{x}-\boldsymbol{x}_j||^2_{\text{per}} - w_j \quad \text{and} \quad ||\boldsymbol{x}-\boldsymbol{x}_i||_{\text{per}} = ||\boldsymbol{x}-\boldsymbol{x}_i|| \quad \forall j \in \{1,\dots,n\}\}. \tag{A.1}$$

Which is slightly different from:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x}-\boldsymbol{x}_i||^2_{\text{per}} - w_i \leq ||\boldsymbol{x}-\boldsymbol{x}_j||^2_{\text{per}} - w_j \quad \forall j \in \{1,\dots,n\}\}. \tag{A.2}$$

Recall that for periodic Laguerre diagrams cells may consist of multiple fragments. Since a cell is computed from the frame of reference of its generator point $\boldsymbol{x}_i$, we may obtain all fragments by considering multiple frames of reference: $\boldsymbol{x}_i + (il_1, jl_2, kl_3)$ for $i, j, k \in \mathbb{Z}$ (Recall Definition 2). For 3D Laguerre diagrams at most 8 frames of reference need to be considered for a single cell, because a single cell may consist of 8 fragments. In some sense L_i can be obtained by imposing the periodic norm on $\tilde{L}_i$. Therefore, we obtain the Laguerre diagram as desired by computing all cells in the required frames of reference. The next step is to intersect these cells $\tilde{L}_i$ with Ω, such that the fragments are obtained as desired. The intersection with Ω is achieved via the function `voro::voronoicell_neighbor::nplane` which allow for intersecting a cell with a half-space. The resulting Laguerre diagram which satisfies the mathematical definition (A.2) is shown in Figure A.1 (b).



(a) (b)

Figure A.1: Comparison of a periodic Laguerre diagram as computed by Voro++ (A.1): (a) and as per the mathematical definition (A.2): (b).

A.2 Computing a cross section of a Laguerre diagram

When computing a cross section of a Laguerre diagram, we use functionalities that are part of the Voro++ library. Once a Laguerre diagram is computed we can then iterate over all cells in the diagram and use the `voro::voronoicell_neighbor::plane_intersects` method to find out whether the cross section plane of choice intersects with the cell. If the plane does intersect with the cell the `voro::voronoicell_neighbor::nplane` method is used to cut the cell with the cross section plane. It is important to keep in mind that we can then no longer use these cells to compute their volume, as part of the cell has been cut off by the cross section plane (because the cross section plane is interpreted as a half-space). We then use the useful property of the `nplane` method that we can keep track of which face of a cell is caused by which cell. For a Laguerre diagram with n cells this means that each face of a cell is associated with a number in $\{1, \ldots, n\}$ (called the plane ID) which indicates which cell is adjacent to that face. By using the `nplane` method with the plane ID $n+1$ it is possible to identify which faces of the cells were caused by the cross section plane and therefore are in the cross section plane. Finally, the `voro::voronoicell_neighbor::vertices` method is used to obtain the vertices of the faces which are in the cross section and the `voro::voronoicell_neighbor::face_areas` method to compute the areas of these faces.

A.3 Transforming vertices of cross section cells into 2D

Now that the vertices of the cells in the cross section have been computed we would like to transform the plane that was used for the cross section into 2D, in order to be able to provide visualizations of the cross sections. Mathematically, given points (x, y, z) in the plane P:

$$P = \{(x, y, z) \in \mathbb{R}^3 : a(x - x_0) + b(y - y_0) + c(z - z_0) = 0\}.$$

