8-2015

Polygon Distances with Applications to High Performance Liquid Chromatography

Vito Capuano
Clemson University, vcapuan@clemson.edu

Polygon Distances with Applications to High Performance Liquid Chromatography

---

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical Sciences

---

by
Vito Michael Capuano
August 2015

Accepted by:
Dr. Michael Burr, Committee Chair
Dr. Christopher Cox
Dr. Kenneth Marcus

Abstract

Analytical chemistry uses high performance liquid chromatography (HPLC) to separate desired macromolecules from a fluid. This thesis is concerned with monolithic column environments used in HPLC separations. The monolithic column environment under consideration consists of many long polymer fibers suspended in a tube. Between the fibers are voids (interstices) which regulate the separation process. The goal of this thesis is both to identify interstices of a size such that the interstice contributes to the separation process, and to identify the boundaries of the fibers along such interstices. We model the cross section of the column environment with a collection of polygons. We define the desired interstices and boundaries in the model. We implement (using C++) an algorithmic approach to identify the desired boundaries from which we generate images and summarize relevant data.

Dedication

To Vanessa, who inspires me.

Acknowledgments

I would like to thank my committee for sharing their knowledge, resources, and feedback. I am especially thankful for the mirth, comments, and ken of my advisor, Michael Burr. He has helped me learn how to reason precisely and write with clarity. Thanks also to my fellow graduate students, especially Merle Glick, with whom I could talk about mathematics or whatever came to mind. I thank my family, especially my wife Vanessa, for their love and support. I thank my late grandfather, Paul Gallagher, whose life exhibited the virtue of diligence, I could not have made it this far without his example.

Table of Contents

List of Figures

List of Symbols

Denoting Points, Sets and Sets of Sets

Font	Use
lower case roman	Points
$CAPTIAL\ ROMAN$	Set of points
$\mathcal{CALLIGRAPHIC}$	Set of sets

New Distances and Identifications

Notation	Meaning
D_X^-	Points in the polygons of $\mathcal{D}$, provided the polygons are disjoint from X
$P_{0,\tau}(\mathcal{D})$	Identifies points of fiber-polygons which are within τ of another polygon
$\mathcal{S}_{0,\tau}(\mathcal{D})$	Maximal line segments of $P_{0,\tau}(\mathcal{D})$
$\ell_{0,\tau}(\mathcal{D})$	Computes the total length of the line segments in $\mathcal{S}_{0,\tau}(\mathcal{D})$
$A_{0,\tau}(\mathcal{D})$	Identifies points in the interstices based on $\mathcal{S}_{0,\tau}(\mathcal{D})$
$\mathcal{A}_{0,\tau}(\mathcal{D})$	Maximal polygons of $A_{0,\tau}(\mathcal{D})$
$a_{0,\tau}(\mathcal{D})$	Computes the total area of the polygons in $\mathcal{A}_{0,\tau}(\mathcal{D})$

Set Theory Notation

Notation	Meaning
$x \in X$	x is an element of the set X
$\emptyset$	The unique set with no elements
$X \subseteq Y$	Every element of X is an element of Y
$X \cup Y$	The set of elements in either X or Y
$X \cap Y$	The set of elements in both X and Y

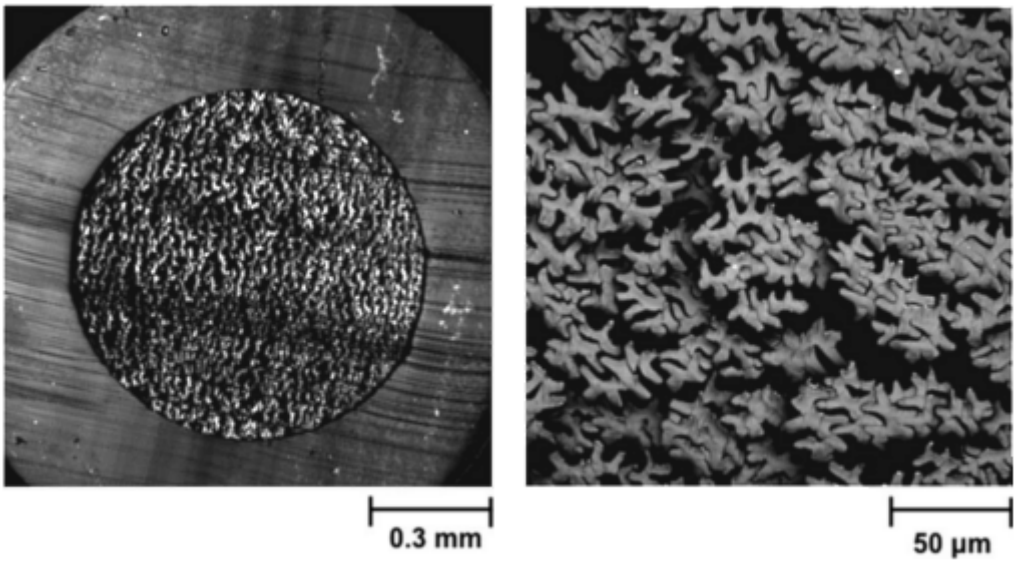
Chapter 1

Introduction

The work presented in this thesis is motivated by active research in analytical chemistry. One aspect of analytical chemistry studies how to improve purification processes using high performance liquid chromatography [13, 14, 16]. This thesis focuses on the column environment used in this purification process. The column environment is described in Section 1.1; a cross section of the column environment can be seen in Figure 1.1. In this thesis, we create a mathematical model to study the column environment. Our model is used to find regions in the column environment which make an appreciable contribution to the purification process. These regions are explained in Section 1.2. We offer two calculations on our model which identify and quantify these regions. To complete these calculations, we use the C++ library CGAL which is discussed in Section 1.3.

1.1 Motivation from Chemistry

The column environment we study is a monolithic column which consists of a tube, a large number of thin polymer fibers that extend the length of the tube, and a fluid. The fluid contains macromolecules which must separated for desired purifications. The fibers inside the tube create and determine the location and sizes of the voids (interstices) through which the fluid may flow. The positions of the numerous fibers do not exhibit a regular



(a) (b)

Figure 1.1: The Column Environment: Cross sectional images of the column environment showing the fibers and interstices, images taken from [14]. In (a), we see the tube containing the fibers. In (b), we see that the fibers have the same shape and are not arranged in any regular pattern. The size of the interstices are irregular and depend on the position of the fibers.

pattern and thus the interstices are varisized, see Figure 1.1. In [14], the author notes that the positions of the fibers are assumed to remain nearly constant throughout the length of the tube. Therefore, we assume all cross sections of the column environment are the same.

The boundaries of the fibers along the interstices and the interstices, are the regions in the column environment which contribute to the purification process. The size of the interstices determines if the molecules in the interstice diffuse to the boundary of the fiber, a crucial step in the purification process. When a macromolecule just barely fits into an interstice, it experiences the most amount of interaction with the surrounding fibers. When a macromolecule experiences large amounts of interaction, the macromolecule passes though the column relatively slowly and diffuses more rapidly to the boundary of a fiber. Interstices of large size allow the macromolecule to pass through the column more rapidly, which results in less interaction with the surrounding fibers. Due to the reduced interaction,

the macromolecule takes longer to diffuse to the boundary of a fiber, or is unlikely to do so.

The size of some interstices may be too small or too large to separate the desired macromolecules in the fluid. Given a macromolecule, we let τ_{min} be the size of an interstice for which the macromolecule is too large to enter; and, we let τ_{max} be the smallest size of an interstice for which the macromolecule is unlikely to diffuse, because the interstice is too large. Therefore, interstices of size smaller than τ_{min} or larger than τ_{max} are unlikely to contribute to the separation process. We call the interstices of sizes between τ_{min} and τ_{max} channels. The boundaries of the fibers along the channels are called the channel-walls.

Our model of a column environment is of a representative cross section of the column wall, fibers, and interstices under the assumption that all cross sections of the column environment are the same, i.e., we study the column environment by studying a cross section. Our first computation identifies the channel-walls and calculates the lengths of the channel-walls. Our second computation identifies the channels and calculates the total area of the channels. The latter computation determines the area of the interstices through which macromolecules flow and may diffuse to a channel-wall.

1.2 A Mathematical Model

In this section, we present our model and calculations. Our model consists of a set of polygons, see Figure 1.2. A regular polygon is used to approximate the column wall. Polygons are used to represent the boundaries of the fibers' cross sections and take the shape exhibited in Figure 1.1. These polygons are called fiber-polygons.

Notation 1.2.1. *We use $\mathcal{CALLIGRAPHIC}$ font to indicate sets of sets, CAPITAL ROMAN letters to denote sets of points, and lower case roman letters to denote points.*

Definition 1.2.2. Let C be a regular polygon representing the inner boundary of the column wall in the cross section. Let $\mathcal{P}$ be a collection of fiber-polygons. Then $\mathcal{P} := \bigcup\{P_i\}$ where P_i is a fiber-polygon. Let P be the points from the collection of fiber-polygons. Then $P := \bigcup P_i$, where the union is taken over all fiber-polygons P_i. Let $\mathcal{D}$ be the collection of

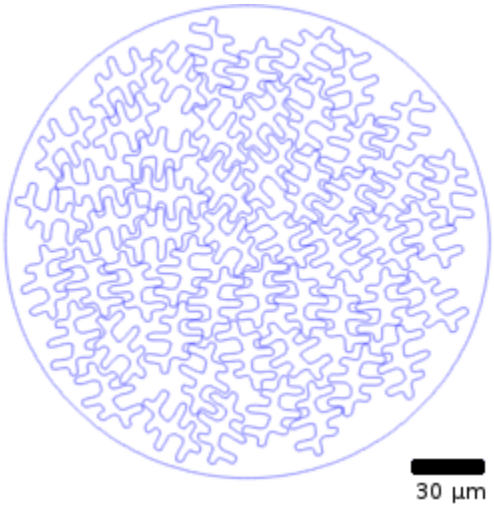


Figure 1.2: The column wall is represented by a polygon which approximates a circle. The disjoint fiber-polygons are arranged in a similar way as the fibers in Figure 1.1. The inner column wall has a diameter of 215 μm. The polygon-fibers have a diameter, cf. Definition 2.2.2, of 30 μm. The data for this diagram is from [7].

polygons which includes the fiber-polygons and the regular polygon in the plane representing the inner boundary of the column wall. We call this collection of polygons the domain. The domain is given by $\mathcal{D} := \mathcal{P} \cup \{C\}$.

Definition 1.2.3. Let X be any geometric object (point, line segment or polygon) or collection of geometric objects that can be formed from the objects in the domain $\mathcal{D}$. We define the collection of polygons in $\mathcal{D}$ which are disjoint from X to be

$$D_X^- := \bigcup_{\substack{P_i \in \mathcal{D} \\ P_i \cap X = \emptyset}} P_i$$

Our next definition, studied further in Section 2.4, identifies the channel-walls, which are boundaries of the fibers necessary for the separation process; that is, the portion of the boundaries of the polygons which form interstices of size between τ_{min} and τ_{max}.

Definition 1.2.4. Let P and $\mathcal{D}$ be given as in Definition 1.2.2. We define

$$P_{\tau_{min},\tau_{max}}(\mathcal{D}) := \left\{ x \in P : \ \tau_{min} < d(x, D_x^-) \leq \tau_{max} \right\}.$$

Observe that $P_{\tau_{min},\tau_{max}}(\mathcal{D})$ can be expressed as a union of maximal disjoint line segments $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$.