Now we would like to transform the points in this plane into the xy-plane by means of a translation and a rotation. This can be achieved by computing $R(\boldsymbol{x} - \boldsymbol{x}_0)$ for $\boldsymbol{x} \in P$ where $\boldsymbol{x}_0 = (x_0, y_0, z_0)^T$. The rotation matrix R that is required is the matrix for a rotation around an axis $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^T$ (which is a unit vector) with angle θ which is given by:

$$R = \begin{pmatrix} \omega_1^2(1-\cos\theta) + \cos\theta & \omega_1\omega_2(1-\cos\theta) - \omega_3\sin\theta & \omega_1\omega_3(1-\cos\theta) + \omega_2\sin\theta \\ \omega_1\omega_2(1-\cos\theta) + \omega_3\sin\theta & \omega_2^2(1-\cos\theta) + \cos\theta & \omega_2\omega_3(1-\cos\theta) - \omega_1\sin\theta \\ \omega_1\omega_3(1-\cos\theta) - \omega_2\sin\theta & \omega_2\omega_3(1-\cos\theta) + \omega_1\sin\theta & \omega_3^2(1-\cos\theta) + \cos\theta \end{pmatrix}. \tag{A.3}$$

A derivation of (A.3) using a formula known as Rodrigues' formula can for example be found in the proof of Proposition 2.5. in [43]. For every point $\boldsymbol{x}$ in P the point $\boldsymbol{x}_0$ may be subtracted such that a plane is obtained which passes through the origin. The rotation axis $\boldsymbol{\omega}$ can be found by using the fact that $\boldsymbol{a} = (a, b, c)^T$ is a normal vector of P. The rotation axis of interest is a unit vector which is orthogonal to both $\boldsymbol{a}$ and $\boldsymbol{e}_z = (0, 0, 1)^T$, hence we may take:

$$\boldsymbol{\omega} = \frac{\boldsymbol{a} \times \boldsymbol{e}_z}{||\boldsymbol{a} \times \boldsymbol{e}_z||} = \frac{1}{\sqrt{a^2 + b^2}}(b, -a, 0)^T.$$

This means that:

$$\omega_1 = \frac{b}{\sqrt{a^2 + b^2}}, \quad \omega_2 = -\frac{a}{\sqrt{a^2 + b^2}}, \quad \omega_3 = 0. \tag{A.4}$$

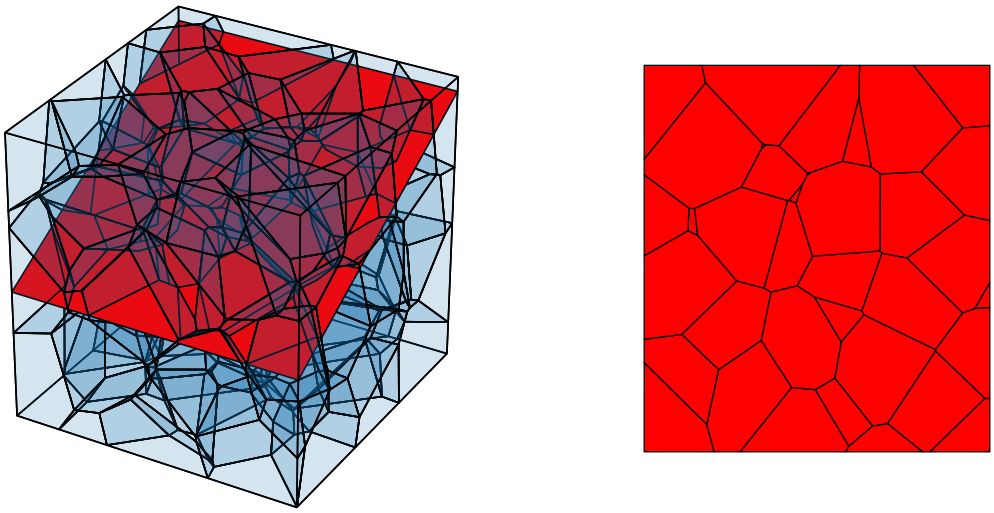
Since $\omega_3 = 0$ equation (A.3) simplifies to:

$$R = \begin{pmatrix} \omega_1^2(1-\cos\theta) + \cos\theta & \omega_1\omega_2(1-\cos\theta) & \omega_2\sin\theta \\ \omega_1\omega_2(1-\cos\theta) & \omega_2^2(1-\cos\theta) + \cos\theta & -\omega_1\sin\theta \\ -\omega_2\sin\theta & \omega_1\sin\theta & \cos\theta \end{pmatrix}. \tag{A.5}$$

The angle of rotation θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{e}_z$ and it can easily be shown that $\cos\theta$ and $\sin\theta$ may therefore be computed as:

$$\cos\theta = \frac{c}{\sqrt{a^2 + b^2 + c^2}}, \quad \sin\theta = \sqrt{\frac{a^2 + b^2}{a^2 + b^2 + c^2}}. \tag{A.6}$$

In the implementation the expressions in (A.4) and (A.6) are plugged into the expression for rotation matrix R as in (A.5). We now conclude that a vertex $\boldsymbol{x}$ of a cross section cell in P is transformed into the xy-plane by computing: $R(\boldsymbol{x} - \boldsymbol{x}_0)$. After this transformation the z-component of every vertex may be discarded such that the vertices are now in $\mathbb{R}^2$. Keep in mind that throughout these computations we assumed that we are not in the case that P is already parallel to the xy-plane: $a = 0$, $b = 0$, $c \neq 0$, in that case all points $\boldsymbol{x}$ in P can be transformed into the xy-plane by the translation: $\boldsymbol{x} - \boldsymbol{x}_0$. An example visualizing the computations discussed in this section and in A.2 are visualized in Figure A.2.



(a) (b)

Figure A.2: (a): An example of a cross section visualized in red, in a transparent Laguerre diagram. This cross section is computed as described in section A.2. (b) The same cross section after transforming it into $\mathbb{R}^2$ using the method as described in section A.3.

B

Appendix: Algorithm 1 as a sphere packing algorithm

When using Algorithm 1 with a small value of ε we generate a centroidal Laguerre diagram which has cells that are quite round, considering they are polyhedrons. We may compute the largest inscribed sphere for each cell of the Laguerre diagram to generate a sphere packing. In this section we show how to find the largest inscribed sphere for a single cell via a small linear programming problem. By solving this problem for each cell we find spheres which are guaranteed not to overlap. Because algorithm 1 also allows periodic boundary conditions we can also generate a sphere packing with periodic boundary conditions. During simulations we tended to observe packing densities between 0.5 and 0.6 (packing density $= \frac{\text{sum of sphere volumes}}{\text{volume of domain}}$). We now simply state how to compute these spheres, we do not analyze whether this is a good way to generate a sphere packing as this is out of scope.

Finding the largest inscribed sphere inside of a polyhedron can be described as a linear programming problem (LP problem). A simple derivation of the LP problem which we state in a moment can be found in [44, p. 148, 149]. Let P be a polyhedron with m faces which is given by:

$$P = \left\{ \boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{a}_j^T \boldsymbol{x} \leq b_j \text{ for } j = 1, \ldots, m \right\}.$$

Where $\boldsymbol{a}_j \in \mathbb{R}^3$, $b_j \in \mathbb{R}$ for $j = 1, \ldots, m$. Then, the largest inscribed sphere can be found by solving the following LP problem:

$$\begin{array}{ll} \text{maximize:} & r \\ \text{subject to:} & \boldsymbol{a}_j^T (x, y, z)^T + r||\boldsymbol{a}_j|| \leq b_j, \quad j = 1, \ldots, m \end{array}. \tag{B.1}$$

The solution of this problem is a sphere with center $(x, y, z) \in \mathbb{R}^3$ and radius r, which is inscribed in P. This is a small LP problem with 4 variables (x, y, z and r) and the number of constraints is equal to the number of faces of the polyhedron. We now show how we may use this to find the largest inscribed sphere inside of a cell of a Laguerre diagram.