Definition 1.2.5. The sum of the lengths of the line segments in $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$ quantifies the amount of the fiber boundaries where there is likely interaction with macromolecules, we refer to this value as the perimeter-distance of the domain,

$$\ell_{\tau_{min},\tau_{max}}(\mathcal{D}) := \sum_{S_i \in \mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})} \text{length}(S_i)$$

Definition 1.2.6. In Section 5.1, we identify a collection, $\mathcal{A}$, of disjoint polygons which approximate the collection of channels. The total area of these polygons quantifies amount of cross sectional area through which macromolecules in a fluid experience interaction, we refer to this value as the channel-area of the domain,

$$a_{\tau_{min},\tau_{max}}(\mathcal{D}) := \sum_{A_i \in \mathcal{A}} \text{area}(A_i).$$

1.3 Programming

To implement the calculations developed in this thesis, I used the Computational Geometry Algorithms Library (CGAL) [2], an open source software library written in the programming language C++. The classes of CGAL represent geometric objects such as points, line segments, circles, and polygons. The geometric operations on these objects include the intersection of lines and circles, the construction of supporting lines for line segments and perpendicular lines to a given line, the computation of the length of line segments, the construction of convex hulls of given points, and the construction of the union and intersection of polygons. In this thesis, these operations are used to compute $\ell_{\tau_{min},\tau_{max}}(\mathcal{D})$ and $a_{\tau_{min},\tau_{max}}(\mathcal{D})$.

Due the amount of computation needed to collect the data for Figure 4.4, I used the Clemson University Palmetto Cluster [4]. My real computation time was substantially re-

duced due to the parallelization effects of the job array feature which runs multiple instances of the same program simultaneously.

The computations used to produce Figure 4.4 used the number type CORE in CGAL. CORE is an exact geometric and arithmetic computation library. CORE is maintained by the Exact Geometric Computation Group within the Computer Science Department of the Courant Institute for Mathematical Sciences in NYU [3].

The images in this thesis depicting the model and fiber-polygons were rendered using Qt4 [1].

Chapter 2

Distances

In Chapter 1, we explained how the channel-walls of the domain represent the boundaries of the cross section necessary for macromolecule separation. We show that minimum, maximum and Hausdorff distances are not useful for identifying the channel-walls, identifying the channels, or calculating the length or area of these regions. We exhibit the limitation for each in Figures 2.1, 2.2, and 2.3, respectively. We offer a mathematical definition for the identification of channel-walls and perimeter-distance in Section 2.4.

2.1 Minimum Distance

Definition 2.1.1. Let M and N be compact sets. The *minimum distance* between M and N is defined by

$$d(M, N) = \min_{x \in M, y \in N} d(x, y).$$

This distance is useful for identifying when two objects are far apart, but this calculation fails to capture geometric information (e.g., shape and size) that is not represented by the nearest points of these objects.

Example 2.1.2. Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be disjoint fiber-polygons. If the minimum distance between the polygons is large enough, (i.e., $d(P_1, P_2) \geq \tau_{max}$), then there are no points on either polygon such that a point on the other polygon is nearby, and

thus there are no channels, as shown in Figure 2.1(a).

Example 2.1.3. A limitation to minimum distance is that it is not sensitive to differences in channel size due to changes in position or orientation. Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be disjoint fiber-polygons. Let P_3 and P_4 be disjoint fiber-polygons where $d(P_1, P_2) = d(P_3, P_4) < \tau_{max}$. In Figures 2.1(b) and 2.1(c), the polygons are nearly touching and the minimum distances are the same. Under these conditions, the fiber-polygons P_1 and P_2 can be arranged such that the interstice of size smaller than τ_{max} has a large area, while P_3 and P_4 can be arranged to form an interstice of significantly smaller area. Hence, minimum distance is not able to distinguish between channels of large or small area.

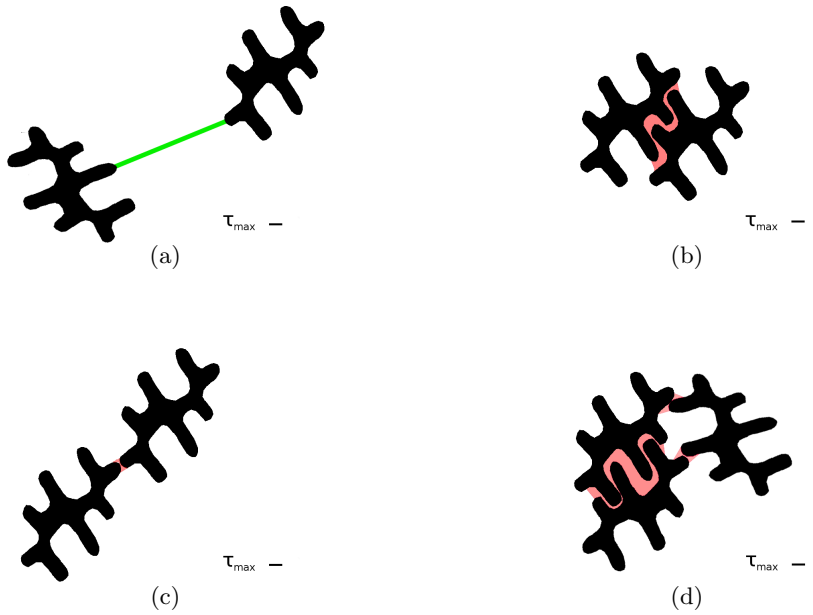


Figure 2.1: Examples of Minimum Distance: In these figures, τ_{max} is specified as the length of the line segment in each of the figures and $\tau_{min} = 0$. In (a), the minimum distance is greater than τ_{max}; we see that no points on the boundary of either polygon are close to the other polygon. The minimum distances between the polygons in (b) and (c) are the same, but the area of the shaded region in (b), the interstice of size less than τ_{max}, is significantly greater than the shaded region in (c). The difference in area of the these shaded regions, and length of the channel-walls, demonstrates the limitation of minimum distance calculations. This limitation becomes more apparent when several polygons are near to one another, as in (d).

2.2 Maximum Distance

Definition 2.2.1 (cf. [10])**.** *Maximum distance* between compact sets M and N is defined to be

$$d_{\max}(M,N) = \max_{x\in M, y\in N} d(x,y).$$

This distance, like minimum distance, is useful for identifying when objects are far apart. In Example 2.2.5, we see how these calculations fail to capture geometric information that is not represented by the two furthest points on the geometric objects.

Definition 2.2.2. Let X be a compact set. The *diameter* of X is given by the distance between the points in X which are the furthest apart. We denote the diameter of X to be

$$Diam(X) := \max_{x,y\in X} d(x,y).$$

Maximum distance can be used to find when fiber-polygons are far apart as in the following lemma:

Lemma 2.2.3. *Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be polygons of the same diameter. If $\tau_{max} + 2(Diam(P_1)) < d_{max}(P_1,P_2)$, then $\tau_{max} < d(P_1,P_2)$.*

Proof. Let v_1 and w_1 be points on P_1 and P_2 respectively which give the maximum distance between P_1 and P_2. Let v_2 and w_2 be points on P_1 and P_2, respectively, which give the minimum distance between P_1 and P_2. By the triangle inequality, the length of the direct path from v_1 and w_1 is the same length or longer than the one given by $v_1 \to v_2 \to w_2 \to w_1$. Equivalently,

$$d_{max}(P_1,P_2) \le d(P_1,P_2) + d(v_1,v_2) + d(w_2,w_1).$$

Given that $Diam(P_1) = Diam(P_2)$ and the definition of diameter of sets, it follows that $d(v_1,v_2), d(w_2,w_1) \le Diam(P_1)$. Therefore,

$$d_{max}(P_1,P_2) \le d(P_1,P_2) + 2Diam(P_1).$$

From the statement of the lemma, $\tau_{max} + 2Diam(P_1) < d_{max}(P_1, P_2)$. Hence,

$$\tau_{max} + 2Diam(P_1) < d(P_1, P_2) + 2Diam(P_1).$$

Therefore,

$$\tau_{max} < d(P_1, P_2),$$

as desired. □

Example 2.2.4. Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be disjoint fiber-polygons, so $Diam(P_1) = Diam(P_2)$. Then if we position P_1 and P_2 such that $\tau_{max} + 2(Diam(P_1)) \leq d_{max}(P_1, P_2)$ then the distance between a pair of points taken from each polygon is greater than τ_{max}, so there is no channel, as shown in Figure 2.2(a).

Example 2.2.5. Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be disjoint fiber-polygons. Let P_3 and P_4 be disjoint fiber-polygons where $d_{max}(P_1, P_2) = d_{max}(P_3, P_4)$. The length of the interstices boundaries along the polygons P_1 and P_2, see Figure 2.1(b), may be significantly longer than the length of the interstice along the channel between P_3 and P_4, cf. Figure 2.1(c).

2.3 Hausdorff Distance

Definition 2.3.1. Let M and N be compact sets. The *Hausdorff distance* between M and N is defined as

$$d_H(M, N) = \max \left\{ \max_{x \in M} \min_{y \in N} d(x, y),\ \max_{y \in N} \min_{x \in M} d(x, y) \right\}.$$

In Lemma 2.5.4, we exhibit a similarity between Hausdorff distance and our calculation, perimeter-distance. Hausdorff distance has been utilized in several computational applications including image pattern recognition [11]. Hausdorff distance has also been studied to find distances between two spatially separate objects [8]. By definition, for two

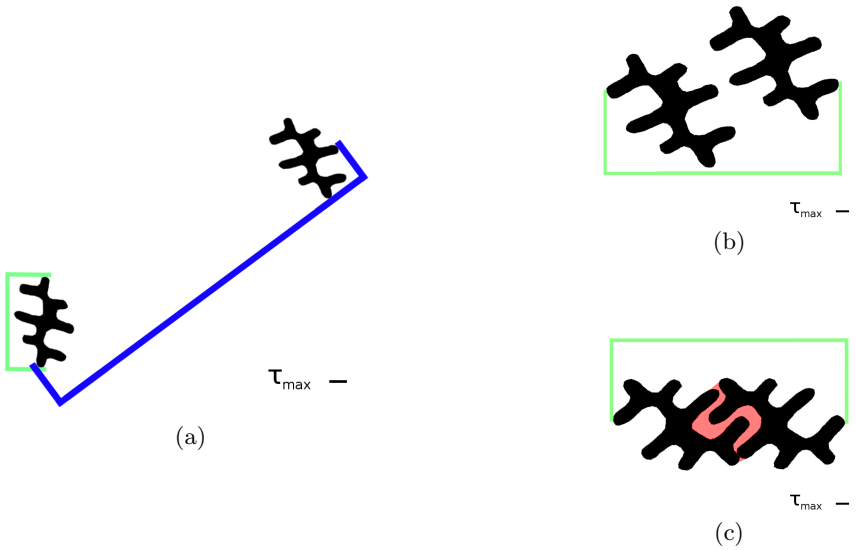
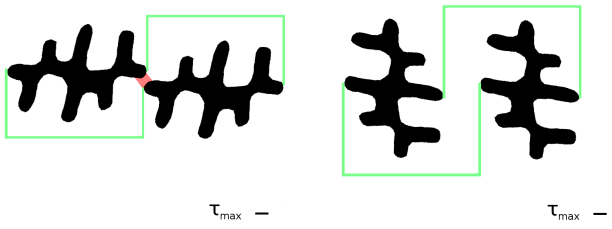


Figure 2.2: Examples of Maximum Distance: These figures show that maximum distance provides limited information about the geometry of neighboring polygons. In these figures, τ_{max} is given as the length of the line segment in each of the figures and $\tau_{min} = 0$. The shaded regions shows the space bounded by the channel-walls. In Figure (a), we see that when the maximum distance (shown with large calipers) between two polygons is large enough there is no channel. A sufficient condition that no channel is present between a pair of polygons is given by $\tau_{max} + Diam(P_1) + Diam(P_2) \leq d_{max}(P_1, P_2)$. The diameter of the fiber-polygons are equal, shown with the smaller calipers in (a). The limitations of maximum distance are demonstrated by comparing (b) and (c). The maximum distances of the polygons (shown with calipers) are the same, but in (c) there are points along the boundaries of the polygons which are within τ_{max} of another polygon (identified by the shaded region). However, in (b), no such points along the boundaries exists, and thus no channels exist.

sets X and Y, $d(X,Y) \leq d_H(X,Y) \leq d_{max}(X,Y)$. For these reasons one might be inclined to believe that Hausdorff distance could be used to identify the channels in the domain and would not suffer from the same issues found with maximum and minimum distance. However, this distance also fails to be sensitive to significant differences in the area of the channels and the length of the channel-walls.