Suppose we have a (approximately) centroidal Laguerre diagram with domain $\Omega = [0, l_1] \times [0, l_2] \times [0, l_3]$, with generator points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \Omega$ and corresponding weights: $w_1, \ldots, w_n \in \mathbb{R}$. A centroidal Laguerre diagram does not have empty cells, and we need non-empty cells if we want to be able to inscribe spheres. Let us now consider what (B.1) looks like when considering

a cell L_i (for some $i \in \{1, \ldots, n\}$) of a Laguerre diagram with or without periodic boundary conditions.

B.1 A Laguerre diagram without periodic boundary conditions

Recall that L_i is given by:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x} - \boldsymbol{x}_i||^2 - w_i \leq ||\boldsymbol{x} - \boldsymbol{x}_j||^2 - w_j \ \ \forall j \in \{1, \ldots, n\}\}.$$

While we do not know the neighbor information of the cell a priori, once a Laguerre diagram is computed we do have this information. For a cell L_i we define $N_i \subset \{1, \ldots, n\}$ as the index set of neighbors of L_i, this means that if $j \in N_i$ then L_j is a neighbor of L_i. We may then simply consider:

$$L_i = \{\boldsymbol{x} \in \Omega : \ ||\boldsymbol{x} - \boldsymbol{x}_i||^2 - w_i \leq ||\boldsymbol{x} - \boldsymbol{x}_j||^2 - w_j \ \ \forall j \in N_i\}.$$

Via straightforward computation we see that:

$$\begin{aligned} ||\boldsymbol{x} - \boldsymbol{x}_i||^2 - w_i \leq ||\boldsymbol{x} - \boldsymbol{x}_j||^2 - w_j &\iff \boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{x}_i^T\boldsymbol{x} + \boldsymbol{x}_i^T\boldsymbol{x}_i - w_i \leq \boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{x}_j^T\boldsymbol{x} + \boldsymbol{x}_j^T\boldsymbol{x}_j - w_j \\ &\iff 2(\boldsymbol{x}_j - \boldsymbol{x}_i)^T\boldsymbol{x} \leq \boldsymbol{x}_j^T\boldsymbol{x}_j - \boldsymbol{x}_i^T\boldsymbol{x}_i + w_i - w_j. \end{aligned}$$

Hence, we define $\boldsymbol{a}_j := 2(\boldsymbol{x}_j - \boldsymbol{x}_i)$ and $b_j := \boldsymbol{x}_j^T\boldsymbol{x}_j - \boldsymbol{x}_i^T\boldsymbol{x}_i + w_i - w_j$ for $j \in N_i$. Lastly, we need to add constraints for the case that a face of the cell is not caused by a neighboring cell, but by the domain. We conclude that we may solve the following LP problem:

$$\begin{array}{ll} \text{maximize:} & r \\ \text{subject to:} & \boldsymbol{a}_j^T(x, y, z)^T + r||\boldsymbol{a}_j|| \leq b_j, \quad j \in N_i \\ & x - r \geq 0, \quad x + r \leq l_1 \\ & y - r \geq 0, \quad y + r \leq l_2 \\ & z - r \geq 0, \quad z + r \leq l_3 \end{array} . \tag{B.2}$$

In practice, when a cell has faces which are caused by the domain we even know which of the six sides of our brick-shaped domain caused these faces. Therefore, we do not necessarily need to add all of the six additional constraints for each side of the domain:

$$\begin{aligned} x - r \geq 0, \quad x + r \leq l_1 \\ y - r \geq 0, \quad y + r \leq l_2 \\ z - r \geq 0, \quad z + r \leq l_3. \end{aligned}$$

We may simply use the constraints which correspond with sides of the domain which are coplanar with faces of the cell.