Example 2.3.2. Let $\tau_{max} > 0$ and $\tau_{min} = 0$. Let P_1 and P_2 be disjoint fiber-polygons. Let P_3 and P_4 be disjoint fiber-polygons where $d_H(P_1, P_2) = d_H(P_3, P_4)$. In Figures 2.3(a) and 2.3(b), the Hausdorff distances are the same. In Figure 2.3(a), a channel is present

and there are pairs of points from P_1 and P_2 which are within τ_{max} of the other polygon. All pairs of points taken from P_3 and P_4 may have a distance greater than τ_{max}, so no channel is present, as in Figure 2.3(b). These differences in geometry demonstrate similar limitations as with minimum and maximum distance calculations.



(a) (b)

Figure 2.3: Examples of Hausdorff Distance: In these figures, τ_{max} is given as the length of the line segment in both of the figures, and $\tau_{min} = 0$. All the calipers in (a) and (b) measure the same length, they measure the distance between two points where the first point is on one fiber-polygon that is furthest point from other fiber-polygon and where the second point comes from the fiber-polygon which does not contain the first point and is the closest to the first point. Since all the calipers show the same distance, the Hausdorff distance between the fiber-polygons is the same for both (a) and (b). However, in (a) there is a shaded region which denotes the presence of a channel, in (b) no channel is present. Therefore, Hausdorff distance does not identify where separation will occur in the column environment.

2.4 Perimeter Distance

For a domain $\mathcal{D}$ and set of the points in the fiber-polygons P, the calculation of perimeter-distance, $P_{\tau_{min},\tau_{max}}(\mathcal{D}) \subseteq P$, begins by identifying points which satisfy the following conditions. A point $p \in P$ is included in $P_{\tau_{min},\tau_{max}}(\mathcal{D})$ if

1. The distance to some other polygons is less than or equal to τ_{max}, and,

2. The distance to all other polygons is greater than τ_{min}.

We formalize this with the following remarks:

Remark 2.4.1. Let X and Y be sets and $\tau_{min} > 0$. $\tau_{min} < d(X,Y)$ if and only if $\tau_{min} < d(x,y)$ for all $x \in X$ and for all $y \in Y$.

Remark 2.4.2. Let X and Y be sets and $0 < \tau_{min} < \tau_{max}$. We say $d(X,Y) \leq \tau_{max}$, if there exist $x \in X$ and $y \in Y$ such that $d(x,y) \leq \tau_{max}$.

Observe, the negation of $\tau_{max} < d(X,Y)$ is $d(X,Y) \leq \tau_{max}$.

Definition 2.4.3. Let $0 < \tau_{min} < \tau_{max}$. Let P and $\mathcal{D}$ be given as in Definition 1.2.2. Based on Remarks 2.4.1 and 2.4.2, we define the *channel-walls* of $\mathcal{D}$ to be

$$P_{\tau_{min},\tau_{max}}(\mathcal{D}) := \left\{x \in P : \ \tau_{min} < d(x, D_x^-) \leq \tau_{max}\right\}$$

where D_x^- is given as in Definition 1.2.3.

Recall from Section 1.2 that $P_{\tau_{min},\tau_{max}}(\mathcal{D})$ can be expressed as a union of maximal disjoint line segments $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$.

Definition 2.4.4. We define *perimeter-distance* of $\mathcal{D}$ to be

$$\ell_{\tau_{min},\tau_{max}}(\mathcal{D}) := \sum_{S_i \in \mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})} \text{length}(S_i).$$

In Proposition 3.3.1, the expression in Definition 2.4.4 is separated into two simpler calculations.

Example 2.4.5. Let C be the axis-aligned square in the first quadrant with vertices at the origin and the point $(10, 10)$, see Figure 2.4. Let P_0, P_1 be the axis-aligned rectangles where P_0 has vertices

$$\{(5.5, 1), (8.5, 1), (8.5, 4), (5.5, 4)\},$$

and, P_1 has vertices

$$\{(5.5, 4.5), (5.5, 6.5), (2.5, 6.5), (2.5, 4.5)\}.$$

Let $\mathcal{P}$ consist of the two rectangles P_0 and P_1, $\mathcal{P} = \{P_0\} \cup \{P_1\}$. Let $\mathcal{D} = \mathcal{P} \cup \{C\}$. Then,

$$\begin{aligned}\mathcal{S}_{1,2}(\mathcal{D}) &= \{\overline{(17/2,1),(17/2,4)}, \overline{((11+\sqrt{3})/2,4),((11+\sqrt{15})/2,4)}, \overline{(8,4),(17/2,4)}, \overline{(11/2,1),(11/2,2)},\\ &\quad \overline{(11/2,5/2),(11/2,7/2)}, \overline{((11-\sqrt{15})/2,9/2),((11-\sqrt{3})/2,9/2)}, \overline{(11/2,5),(11/2,6)}\}.\end{aligned}$$

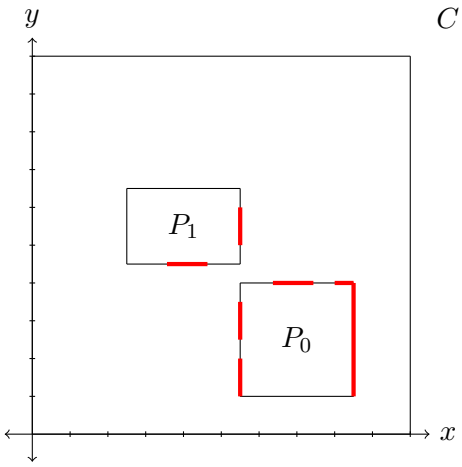


Figure 2.4: Perimeter Distance Example: In this figure, the two smaller rectangles, P_1 and P_2, are polygons in $\mathcal{P}$. The large square is C. The line segments in $\mathcal{S}_{1,2}(\mathcal{D})$ are shown as the bold lines. The sum of the length of these line segments is value of the perimeter-distance $\ell_{(1,2)}(\mathcal{D})$.

Definition 2.4.6. Let $\tau_{min} = 0$ and $0 < \tau_{max}$. Let P and $\mathcal{D}$ be given as in Definition 1.2.2. We define the *interstice-walls* of $\mathcal{D}$ of size τ_{max} to be $P_{0,\tau_{max}}(\mathcal{D})$.

Example 2.4.7. Let $\mathcal{P}$ be a collection of disjoint fiber-polygons. Let $\tau_{min} = 0$. Then the points which are part of the line segments on the boundaries of the polygons whose length is summed to compute perimeter distance are points along the boundary of the polygons which are within τ_{max} of another polygon, Figure 2.5(a) shows these boundaries using CGAL and Qt4.

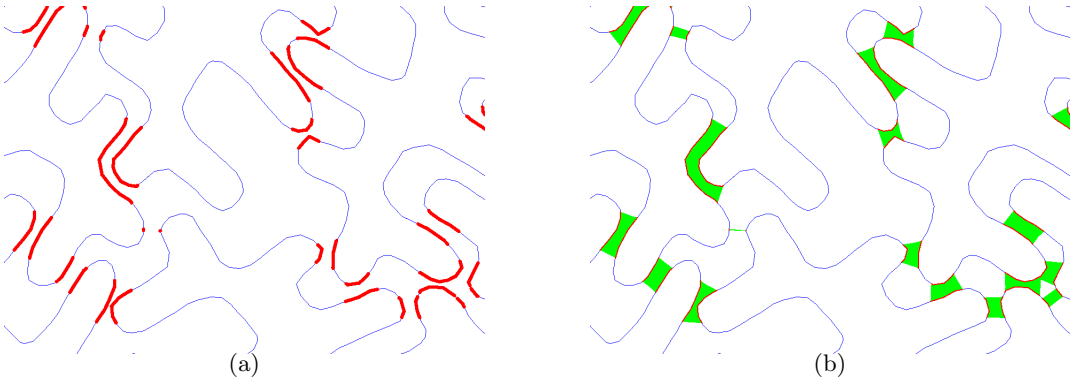


Figure 2.5: Perimeter Distance: In (a), we see segments which are used to compute perimeter-distance performed on many fiber-polygons where $\tau_{min} = 0$. In (b), we show how these lines connect to form channels, the method illustrated here is discussed in Chapter 5.

2.5 Properties and Inequalities for Perimeter Distance

In the following proposition, we obtain a lower bound on the radius of a disk shaped macromolecule that touches two channel-walls. In Proposition 2.5.2, we also find an upper bound for the distance between channel-walls.

Proposition 2.5.1. *Let R_0 and R_1 be segments in $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$. Let B be a disk with radius r. If $B \cap R_0$, $B \cap R_1 \neq \emptyset$, then $r \geq \tau_{min}/2$.*

Proof. Let $p_0 \in B \cap R_0$ and $p_1 \in B \cap R_1$. Let h be the center of the disk B. By definition of the diameter of B, $d(p_0, p_1) \geq \tau_{min}$. On the other hand, $d(p_0, h) \leq r$ and $d(p_1, h) \leq r$. By the triangle inequality, $d(p_0, p_1) \leq d(p_0, h) + d(p_1, h)$. Therefore,

$$\tau_{min} \leq d(p_0, p_1) \leq d(p_0, h) + d(p_1, h) \leq 2r.$$

It follows that $r \geq \tau_{min}/2$, as desired. □

In the following proposition, we obtain an upper bound on the radius of a disk shaped macromolecule that touches two channel-walls.

Proposition 2.5.2. *Let S_0 and S_1 be segments in $\mathcal{P}$ with respective supporting lines l_0 and l_1. Consider the line segment $R_0^* := \{x \in l_0 : d(x, l_1) \leq \tau_{max}\}$ which consists of the points on the supporting line l_0 which are within τ_{max} of the supporting line l_1. Let R_0^* be the maximal segment of $\{x \in l_0 : d(x, l_1) \leq \tau_{max}\}$, and let R_1^* be defined similarly. Let B be a disk with radius r which is tangent to l_0 and l_1 such that $B \cap R_0^*$, $B \cap R_1^* \neq \emptyset$. If $l_0 || l_1$, then $r \leq \tau_{max}/2$. If l_0 and l_1 are not parallel, then*

$$r \leq \frac{\tau_{max}}{2} \cdot \max\left\{\sec^2\left(\frac{\theta}{2}\right),\ \csc^2\left(\frac{\theta}{2}\right)\right\},$$

where θ is the angle formed by l_0 and l_1, see Figure 2.6(b).

Proof. If $l_0 || l_1$, then either ($R_0^* = l_0$ and $R_1^* = l_1$) or $R_0^* = R_1^* = \emptyset$. It was given that $B \cap R_0^* \neq \emptyset$, so ($R_0^* = l_0$ and $R_1^* = l_1$). Since $R_0^* = l_0$ and $R_1^* = l_1$, then $d(l_0, l_1) \leq \tau_{max}$. It was given that B is tangent to the lines l_0 and l_1, and we have assumed $l_0 || l_1$; therefore, the radius of B is half the distance between the lines, i.e., $r = d(l_0, l_1)/2$. Hence, $r \leq \tau_{max}/2$.

If l_0 is not parallel to l_1, let θ be the measure of the acute angle formed between l_0 and l_1 and let O be the intersection of l_0 and l_1. Let h be the center of the disk B. Let q be the point of tangency of the disk with l_0, i.e., $q = B \cap l_0$. Define $l_{1_q}^{\perp}$ be the line perpendicular to l_1 that passes through q. Let m be the intersection of l_1 and $l_{1_q}^{\perp}$.

$\angle Oqm = \pi/2 - \theta$ because $\angle Omq$ is a right triangle. By definition, $\sin\theta = \frac{d(q,m)}{d(O,q)}$. By the construction of R_0^*, we have $d(q, m) \leq \tau_{max}$, since m is the closest point on l_1 to q. Hence,

$$d(O, q) \leq \frac{\tau_{max}}{\sin\theta}.$$

Since h is the center of the circle tangent to l_0 and l_1, symmetry implies that the line segment from O to h is the bisector of the angle formed between l_1 and l_0 and $\angle Oqh$ is a right angle. Therefore, $\frac{d(h,q)}{d(O,q)} = \tan(\theta/2)$. Equivalently,

$$\frac{r}{d(O, q)} = \tan(\theta/2).$$

By combining the previous equality with the previous inequality, we have bounded the length of the radius; $r \leq \tan(\theta/2)(\tau_{max}/\sin\theta)$. Equivalently,

$$r \leq \frac{\tau_{max}}{2}\sec^2(\theta/2)$$

using the double angle formula, which immediately gives the desired inequality. Using the same trigonometry arguments for the supplementary angle $\pi - \theta$, we obtain $r \leq \frac{\tau_{max}}{2}\csc^2\left(\frac{\theta}{2}\right)$. □

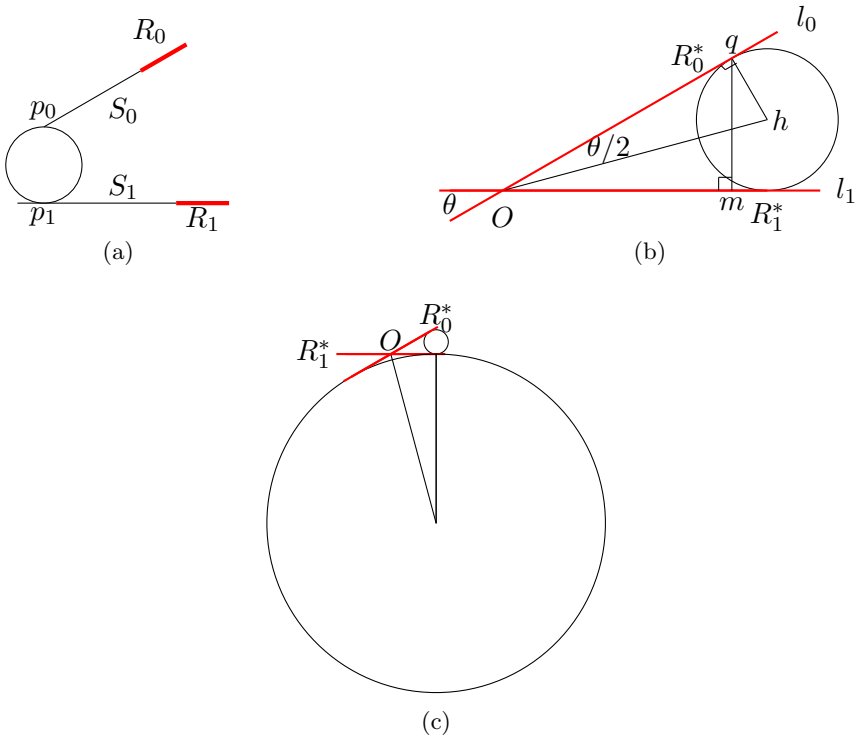


Figure 2.6: Separation of Nearby Segments in $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$: In these figures, the circles are the boundaries of disks. In (a), the circle has a radius which is too small to touch both red segments from $\mathcal{S}_{\tau_{min},\tau_{max}}$; however, the circle does touch both supporting edges S_0 and S_1. In (b) and (c), the disks are tangent to the supporting lines of R_0^* and R_1^*, the points along these line segments are within τ_{max} of the other; and, we illustrate some of the steps in the proof of Proposition 2.5.1.

The following proposition allows for a simpler calculation of $\ell_{\tau_{min},\tau_{max}}(\mathcal{D})$.

Proposition 2.5.3. *Let $0 < \tau_{min} < \tau_{max}$ and let P and $\mathcal{D}$ be given as in Definition 1.2.2.*

$$P_{\tau_{min},\tau_{max}}(\mathcal{D}) = P_{0,\tau_{max}}(\mathcal{D}) \setminus P_{0,\tau_{min}}(\mathcal{D})$$

where the right hand side is a set difference.

Proof. Let $x \in P_{\tau_{min},\tau_{max}}(\mathcal{D})$. Equivalently, $\tau_{min} < d(x, D_x^-) \leq \tau_{max}$. This is true if and only if both $x \in P_{0,\tau_{max}}(\mathcal{D})$ and $x \notin P_{0,\tau_{min}}(\mathcal{D})$, since x and D_x^- are disjoint and by definition $\tau_{min} < d(x, D_x^-)$. Equivalently, $x \in P_{0,\tau_{max}}(\mathcal{D}) \setminus P_{0,\tau_{min}}(\mathcal{D})$. Therefore, we have the equality $P_{\tau_{min},\tau_{max}}(\mathcal{D}) = P_{0,\tau_{max}}(\mathcal{D}) \setminus P_{0,\tau_{min}}(\mathcal{D})$. □

In the following proposition, we see that the upper bound on nearby channel-walls is similar in form to Hausdorff distance, cf. Definition 2.3.1.

Proposition 2.5.4. *Let P and $\mathcal{D}$ be given as in Definition 1.2.2.*

$$w = \max_{x \in P} \min_{y \in D_x^-} d(x, y)$$

is the smallest value such that $P_{0,w}(\mathcal{D}) = P$.

Proof. Let $x \in P$. By the definition of w there exists $y \in D_x^-$ such that $d(x, y) \leq w$. Hence, $x \in P_{0,w}(\mathcal{D})$. Thus, $P \subseteq P_{0,w}(\mathcal{D})$. By the definition of $P_{0,w}(\mathcal{D})$, it follows that $P_{0,w}(\mathcal{D}) \subseteq P$. Therefore, $P_{0,w}(\mathcal{D}) = P$. No u such that $u < w$ can satisfy $P_{0,u}(\mathcal{D}) = P$, for by the definition of w there exists a $x \in P$ such that $d(x, D_x^-) = w > u$. So, $x \notin P_{0,u}(\mathcal{D})$. Therefore, w is the smallest value such that $P_{0,w}(\mathcal{D}) = P$. □

This result immediately can be understood in terms of perimeter-distance in the following corollary:

Corollary 2.5.5. *Let $\tau > 0$ where $\mathcal{P}$ and $\mathcal{D}$ as in Definition 1.2.2. Let w be defined as in Proposition 2.5.4, w is the smallest value such that*

$$\ell_{0,w}(\mathcal{D}) = Perimeter(\mathcal{P})$$

where $Perimeter(\mathcal{P})$ is the sum total of the perimeters of the polygons in $\mathcal{P}$.

Proof. The equality follows from Proposition 2.5.4. □

In Chapter 1, we determined that macromolecule separations for the purpose of purification require channels of size between τ_{min} and τ_{max}. It was also stated the molecules diffuse to the channel-walls. In this chapter, we have seen how existing distances, such as minimum, maximum and Hausdorff, are not able to adequately identify channels or channel-walls. In Section 2.4, we developed $P_{\tau_{min},\tau_{max}}(\mathcal{D})$ which identifies the channel-walls. We also developed $\ell_{\tau_{min},\tau_{max}}(\mathcal{D})$ which calculates the length of the channel-walls. By determining the total length of the channel-walls, we have a quantification of the utility of the column environment. In Chapter 5, we see how our identification of channel-walls of the domain is consistent with our identification of the channels of the domain. In Chapter 4, we provide algorithms for computing perimeter-distance and one approach to reduce the computational resources needed to compute perimeter-distance.

Chapter 3

Theoretical Computation of Perimeter Distance

In the previous chapter, we created a precise definition for perimeter-distance which is the length of the boundaries of the fiber-polygons which represent the boundaries of the cross sections of the fibers in the column environment to which the macromolecules may diffuse during the separation process. Our definition, however, offers little suggestion on how to compute perimeter-distance. In this chapter, our goal is to show how to use Minkowski sums to simplify the calculation of perimeter-distance. The actual implementation of perimeter-distance uses the theoretical results in this chapter in an efficient algorithm for computing perimeter-distance. In Section 3.1, we fix notation, which is analogous to that in CGAL, that makes the proofs of this chapter more streamlined. In Section 3.2, we see that the computation of perimeter-distance can be reduced to computations between pairs of edges on the fiber-polygons. In Proposition 3.2.6, we see that these calculations involve only operations on line segments and circles. This, in turn, makes for a manageable implementation of perimeter-distance, discussed in Chapter 4.

3.1 Notation

In this section we fix notation, similar to that used in CGAL, for finding endpoints of line segments. We also fix a total order on line segments.

Notation 3.1.1. *Let $p_1 = (x_0, y_0)$ and $p_2 = (x_1, y_1)$ be points in the plane.*

We say $p_0 < p_1$, if $x_0 < y_0$ or ($x_0 = x_1$ and $y_0 < y_1$). We define $>$ similarly; $\leq$ and $\geq$ are defined in the usual way.

Observe that this is a total order on points in the plane.

Notation 3.1.2. *Let S be a line segment. We denote the two endpoints of S by $S[0]$ and $S[1]$ where $S[0] \leq S[1]$.*

Example 3.1.3. Let $S = \overline{(20, 0)(20, -1)}$, then $S[0] = (20, -1)$.

Notation 3.1.4. *Let S_0 and S_1 be line segments. We say $S_0 < S_1$, if $S_0[0] < S_1[0]$ or ($S_0[0] = S_1[0]$ and $S_0[1] < S_1[1]$).*

Observe that this is a total order on line segments in the plane.

Definition 3.1.5 (cf. [5])**.** A set is *convex* if it contains all the line segments between any pair of its points.

Definition 3.1.6 (cf. [5])**.** Let X be a bounded set in the plane, we define the *convex hull* to be

$$CH(X) = \bigcap_{\substack{X \subseteq H \\ H \text{ is convex}}} H.$$

3.2 Minkowski Sum

In this section, we use Minkowski sums to build up a foundation for a more practical way to compute perimeter-distance.