B.2 A Laguerre diagram with periodic boundary conditions

When dealing with periodic boundary conditions things change slightly. For one thing, faces of cells are always caused by neighboring cells and never by the domain. Other than that, for the cell L_i we can no longer just consider the generator points $\boldsymbol{x}_j$ with $j \in N_i$, we may need a

different "version" of these points. Generator points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ may be close w.r.t. the periodic norm, but they may be much further away w.r.t. the Euclidean norm.

Let $j \in N_i$, we compute:

$$(u, v, w) = \underset{(u,v,w)\in\mathbf{Z}^3}{\arg\min} ||\boldsymbol{x}_i - \boldsymbol{x}_j + (ul_1, vl_2, wl_3)||.$$

Then we define: $\tilde{\boldsymbol{x}}_j = \boldsymbol{x}_j - (ul_1, vl_2, wl_3)$, this particular choice means that: $||\boldsymbol{x}_i - \boldsymbol{x}_j||_{\text{per}} = ||\boldsymbol{x}_i - \tilde{\boldsymbol{x}}_j||$. We could call $\tilde{\boldsymbol{x}}_j$ the version of $\boldsymbol{x}_j$ that is closest to $\boldsymbol{x}_i$. Now, we define $\boldsymbol{a}_j := 2(\tilde{\boldsymbol{x}}_j - \boldsymbol{x}_i)$ and $b_j := \tilde{\boldsymbol{x}}_j^T \tilde{\boldsymbol{x}}_j - \boldsymbol{x}_i^T \boldsymbol{x}_i + w_i - w_j$ for $j \in N_i$. We may then simply solve the LP problem:

$$\begin{array}{ll} \text{maximize:} & r \\ \text{subject to:} & \boldsymbol{a}_j^T (x, y, z)^T + r||\boldsymbol{a}_j|| \leq b_j, \quad j \in N_i \end{array} . \tag{B.3}$$

Remark. *A demo of this sphere packing algorithm is included in the vorostereology package in the examples folder.*

C

Appendix: Proof of Lemma 1

Proof. In order to show that $L_i \cap P = L_i^P$ for all $i \in I$ it is sufficient to show:

$$||\boldsymbol{x}_i - \boldsymbol{x}||^2 - w_i = ||\boldsymbol{x}_i^P - \boldsymbol{x}||^2 - w_i^P \quad \text{for all } i \in I,\, \boldsymbol{x} \in P.$$

Therefore, let $i \in I$ and $\boldsymbol{x} \in P$. Write: $\boldsymbol{a} = (a, b, c)^T$, $\boldsymbol{x} = (x, y, z)^T$ and $\boldsymbol{x}_i = (x_i, y_i, z_i)^T$. Via direct computation it follows that:

$$\begin{aligned}
||\boldsymbol{x}_i^P - \boldsymbol{x}||^2 = &\quad \left|\left|\boldsymbol{x}_i - \boldsymbol{x} - \frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}\right|\right|^2 \\
= &\quad \left(x_i - x - \frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}a\right)^2 + \left(y_i - y - \frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}b\right)^2 + \\
&+ \left(z_i - z - \frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}c\right)^2 \\
= &\quad (x_i - x)^2 - 2(x_i - x)\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}a + \left(\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}a\right)^2 + \\
&+ (y_i - y)^2 - 2(y_i - y)\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}b + \left(\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}b\right)^2 + \\
&+ (z_i - z)^2 - 2(z_i - z)\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}c + \left(\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}c\right)^2 \\
= &\quad ||\boldsymbol{x}_i - \boldsymbol{x}||^2 - 2\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}) + \left(\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\right)^2 \boldsymbol{a}^T\boldsymbol{a} \\
= &\quad ||\boldsymbol{x}_i - \boldsymbol{x}||^2 - 2\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}) + \frac{\left(\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)\right)^2}{\boldsymbol{a}^T\boldsymbol{a}} \\
= &\quad ||\boldsymbol{x}_i - \boldsymbol{x}||^2 - 2\frac{\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)}{\boldsymbol{a}^T\boldsymbol{a}}\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0) + \frac{\left(\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)\right)^2}{\boldsymbol{a}^T\boldsymbol{a}} \qquad (\boldsymbol{a}^T\boldsymbol{x} = \boldsymbol{a}^T\boldsymbol{x}_0)
\end{aligned}$$