Definition 3.2.1. Let A and B be sets, then the *Minkowski sum* of A and B is the set

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

Remark 3.2.2. It follows immediately from the definition of the Minkowski sum that

$$A + (B \cup C) = (A + B) \cup (A + C).$$

We can obtain more useful geometric interpretation for certain Minkowski sums. Let S be a line segment and $\tau > 0$, define B_τ to be the disk centered at the origin of radius τ. Then, geometrically, we know that

$$B_\tau + S = \{x \in \mathbb{R}^2 : d(x, S) \leq \tau\}, \tag{3.1}$$

see Figure 3.1.

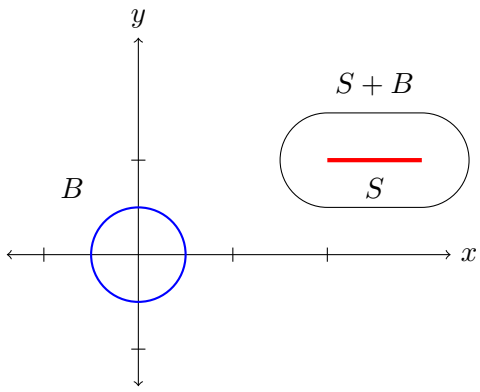


Figure 3.1: Minkowski Sum: The Minkowski sum S+B is shown, where B is the unit disk centered at the origin and S is the line segment.

Proposition 3.2.3. *Let $\tau > 0$ where P and $\mathcal{D}$ as in Definition 1.2.2. Let B_τ be a closed disk of radius τ centered at the origin. Then*

$$P_{0,\tau}(\mathcal{D}) = \bigcup_{P_i \in \mathcal{P}} ((D^-_{P_i} + B_\tau) \cap P_i).$$

Proof. Let $x \in \bigcup_{P_i \in \mathcal{P}}((D^-_{P_i} + B_\tau) \cap P_i)$. Without loss of generality, we may assume $x \in (D^-_{P_i} + B_\tau) \cap P_i$. Equivalently, there exists y in $D^-_{P_i}$ such that $d(x, y) \leq \tau$. By our assumption, $D^-_{P_i} = D^-_x$; so, $d(x, D^-_x) \leq \tau$. Because $x \cap D^-_x = \emptyset$, we have that $0 < d(x, D^-_x) \leq \tau$. And, $0 <$

$d(x, D_x^-) \leq \tau$ if and only if $x \in P_{0,\tau}(\mathcal{D})$. Therefore, $\bigcup_{p_i \in P}((D_{P_i}^- + B_\tau) \cap P_i) = P_{0,\tau}(\mathcal{D})$. $\square$

Proposition 3.2.4. *Let B_τ be a disk centered at the origin of radius τ and S a line segment. The boundary of $B_\tau + S$ is given by two perpendicular translations of the line segments and two half circles of radius τ centered at S[0], S[1].*

Proof. Let $\overrightarrow{d}$ be a vector in the plane. We use the fact that the boundary points of the Minkowski sum which are maximal in the direction $\overrightarrow{d}$ are given by the sum of points on the boundary of each object maximal in the direction $\overrightarrow{d}$, see [9]. Each point in the line segment S is maximal in both perpendicular directions to the supporting line of S, see Figure 3.2(a); we let $\pm S_d$ be unit vectors in the perpendicular directions to the supporting line of S. In these directions, there are unique maximal points in B_τ, since the boundary of B_τ is a circle, see Figure 3.2(b). Hence, we obtain translations of S by a magnitude of τ.

$S[0]$ is maximal in the direction of any $\overrightarrow{d}$ such that the dot product of $\overrightarrow{d}$ and $\overrightarrow{S[1]S[0]}$ is positive, see Figure 3.2(b). These directions correspond to the half-circle $H_{S[0]} \subseteq B_\tau$. $H_{S[1]}$ is defined similarly. Then $H_{S[0]} + S[0]$ and $H_{S[1]} + S[1]$ are half circles which complete the closed boundary of $B_\tau + S$, see Figure 3.2(c). $\square$

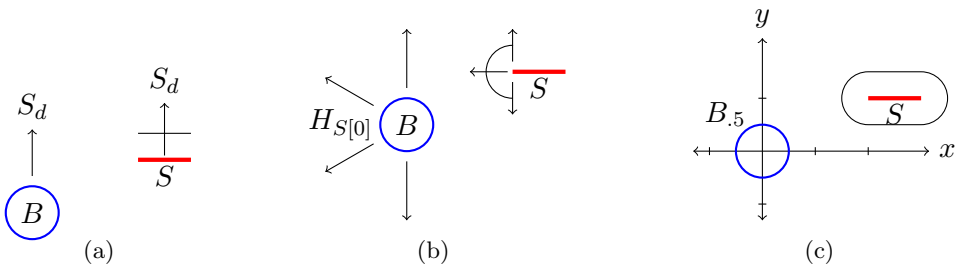


Figure 3.2: The Boundary of a Minkowski Sum: In Proposition 3.2.4, we describe the boundary of the Minkowski sum of the line segment S and the disk B. In (a), we can see how all points along the line segment S are maximal in the two perpendicular directions to S, we show one perpendicular direction S_d. For each point in the segment S, we add the point of the circle which is maximal in the same direction to form the perpendicular translation of the line segment S. In (b), we see that the endpoints are also maximal in the direction of the line segment, and all positive combinations, as shown. In (c) we complete the process for the disk of radius .5, $B_{.5}$, and the line segment $S = \overline{(2,1),(3,1)}$.

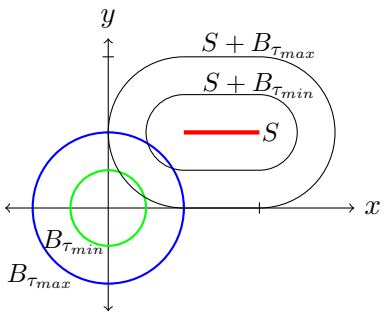


Figure 3.3: Minkowski Sums of a Line Segment and Two Disks: The boundary of $B_{\tau_{max}}$ is the outer blue circle, the boundary of $B_{\tau_{min}}$ is the inner circle, shown in green, and S is the red line segment. The Minkowski sum $S + B_{\tau_{max}}$ contains $S + B_{\tau_{min}}$. The region bound between the two pill-shapes are the points whose distance from S is between τ_{min} and τ_{max}.

Proposition 3.2.5. *Let Q be a closed and bounded convex set which has a nontrivial interior. Let S be a line segment. Let l be the supporting line segment of S and ∂Q be the boundary of Q. Then, ∂Q determines $S \cap Q$ in the following way.*

1. *If $l \cap \partial Q$ is a point, then $S \cap Q$ is either the same point or the empty set.*
2. *If $l \cap \partial Q$ contains exactly two points p and q then $S \cap Q \subseteq \overline{pq}$.*
3. *$l \cap \partial Q$ contains more than two points if and only if $l \cap Q$ is forms an edge of ∂Q and $S \cap Q \subseteq l \cap \partial Q$.*

Proof. We observe that $S \cap Q \subseteq l \cap Q$. Since Q is convex and bounded, $l \cap Q = CH(l \cap \partial Q)$.

Assume first $l \cap \partial Q$ is a point. Because $l \cap Q = CH(l \cap \partial Q)$, $l \cap Q$ is a point. Since $S \cap Q \subseteq l \cap Q$, $S \cap Q$ is either the point $l \cap \partial Q$ or the empty set.

Assume second that $l \cap \partial Q$ contains exactly two points p and q. Then since $l \cap Q = CH(l \cap \partial Q)$, $\overline{pq} = l \cap Q$. Since $S \cap Q \subseteq l \cap Q$, $S \cap Q \subseteq \overline{pq}$.

For the final statement, the converse direction is obvious. If $l \cap \partial Q$ is a line segment then it contains the two endpoints and the midpoint.

On the other hand, if $R := l \cap \partial Q$ contains more than two points. Define $R[0] := \min\{l \cap \partial Q\}$ the minimal element of the set R and $R[1] := \max\{l \cap \partial Q\}$ the maximal element. Then let $r^* \in R$ be the third guaranteed point where $R[0] < r^* < R[1]$. We

know that $\overline{R[0], R[1]} \subseteq Q$. Recall that Q is not a line segment. Therefore, we may let x be a point not on the line l and in the interior of Q. Then x is in the interior of one of the half planes determined by l. If $l \cap \partial Q$ is not an edge of ∂Q, then there exists y in the interior of Q in the other half plane determined by l. But then r^* is in the interior of $CH(\{R[0], R[1], x, y\})$. But, $CH(\{R[0], R[1], x, y\}) \subseteq Q$ and thus r^* is in the interior of Q, a contradiction. Therefore, $l \cap Q = l \cap \partial Q$ is an edge of Q. Since $S \cap Q \subseteq l \cap Q$, it follows that $S \cap Q \subseteq l \cap \partial Q$. □

We have seen in Proposition 3.2.3 that perimeter-distance can be computed by intersections of line segments and Minkowski sums of disks with line segments. In the following proposition, we see that these computations can be performed with geometric operations on lines and circles.

Proposition 3.2.6. *Let $\tau > 0$, B_τ be the disk centered are the origin of radius τ. Let S_1 and S_2 be line segments. Then $S_1 \cap (B_\tau + S_2)$ can be determined using a ruler and compass.*

Proof. The boundary of $B_\tau + S_2$, $\partial(B_\tau + S_2)$, is given by two line segments and two half circles by Proposition 3.2.4. We consider two cases $S_1 \cap \partial(B_\tau + S_2)$ is the empty set or is not the empty set. If the intersection is the empty set then by the convexity of $B_\tau + S_2$ the segment S_1 is either contained within or entirely outside $(B_\tau + S_2)$.

On the other hand, suppose $S_1 \cap \partial(B_\tau + S_2)$ is not the empty set. Let l be the supporting line of S_1. We now consider three further cases; either $l \cap \partial(B_\tau + S_2)$ consists of one point, two points or more than two points.

If $e := l \cap \partial(B_\tau + S_2)$ consists of one point, then $S_1 \cap (B_\tau + S_2) \subseteq e$ by Proposition 3.2.5. Since $S_1 \cap \partial(B_\tau + S_2)$ is not the empty set, then $S_1 \cap (B_\tau + S_2) = e$.

If $l \cap \partial(B_\tau + S_2)$ consists of two points p and q where $p < q$, then $S_1 \cap (B_\tau + S_2) \subseteq \overline{pq}$ by Proposition 3.2.5. If $S_1 \neq \overline{pq}$, then because $p, q \in \partial(B_\tau + S_2)$ and $S_1 \cap \partial(B_\tau + S_2)$ is not the empty set then either $p < S_1[0]$ or $S_1[1] < q$. Hence, $S_1 \cap (B_\tau + S_2) = \overline{S_1[0]q}$ or $\overline{pS_1[1]}$, receptively.

If $l \cap \partial(B_\tau + S_2)$ consists of more than two points then by Proposition 3.2.5, $S_1 \cap (B_\tau + S_2)$ is a subset of one of the edges of $(B_\tau + S_2)$. Let E be the edge of $(B_\tau + S_2)$ such that $S_1 \cap (B_\tau + S_2) \subseteq E$. First, if $S_1 \neq E$ and if neither endpoint of E is in S, i.e., $E_1[0] < S_1[0]$ and $S_1[1] < E_1[1]$, then $S_1 \cap (B_\tau + S_2) = S_1$. Second, if $S_1 \neq E$ and either $E_1[0] < S_1[0]$ or $S_1[1] < E_1[1]$, then $S_1 \cap (B_\tau + S_2) = \overline{S_1[0], E[1]}$ or $\overline{E[0], S_1[1]}$, respectively. Third, if $S_1 \neq E$, then the last case is $S_1[0] < E_1[0]$, and $E_1[1] < S_1[1]$. Hence, $E = S_1 \cap (B_\tau + S_2)$

We have determined $S_1 \cap (B_\tau + S_2)$ in each case using only intersections of lines and circles. These intersections can be performed using ruler and compass, as desired. □

3.3 Minkowski Sums and Perimeter-Distance

In this section, we show how to reduce perimeter-distance to many simple calculations of finding the intersection of a line segment and a Minkowski sum of a disk and line segment.

Proposition 3.3.1. *Let $0 < \tau_{min} < \tau_{max}$. Let P and $\mathcal{D}$ as in Definition 1.2.2. Then*

$$\ell_{\tau_{min},\tau_{max}}(\mathcal{D}) = \ell_{0,\tau_{max}}(\mathcal{D}) - \ell_{0,\tau_{min}}(\mathcal{D}).$$

Proof. By Proposition 2.5.3,

$$P_{\tau_{min},\tau_{max}}(\mathcal{D}) = P_{0,\tau_{max}}(\mathcal{D}) \setminus P_{0,\tau_{min}}(\mathcal{D}).$$

By Proposition 3.2.3,

$$P_{\tau_{min},\tau_{max}}(\mathcal{D}) = \bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})) \setminus \bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{min}})) \tag{3.2}$$

by Proposition 2.5.3.

We now consider only the first union of the right hand side,

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})).$$

Let $e_{i,j}$ be the jth edge of polygon P_i. Hence,

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} (e_{i,j} \cap (D^-_{P_i} + B_{\tau_{max}})).$$

Let $e_{k,l}$ be the lth edge of polygon $P_k \in D^-_{P_i}$.

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} \left(e_{i,j} \cap \left(\bigcup_{P_k \in D^-_{P_i}} \bigcup_{e_{k,l} \subseteq P_k} e_{k,l} + B_{\tau_{max}} \right) \right).$$

Then,

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} \bigcup_{P_k \in D^-_{P_i}} \bigcup_{e_{k,l} \subseteq P_k} (e_{i,j} \cap (e_{k,l} + B_{\tau_{max}})). \tag{3.3}$$

The constructive process to compute the intersection of a line segment with the Minkowski sum of a line segment and a disk centered at the origin of fixed radius is described in Proposition 3.2.6. The inner union over k and l in Equation (3.3) is over the intersection of several Minkowski sums, $(e_{k,l} + B_{\tau_{max}})$ with the same line segment $e_{i,j}$. We define

$$E_{i,j}(\tau_{min}) := \bigcup_{P_k \in D^-_{P_i}} \bigcup_{e_{k,l} \subseteq P_k} (e_{i,j} \cap (e_{k,l} + B_{\tau_{max}})).$$

By Proposition 4.2.1 we explicitly obtain the maximal disjoint line segments of $E_{i,j}(\tau_{min}) \subseteq e_{i,j}$, $\mathcal{E}_{i,j}(\tau_{min})$.

The same construction yields $\mathcal{E}_{i,j}(\tau_{max})$. In this way, each segment $e_{k,l}$ is used to form two Minkowski sums $e_{k,l} + B_{\tau_{max}}$ and $e_{k,l} + B_{\tau_{min}}$. Observe, each $e_{k,l} + B_{\tau_{max}}$ and $e_{k,l} + B_{\tau_{min}}$ intersect the same line segment, $e_{i,j}$, see Figure 3.4, where each intersection is

contained in some segment from $\mathcal{E}_{i,j}(\tau_{min})$ and $\mathcal{E}_{i,j}(\tau_{max})$, respectively.

Then we have

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{max}})) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} \mathcal{E}_{i,j}(\tau_{max}),$$

and,

$$\bigcup_{P_i \in P} (P_i \cap (D^-_{P_i} + B_{\tau_{min}})) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} \mathcal{E}_{i,j}(\tau_{min}).$$

By substitution into Equation 3.2, we obtain

$$P_{\tau_{min},\tau_{max}}(D) = \bigcup_{P_i \in P} \bigcup_{e_{i,j} \subseteq P_i} (\mathcal{E}_{i,j}(\tau_{max}) \setminus \mathcal{E}_{i,j}(\tau_{min})).$$

Notice that the line segments of $\mathcal{E}_{i,j}(\tau_{max})$ and $\mathcal{E}_{i,j}(\tau_{min})$ are contained on the jth edge of polygon $P_i \in \mathcal{P}$. Because the edges of the fiber-polygons of $\mathcal{P}$ are pairwise disjoint,

$$\ell_{\tau_{min},\tau_{max}}(D) = \sum_i \sum_j \text{length}(\mathcal{E}_{i,j}(\tau_{max})) - \sum_i \sum_j \text{length}(\mathcal{E}_{i,j}(\tau_{min})).$$

Equivalently,

$$\ell_{\tau_{min},\tau_{max}}(D) = \sum_{S_k \in \bigcup_{i,j} \mathcal{E}_{i,j}(\tau_{max})} \text{length}(S_k) - \sum_{S_l \in \bigcup_{i,j} \mathcal{E}_{i,j}(\tau_{min})} \text{length}(S_l).$$

By Definition 2.4.4,

$$\ell_{\tau_{min},\tau_{max}}(\mathcal{D}) = \ell_{0,\tau_{max}}(\mathcal{D}) - \ell_{0,\tau_{min}}(\mathcal{D}).$$

□

We now have a construction to find the total length of the channel-walls. The length of the interstice walls which form interstices which are narrow enough to allow macromolecules to diffuse to the walls is calculated first. Then the length of the interstice walls which form interstices which are too narrow to allow macromolecules to diffuse to the

walls is calculated is subtracted from the previous sum. Using this computation, we have a quantification of the utility of the column environment.

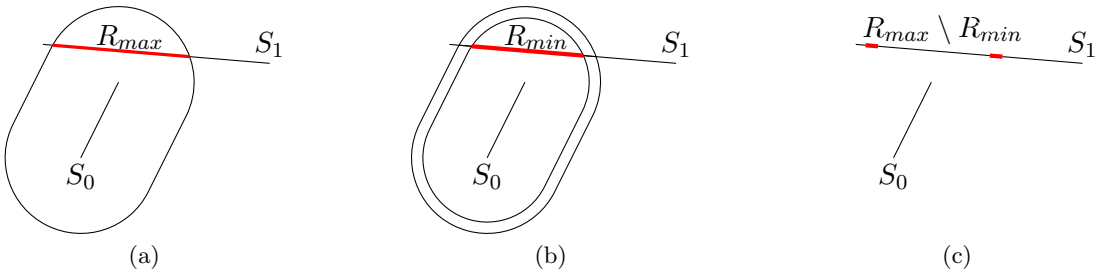


Figure 3.4: Computing Segments for $P_{\tau_{min},\tau_{max}}(\mathcal{D})$: In these figures, two segments are shown where S_0 is an edge taken from a fiber-polygon or the column wall and S_1 is taken from a disjoint fiber-polygon. In (a), the boundary of the Minkowski sum of the line segment S_0 and a disk centered at the origin of radius τ_{max} is shown. R_{max} (in red) is the part of the line segment S_1 which is within τ_{max} of the line segment S_0, i.e., R_{max} is included in $P_{0,\tau_{max}}(\mathcal{D})$. In (b), the boundary of the previous Minkowski sum along with the boundary of a smaller Minkowski sum is shown. The smaller boundary is of the Minkowski sum of the same line segment, S_0, with a disk centered at the origin of radius τ_{min}, i.e., R_{min} is the part of S_1 which is within τ_{min} of the line segment S_0. R_{min} is included in $P_{0,\tau_{min}}(\mathcal{D})$. Moreover, $R_{min} \subseteq R_{max}$. In (c), we show the maximal segments of $R_{max} \setminus R_{min}$.

Chapter 4

Practical Computation of Perimeter-Distance

Equation (3.3), under the conditions of Proposition 3.3.1, and Proposition 3.2.6 together provide an indication that our definition of perimeter-distance can be computed algorithmically. If we implement this directly with a brute force approach, there are a large number of intersections which are unnecessary. We provide a heuristic in Section 4.1 to reduce the computational resources needed to compute perimeter-distance. In Section 4.2 we provide the important functions used to compute perimeter-distance on the model and the results of these computations.

4.1 A Heuristic for Perimeter-Distance

Following Equation (3.3), the number of intersections of a line segment and a Minkowski sum of a disk and line segment needed to compute perimeter-distance is $O(n^2m^2)$ where n is the number of fiber-polygons in $\mathcal{P}$ and m is the number of edges in each fiber-polygon. We describe a comparatively cheap test between pairs of polygons in $\mathcal{P}$, consisting of $O(n^2)$ tests, to reduce the number of intersections of a line segment and a Minkowski sum of a disk and line segment from $O(n^2m^2)$ to $O(nm^2)$. The test is based on the condition

in Corollary 4.1.8 which gives a sufficient spacing for fiber-polygons to exclude them from perimeter-distance calculations.

Definition 4.1.1. Let X be a bounded subset of the plane. An axis-aligned rectangle in the plane is determined by two points, the lower left and upper right corner, LL and UR respectively. We denote the axis-aligned rectangle determined by points LL and UR by $R(LL, UR)$. We define

$$BBox(X) := \bigcap_{\substack{X \subseteq B \\ B \text{ is an axis-} \\ \text{aligned rectangle}}} B$$

and we denote the center of $BBox(X)$ by c_X.

Example 4.1.2. Let C be a circle centered at the origin with radius r. $BBox(C) = R((-r, r), (r, -r))$.

Remark 4.1.3. If X and Y are bounded sets in the plane, then

$$d_{max}(X, Y) \leq d_{max}(BBox(X), BBox(Y)).$$

Lemma 4.1.4. *Let τ, s, α and $\beta > 0$. If*

$$\tau^2 < \frac{s^2 - 3\alpha^2 - 3\beta^2}{3}$$

then

$$\tau < s - \alpha - \beta.$$

Proof. $\tau^2 < \frac{s^2 - 3\alpha^2 - 3\beta^2}{3}$ if and only if

$$3\left(\tau^2 + \alpha^2 + \beta^2\right) < s^2.$$

By the Cauchy-Schwartz inequality, $(\tau \cdot 1 + \alpha \cdot 1 + \beta \cdot 1)^2 \leq (\tau^2 + \alpha^2 + \beta^2)(1^2 + 1^2 + 1^2)$. Therefore,

$$(\tau + \alpha + \beta)^2 \leq 3\left(\tau^2 + \alpha^2 + \beta^2\right).$$

Hence, $(\tau + \alpha + \beta)^2 < s^2$. Because all the terms are positive, we obtain $(\tau + \alpha + \beta) < s$ Therefore, $\tau < s - \alpha - \beta$. □

Proposition 4.1.5. *Let X and Y be bounded sets in the plane. Let $BBox(X)$ and $BBox(Y)$ have centers c_X and c_Y. If*

$$\tau^2 < \frac{d(c_X, c_Y)^2 - 3d_{max}(c_X, BBox(X))^2 - 3d_{max}(c_Y, BBox(Y))^2}{3},$$

then,

$$\tau < d(X, Y).$$

Proof. Let C_X be the disk with center c_X and radius $d_{max}(c_X, BBox(X))$, let C_Y and c_Y be defined similarly. $BBox(X) \subseteq C_X$ because c_X is the center of the axis-aligned rectangle $BBox(X)$, see Figure 4.1. Similarly, $BBox(Y) \subseteq C_Y$. Therefore, $X \subseteq C_X$ and $Y \subseteq C_Y$. Hence, $d(C_X, C_Y) \leq d(X, Y)$.