$$
\begin{aligned}
&= ||\boldsymbol{x}_i - \boldsymbol{x}||^2 - \frac{\left(\boldsymbol{a}^T(\boldsymbol{x}_i - \boldsymbol{x}_0)\right)^2}{\boldsymbol{a}^T\boldsymbol{a}} \\
&= ||\boldsymbol{x}_i - \boldsymbol{x}||^2 - w_i + w_i^P.
\end{aligned}
$$

We used that $\boldsymbol{a}^T\boldsymbol{x} = \boldsymbol{a}^T\boldsymbol{x}_0$, which holds since $x \in P$. By subtracting w_i^P on both sides of the equation we obtain the desired equality. □

D

Appendix: Partial derivatives of the log-likelihood

Recall that we defined:

$$b_{ij} := \log(a_j) - \frac{2}{3}\mu - \frac{2}{9}\sigma^2 - \log\left(a_i^{(1)}\right) \quad \text{and} \quad \beta := \frac{1}{\frac{8}{9}\sigma^2}.$$

Via straightforward computation we compute the partial derivatives of the log-likelihood:

$$\frac{\partial}{\partial\mu} l(\mu,\sigma) = \sum_{j=1}^{m} \frac{\frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right) \frac{\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)}{\frac{8}{9}\sigma^2}\frac{4}{3}}{\frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)} \tag{D.1}$$

$$= \sum_{j=1}^{m} \frac{\sum_{i=1}^{N} \exp\left(-b_{ij}^2\beta\right) \frac{4}{3} b_{ij}\beta}{\sum_{i=1}^{N} \exp\left(-b_{ij}^2\beta\right)} \tag{D.2}$$

$$\frac{\partial}{\partial\sigma} l(\mu,\sigma) = -\frac{m}{2}\frac{1}{2\pi\left(\frac{4}{9}\sigma^2\right)} 2\pi\frac{8}{9}\sigma + \sum_{j=1}^{m} \frac{\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\partial}{\partial\sigma}\exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)\right\}}{\frac{1}{N}\sum_{i=1}^{N} \exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)} \tag{D.3}$$

$$= -m\beta\frac{8}{9}\sigma + \sum_{j=1}^{m} \frac{\sum_{i=1}^{N} \exp\left(-b_{ij}^2\beta\right)\left(\frac{4}{9}\sigma^2 b_{ij} + b_{ij}^2\right)\frac{16}{9}\sigma\beta^2}{\sum_{i=1}^{N} \exp\left(-b_{ij}^2\beta\right)}. \tag{D.4}$$

Note that computing the derivative in the numerator in equation (D.3) yields:

$$\begin{aligned}
&\frac{\partial}{\partial\sigma}\exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)= \\
&=\exp\left(-\frac{\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2}{\frac{8}{9}\sigma^2}\right)\cdot \\
&\quad\cdot\frac{\frac{4}{9}\sigma^2\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)\frac{16}{9}\sigma+\left(\log(a_j)-\frac{2}{3}\mu-\frac{2}{9}\sigma^2-\log\left(a_i^{(1)}\right)\right)^2\frac{16}{9}\sigma}{4\left(\frac{4}{9}\sigma^2\right)^2}
\end{aligned} \tag{D.5}$$

E

Appendix: Computations related to Wicksell's corpuscle problem

In this section we rewrite the relationship as given by Wicksell [15], as also mentioned in section 8.1:

$$\phi(x) = \frac{x}{\mathbb{E}(R)} \int_x^\infty \frac{f(r)}{\sqrt{r^2 - x^2}} \mathrm{d}r. \tag{E.1}$$

Here, $\mathbb{E}(R)$ is the average radius of the 3D spheres:

$$\mathbb{E}(R) = \int_0^\infty r f(r) \mathrm{d}r.$$

Let us recall equation (8.1), which is the result of transforming $r \mapsto \pi r^2$ in equation (E.1) such that we have the density of the observed areas instead of the observed radii:

$$\tilde{f}_A(a) = \frac{1}{2\pi\mathbb{E}(R)} \int_{\sqrt{\frac{a}{\pi}}}^\infty \frac{f(r)}{\sqrt{r^2 - \frac{a}{\pi}}} \mathrm{d}r. \tag{E.2}$$

Via a change of variables we will now express $f(r)$ in (E.2) in terms of f_V which we define as the probability density function which describes the volume distribution of the spheres. Hence, we substitute:

$$r = \left(\frac{3v}{4\pi}\right)^{\frac{1}{3}} \iff v = \frac{4}{3}\pi r^3.$$

Once again, we use (7.7) and we obtain:

$$f(r) = f_V\left(\frac{4}{3}\pi r^3\right) 4\pi r^2.$$

Recall the definition of the constant $a_{\max}$:

$$a_{\max} := \frac{\pi}{\left(\frac{4}{3}\pi\right)^{\frac{2}{3}}} = \sqrt[3]{\frac{9\pi}{16}}.$$

This means that:

$$\begin{aligned}
\mathbb{E}(R) &= \int_0^\infty r f(r) \mathrm{d}r \\
&= \int_0^\infty r f_V\left(\frac{4}{3}\pi r^3\right) 4\pi r^2 \mathrm{d}r \\
&= \int_0^\infty \left(\frac{3v}{4\pi}\right)^{\frac{1}{3}} f_V(v) \mathrm{d}v && \text{substitute } v = \frac{4}{3}\pi r^3. \\
&= \left(\frac{3}{4\pi}\right)^{\frac{1}{3}} \mathbb{E}(V^{\frac{1}{3}}) && \text{where } V \sim f_V. \\
&= \sqrt{\frac{a_{\max}}{\pi}} \mathbb{E}(V^{\frac{1}{3}}).
\end{aligned}$$

And $\tilde{f}_A(a)$ becomes:

$$\begin{aligned}
\tilde{f}_A(a) &= \frac{1}{2\pi\mathbb{E}(R)} \int_{\sqrt{\frac{a}{\pi}}}^\infty \frac{f(r)}{\sqrt{r^2 - \frac{a}{\pi}}} \mathrm{d}r \\
&= \frac{1}{2\pi\sqrt{\frac{a_{\max}}{\pi}}\mathbb{E}(V^{\frac{1}{3}})} \int_{\sqrt{\frac{a}{\pi}}}^\infty \frac{f_V\left(\frac{4}{3}\pi r^3\right)}{\sqrt{r^2 - \frac{a}{\pi}}} 4\pi r^2 \mathrm{d}r \\
&= \frac{1}{2\pi\sqrt{\frac{a_{\max}}{\pi}}\mathbb{E}(V^{\frac{1}{3}})} \int_{\frac{a}{a_{\max}}}^\infty \frac{f_V\left(x^{\frac{3}{2}}\right)}{\sqrt{\frac{x a_{\max}}{\pi} - \frac{a}{\pi}}} 2a_{\max}\sqrt{\frac{x a_{\max}}{\pi}} \mathrm{d}x \quad \text{substitute } x = \left(\frac{4}{3}\pi r^3\right)^{\frac{2}{3}} = \frac{\pi}{a_{\max}} r^2. \\
&= \frac{2a_{\max}}{2\pi\sqrt{\frac{a_{\max}}{\pi}}\mathbb{E}(V^{\frac{1}{3}})} \int_{\frac{a}{a_{\max}}}^\infty \frac{f_V\left(x^{\frac{3}{2}}\right)}{\sqrt{1 - \frac{a}{x a_{\max}}}} \mathrm{d}x \\
&= \frac{3}{4a_{\max}\mathbb{E}(V^{\frac{1}{3}})} \int_{\frac{a}{a_{\max}}}^\infty \frac{f_V\left(x^{\frac{3}{2}}\right)}{\sqrt{1 - \frac{a}{x a_{\max}}}} \mathrm{d}x. && \text{(E.3)}
\end{aligned}$$