By Lemma 4.1.4,

$$\tau < d(c_X, c_Y) - d_{max}(c_X, BBox(X)) - d_{max}(c_Y, BBox(Y)).$$

The minimum distance between circles is the distance from the respective centers minus the radii, hence, $d(c_X, c_Y) - d_{max}(c_X, BBox(X)) - d_{max}(c_Y, BBox(Y)) = d(C_X, C_Y)$. Therefore,

$$\tau < d(C_X, C_Y).$$

We already showed, $d(C_X, C_Y) \leq d(X, Y)$. Thus,

$$\tau < d(X, Y),$$

as desired. □

In the previous proposition, we presented a test to determine when two sets are far

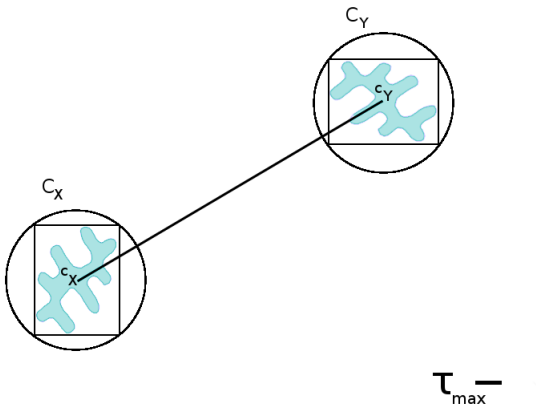


Figure 4.1: Bounding Circles: There are two fiber-polygons shown with bounding boxes of the respective fiber-polygons. The distance between the centers of the bounding boxes c_X and c_Y of the fiber-polygons is easy to compute. In Proposition 4.1.5, we make use of the fact that the disks C_X and C_Y contain both the boxes and fiber-polygons.

apart using bounding boxes. In the remaining part of this section, we show how this test reduces the complexity of the algorithm used when computing perimeter-distance.

Lemma 4.1.6. *Let α be the area of a fiber-polygon. Let $\mathcal{P}$ be a collection of fiber-polygons in the plane. Let $\tau > 0$, and B_τ be a disk of radius τ in the plane. Then, the number of polygons completely contained in the disk bounded above by $\lfloor \frac{\pi\tau^2}{\alpha} \rfloor$.*

Proof. The area of the polygons completely contained in the disk cannot exceed the total area of the disk because the polygons are disjoint. The area of the disk is $\pi\tau^2$. If the fiber-polygons cover the disk, $\lfloor \frac{\pi\tau^2}{\alpha} \rfloor$ is the number of whole fiber-polygons in the disk, our upper bound. □

We can use the above lemma to determine the number of fiber-polygons which intersect a disk of a given size. We use the following result in Corollary 4.1.8.

Lemma 4.1.7. *Let α be the area of a fiber-polygon and η be the diameter of fiber-polygon. Let $\mathcal{P}$ be a collection of fiber-polygons in the plane. Let $\tau > 0$, and B_τ be a disk of radius*

τ in the plane. Then, the number of polygons which intersect the disk B_τ is bounded above by $\lfloor \frac{\pi(\eta+\tau)^2}{\alpha} \rfloor$.

Proof. From any point on a fiber-polygon, all points of the fiber polygon are within η of that point. If a point of the fiber-polygon is inside the disk, B_τ, then all points of the fiber-polygon are within η of the disk, B_τ, see Figure 4.2. But then they are contained entirely inside $B_{\eta+\tau}$, the disk with radius $\eta+\tau$, which has the same center as B_τ. By Lemma 4.1.6, the number of polygons which intersect the disk, B_τ, is bounded above by $\lfloor \frac{\pi(\eta+\tau)^2}{\alpha} \rfloor$, as desired. □

Corollary 4.1.8. *Let $\mathcal{P}$ and $\mathcal{D}$ as in Definition 1.2.2. Let η be the diameter of a fiber-polygon and let α be the area of a fiber polygon. Let $\tau_{max} > 0$. Let P_i be a fiber polygon in $\mathcal{P}$. Let $\mathcal{N}$ be the set of polygons which pass the bounding box heuristic from P_i, $\mathcal{N} := \{P_j \in \mathcal{P} : P_j \neq P_i$ and $d(c_{P_i}, c_{P_j})^2 - 3d_{max}(c_{P_i}, BBox(P_i))^2 - 3d_{max}(c_{P_j}, BBox(P_j))^2 \leq 3\tau_{max}^2\}$, where c_{P_i} is the center of $BBox(P_i)$ and similarly for c_{P_j}. Let N be the number of polygons in $\mathcal{N}$. Then,*

$$N \leq \left\lfloor \frac{\pi(\eta + (3(\tau_{max}^2 + \eta^2/2))^{1/2})^2}{\alpha} \right\rfloor.$$

Proof. Let $P_j \in \mathcal{N}$. If $d(c_{P_i}, c_{P_j})^2 - 3d_{max}(c_{P_i}, BBox(P_i))^2 - 3d_{max}(c_{P_j}, BBox(P_j))^2 \leq 3\tau_{max}^2$ then, $d(c_{P_i}, c_{P_j})^2 \leq 3(\tau_{max}^2 + d_{max}(c_{P_i}, BBox(P_i))^2 + d_{max}(c_{P_j}, BBox(P_j))^2)$. Because the diameter of the fiber-polygons is greater than or equal to the lengths of the sides of the bounding box, it follows that, $d(c_{P_i}, c_{P_j})^2 \leq 3(\tau_{max}^2 + (\sqrt{2}\eta/2)^2 + (\sqrt{2}\eta/2)^2)$. Both sides of the inequality are squares of positive terms therefore, $d(c_{P_i}, c_{P_j}) \leq (3(\tau_{max}^2 + \eta^2))^{1/2}$. This is an upper bound on how far the centers of the boxes can be from one another. Then the disk B_β of radius $(3(\tau_{max}^2 + \eta^2))^{1/2}$ centered at c_{P_i} contains some point of $BBox(P_j)$.

By Lemma 4.1.7, the upper bound on the number of $P_j \in \mathcal{P}$ such that $P_j \cap B_\beta \neq \emptyset$ is given by N where

$$N \leq \left\lfloor \frac{\pi(\eta + (3(\tau_{max}^2 + \eta^2/2))^{1/2})^2}{\alpha} \right\rfloor.$$

□

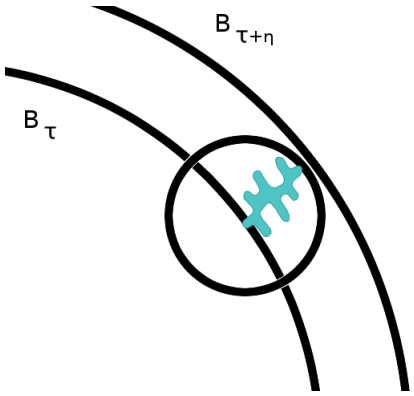


Figure 4.2: Partial Fiber-Polygons in a Disk: The radius of the inner disk, B_τ, is τ. The outer disk, $B_{\tau+\eta}$, has been extended by the diameter of a fiber-polygon, η. This shows that all fiber-polygons which intersect with the inner disk are totally contained in the larger disk.

In Corollary 4.1.8, we showed that for $\tau_{max} > 0$ the number of polygons within τ_{max} of a fiber polygon is a constant, $\mathcal{N}$, which depends on τ_{max}. Therefore, the number of pairs of intersections for every polygon is $\mathcal{N}$ times m^2, where m is the number of edges in the polygon, i.e., the complexity of the algorithm is $O(n \cdot m^2)$ in intersections. Therefore, the complexity of the algorithm is $O(n^2 + nm^2)$, where the first additive term is for the test in Proposition 4.1.5 and the later additive term is for the intersections between pairs of edges. In practice, the test exhibits a great reduction in the computational resources needed to compute perimeter distance.

4.2 Algorithms

In Chapter 3, we showed theoretical methods to compute perimeter-distance. This section contains pseudo-code to illustrate how to compute perimeter-distance algorithmically.

Let S_0 and S_1 be two line segments. Let B_τ be a disk centered at the origin. In Algorithm 3, we compute the intersection of a line segment and the Minkowski sum of a disk and a line segment, $S_0 \cap (B_\tau + S_1)$. This follows Proposition 3.2.6.

In Algorithm 2, we compute the collection of maximal line segments used in Propo-

sition 4.2.1. In Algorithm 1, we see the algorithm of our heuristic, which reduces the number of intersections needed to compute perimeter-distance. The amount of time needed to compute the perimeter-distance $\ell_{\tau_{min},\tau_{max}}(\mathcal{D})$, where the data for $\mathcal{D}$ is taken from [7] is shown in Figure 1.2, without the bounding-box heuristic the computation was projected to take several days using a laptop with a AMD A6-5200 processor. By implementing our heuristic, the length of time to compute $\ell_{\tau_{min},\tau_{max}}(\mathcal{D})$ decreased to about 20 minutes on the same machine for most practical values of τ_{min} and τ_{max}.

Algorithm 1: Perimeter-Distance Algorithm: The first nested loop calls push_back $O(n^2)$ times where n is the number of polygons in $\mathcal{P}$. This algorithm performs a constant number of push_backs that can occur for each polygon. The effect of implementing the test is that Perimeter_Distance() calls intersect_minkowski() $O(nm^2)$ times.

```
Function Perimeter_Distance is
    Input: tausqrd_min, tausqrd_max
    domain //The fiber-polygons and the column wall, D
    Output: perimeter-distance //ℓ_{τmin,τmax}(D)
    for poly in fiber_polygons do
        //In this loop, every fiber-polygon is assigned a list of nearby polygons
        for neighbor_poly in domain do
            if poly ≠ neighbor_poly and
            box_distance_sqrd(poly, neighbor_poly)< tausqrd_max then
                polys_near_to_current_poly.push_back(poly);
            end
        end
        //In this loop, we find S_{τmin,τmax}(D).
        for edge in polygon do
            max = subsegment(polys_near_to_current_poly, edge, tausqrd_max);
            min = subsegment(polys_near_to_current_poly, edge, tausqrd_min);
            edge_segments_in_S_min_max = max_take_away_min(max, min);
            for segment in edge_segments_in_S_min_max do
                segments.push_back(segment);
            end
        end
    end
    Return total_length(segments);
end
```

In the following proposition, we show how to find the maximal line segments in a union of many line segments, as illustrated in Figure 4.3.

Algorithm 2: Subsegment Function: This function uses calculations between pairs of edges to compute perimeter-distance.

```
Function subsegment is
    Input: nearby_polygons, tausqrd, current_edge (using CE for current_edge)
    Output: A collection of line segments which are within tau of a point on a
            polygon in nearby_polygon
    for polygon in nearby_polygons do
        for E in polygon do
            if d(E, CE) ≤ tausqrd then
                subsegs_of_CE.push_back(intersect_minkowski(E,CE,tausqrd));
            end
        end
    end
    Return union_subsegments(order_subsegments(subsegs_of_CE));
end
```

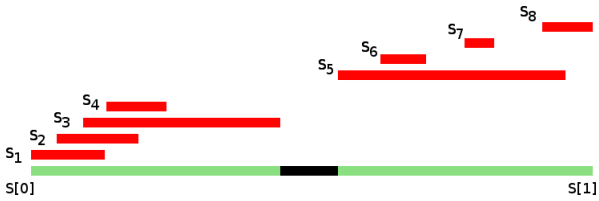


Figure 4.3: The Union of Line Segments: In this figure, we illustrate how to form maximal line segments from a union of overlapping ordered line segments. The segments $S_1, \ldots, S_8$ are subsegments of the lower line segment, S. S has endpoints $S[0]$ and $S[1]$. The lightly green shaded parts of the lower line segment represent the maximal line segments of the union of the smaller line segments. The algorithm to find the maximal line segment is explained in Proposition 4.2.1.