F

Appendix: A lognormal volume distribution

Recall the density $f_{V^{\frac{2}{3}}|I}$ which was defined in (7.8). When assuming that f_V is the density of a lognormal volume distribution the density $f_{V^{\frac{2}{3}}|I}$ corresponds to a lognormal distribution but with different parameters. Recall:

$$f_{V|I}(v) = \frac{v^{\frac{1}{3}} f_V(v)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)}.$$

The density $f_{V^{\frac{2}{3}}|I}$ corresponds to the density of a random variable $V_I^{\frac{2}{3}}$ with $V_I \sim f_{V|I}$. First, we show that $f_{V|I}$ is the density of lognormal distribution when f_V is the density of a lognormal distribution:

$$\begin{aligned}
v^{\frac{1}{3}} f_V(v) &= \frac{v^{\frac{1}{3}}}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(v)-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(v)-\mu)^2}{2\sigma^2} + \frac{1}{3}\log(v)\right) \\
&= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\log^2(v) - 2\mu\log(v) + \mu^2 - \frac{2}{3}\sigma^2\log(v)\right)\right) \\
&= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\log^2(v) - 2\left(\mu + \frac{1}{3}\sigma^2\right)\log(v) + \left(\mu + \frac{1}{3}\sigma^2\right)^2 - \frac{2}{3}\mu\sigma^2 - \frac{1}{9}\sigma^4\right)\right) \\
&= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(\left(\log(v) - \left(\mu + \frac{1}{3}\sigma^2\right)\right)^2\right) + \frac{1}{3}\mu + \frac{1}{18}\sigma^2\right) \\
&= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\log(v) - \left(\mu + \frac{1}{3}\sigma^2\right)\right)^2}{2\sigma^2}\right) \exp\left(\frac{1}{3}\mu + \frac{1}{18}\sigma^2\right).
\end{aligned}$$

And as a result:

$$\begin{aligned} f_{V|I}(v) &= \frac{v^{\frac{1}{3}} f_V(v)}{\mathbb{E}\left(V^{\frac{1}{3}}\right)} \\ &= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\log(v) - \left(\mu + \frac{1}{3}\sigma^2\right)\right)^2}{2\sigma^2}\right) \exp\left(\frac{1}{3}\mu + \frac{1}{18}\sigma^2\right) \frac{1}{\exp\left(\frac{1}{3}\mu + \frac{1}{18}\sigma^2\right)} \\ &= \frac{1}{v\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\log(v) - \left(\mu + \frac{1}{3}\sigma^2\right)\right)^2}{2\sigma^2}\right). \end{aligned}$$

For $V_I \sim f_{V|I}$ we conclude that $V_I \sim$ Lognormal$(\mu + \frac{1}{3}\sigma^2, \sigma^2)$. The expected value $\mathbb{E}\left(V^{\frac{1}{3}}\right)$ was computed via (2.2). Via (2.3) we may conclude that $V_I^{\frac{2}{3}} \sim$ Lognormal$(\frac{2}{3}\mu + \frac{2}{9}\sigma^2, \frac{4}{9}\sigma^2)$. To summarize: if f_V is the density of a lognormal distribution with parameters μ and σ^2 we obtain that $f_{V^{\frac{2}{3}}|I}$ is the density of a lognormal distribution with parameters $\bar{\mu} = \frac{2}{3}\mu + \frac{2}{9}\sigma^2$ and $\bar{\sigma}^2 = \frac{4}{9}\sigma^2$.