Proposition 4.2.1. *Let S be a line segment with supporting line l. Let $S_1 \leq \cdots \leq S_n$ be line segments on the line segment S, i.e., $S_i \subseteq S$ for $i = 1, \ldots, n$. Suppose there are exactly k integers $1 \leq a_1 < \cdots < a_k < n$ such that for all $j = 1, \ldots, k$ and $i = 1, \ldots, a_j$, $S_i[1] < S_{(a_j)+1}[0]$. Then there are $k+1$ maximal line segments in $\bigcup_{i=1}^{n} S_i$.*

Furthermore, let $a_0 = 0$ and $a_{k+1} = n$. We show the starting point of the $k+1$ maximal segments are given by $S_1[0]$ and $S_{(a_j)+1}[0]$ for $j = 1, \ldots, k$ and the terminal point for the $k+1$ maximal segments are given by

$$m_j = \max \{S_i[1]\}_{(a_j)+1 \leq i < a_{(j+1)}}$$

Algorithm 3: The Intersection of a Line Segment and a Minkowski Sum of a Line Segment and a Disk.

```
Function intersect_minkowski is
    Input: minkowski_segment, current_edge, tausqrd
    Output: A line segment which represents the points on current_edge which
            are within tau of minkowski_segment
    MC = The pair of circles with centers at the respective endpoints of
    minkowski_segment and of squared radius tausqrd;
    ME = The pair of segments which form the two straight edges of the
    Minkowski sum of the disk centered at the origin of squared radius tausqrd
    and minkowski_segment;
    if d(ME intersect current_edge)>0 then
        Return ME intersect current_edge;
    else
        if The squared distance of either end points of current_edge to ME is
        less than tausqrd then
            return_points.push_back(respective endpoints)
        end
        if return_points has both endpoints of current_edge then
            Return current_segment
        end
        circle_points = intersections of MC with current_edge;
        for point in circle_points do
            if squared distance from point to current_edge is equal to tausqrd
            then
                return_points.push_back(point);
            end
        end
        return_points.push_back(intersections of ME with current_edge);
        min_point=Min( return_points );
        max_point=Max( return_points );
        Return  The line segment with endpoints min_point and max_point;
    end
end
```

where $j = 0, \ldots, k$*; and,*

$$\bigcup_{i=1}^{n} S_i = \bigcup_{i=0}^{k} \overline{S_{a_i+1}[0]m_i}.$$

Proof. Let $x \in \bigcup_{i=1}^{n} S_i$ without loss of generality let $x \in S_i$ where $S_{a_j+1} \leq S_i \leq S_{a_{(j+1)}}$. Because $S_{a_j+1} \leq S_i$, $S_{a_j+1}[0] \leq S_i[0]$. By the construction of m_j, $S_i[1] \leq m_j$. Therefore, $S_i \subseteq \overline{S_{a_j+1}m_j}$. It follows that $\bigcup_{i=1}^{n} S_i \subseteq \bigcup_{i=0}^{k} \overline{S_{a_i+1}[0]m_i}$.

For the converse containment, let $x \in \bigcup_{i=0}^{k} \overline{S_{a_i+1}[0]m_i}$. Without loss of generality, $x \in \overline{S_{a_i+1}[0]m_i}$. By construction, $S_{a_i+1}[0] \leq x \leq m_i$, but we only consider the strict inequalities for otherwise the containment holds obviously, i.e., $S_{a_i+1}[0] < x < m_i$. By the construction of $\overline{S_{a_i+1}[0]m_i}$, we know that $\bigcup_{q=a_i+1}^{a_{(i+1)}} S_q \subseteq \overline{S_{a_i+1}[0]m_i}$. We show that this containment is an equality to complete our proof.

Assume, for the purpose of a contradiction, $x \notin \bigcup_{q=a_i+1}^{a_{(i+1)}} S_q$. Let $m_i = S_{a_i+h+1}[1]$, then $0 \leq h \leq (a_{(i+1)} - ((a_i) + 1))$.

Then from $x \notin \bigcup_{q=a_i+1}^{a_{(i+1)}} S_q$, it follows that $x < S_{a_i+h+1}[0]$. Let p be the least element of the set $\{a_i+1, \ldots, a_{(i+1)}\}$ such that $x < S_p[0]$. Then $x > S_z[0]$ for all $z = a_i+1, \ldots, p-1$. Since $x \notin \bigcup_{q=a_i+1}^{a_{(i+1)}} S_q$, it follows that $x > S_z[1]$ for all $z = a_i + 1, \ldots, p - 1$. Therefore, $S_z[1] < S_p[0]$ for $z = a_i + 1, \ldots, p - 1$.

From the statement of this proposition, for all $z = 1, \ldots, a_i$, $S_z[1] < S_{a_i+1}[0]$. Hence, $S_z[1] < S_p[0]$ for $z = 1, \ldots, p - 1$.

Because $p \neq a_i$ for $i = 1, \ldots, k$, there are $k + 1$ integers such that $1 \leq a_1 < \cdots < a_{k+1} < n$ such that for all $j = 1, \ldots, k$ and $i = 1, \ldots, a_j$, $S_i[1] < S_{(a_j)+1}[0]$, a contradiction to conditions of this Proposition. Thus $x \in \bigcup_{q=a_i+1}^{a_{(i+1)}} S_q$ and so $x \in \bigcup_{i=1}^{n} S_i$ which gives the equality, $\bigcup_{q=a_i+1}^{a_{(i+1)}} S_q = \overline{S_{a_i+1}[0]m_i}$.

Therefore, $\bigcup_{i=0}^{k} \overline{S_{a_i+1}[0]m_i}$ are the $k + 1$ maximal line segments of $\bigcup_{i=1}^{n} S_i$. □

In Chapter 1, we presented our model of the domain $\mathcal{D}$ in Figure 1.2, data from [7]. In Figure 4.4, we present the values of $\ell_{0,\tau_{max}}(\mathcal{D})$ for $\tau_{max} = 0, .05, \ldots, 15.6\ \mu m$ and we show that $\ell_{0,15.6\ \mu m}(\mathcal{D}) = Perimeter(\mathcal{P})$. In Figure 4.5, we show the marginal increase of $\ell_{0,\tau_{max}}$ for step size .05 μm. The marginal increase in perimeter-distance is $\ell_{\tau_{max}-.05,\tau_{max}}(\mathcal{D})$,

where $0 \leq \tau_{max} \leq 15.6$ by a step size of .05 μm. We see that there are diminishing marginal increasing value for $\ell_{0,\tau_{max}}(\mathcal{D})$ for $2.09 \leq \tau_{max}$.

We can summarize the data shown in Figure 4.5 by using the statistical software program R [12]. In order to use R, we created sample data designed to mimic the marginal increases in perimeter-distance given by $\ell_{\tau_{max},\tau_{max}+.05}(\mathcal{D})$, where $0 \leq \tau_{max} \leq 15.6$ by a step size of .05 μm.

Using the summary command in R, we found that the average interstice size is 3.97 μm and that half of all interstices are smaller than or equal to 3.30 μm, using the mean and median respectively. This information about the channel-walls of the interstices can be used to understand the useful amount of the fiber boundaries where the macromolecules diffuse in the separation process.

We then used the Maximum-likelihood Fitting of Univariate Distributions function, MASS::fitdistr, to fit a chi-squared distribution to our sample data. We plotted a chi-squared probability distribution aligned to the units of perimeter-distance in Figure 4.5 and layered this with the marginal increases in perimeter-distance given by $\ell_{\tau_{max},\tau_{max}+.05}(\mathcal{D})$, where $0 \leq \tau_{max} \leq 15.6$ with a step size of .05 μm in Figure 4.5. The chi-squared parameter provided by MASS::fitdistr was 4.09, using 3 as an initial guess for the parameter. This implies that perimeter-distance is increasing the most around $\tau_{max} = 2.09$ μm, the mode obtained from the parameter estimation.

In this chapter, we have shown how to practically implement a calculation of perimeter-distance. We have used an algorithm that reduces the complexity to achieve significant savings of computational resources. We have shown and explained an extensive summary of perimeter-distance calculations using data from [7].

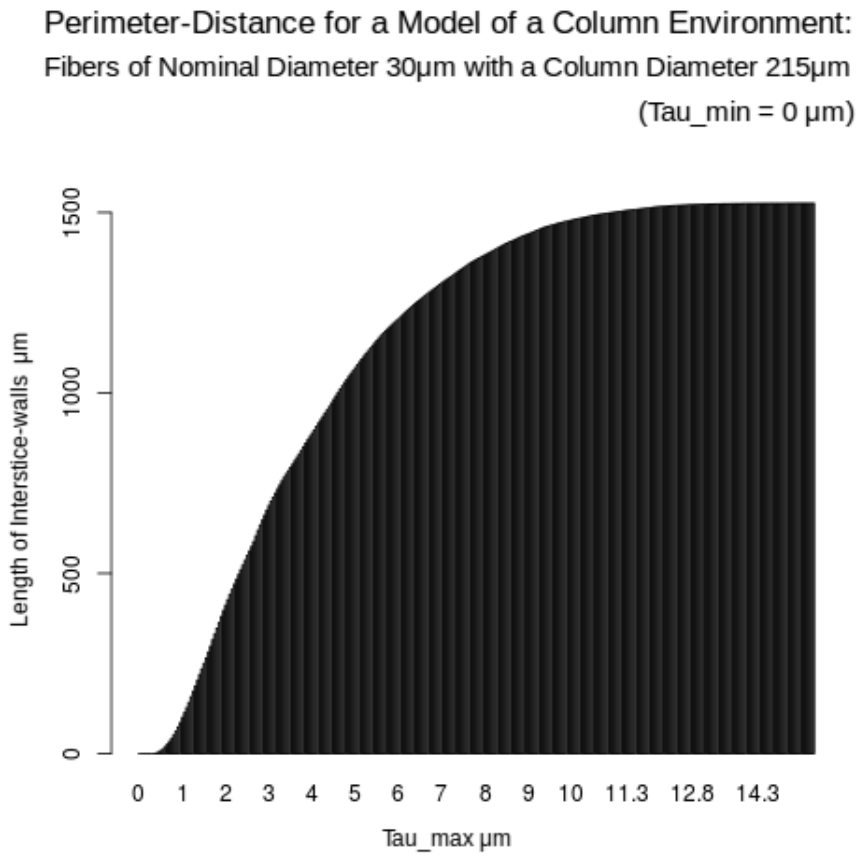


Figure 4.4: Calculations of Perimeter-Distance: In this figure, $\tau_{min} = 0$, and τ_{max} is the input for the graph. We see monotonic increasing behavior of $\ell_{0,\tau_{max}}(\mathcal{D})$ with respect to τ_{max}. The value of $\ell_{0,\tau_{max}}(\mathcal{D})$ remains constant for $15.6 \leq \tau_{max}$. The data used for these calculations is taken from [7], cf. Figure 1.2.

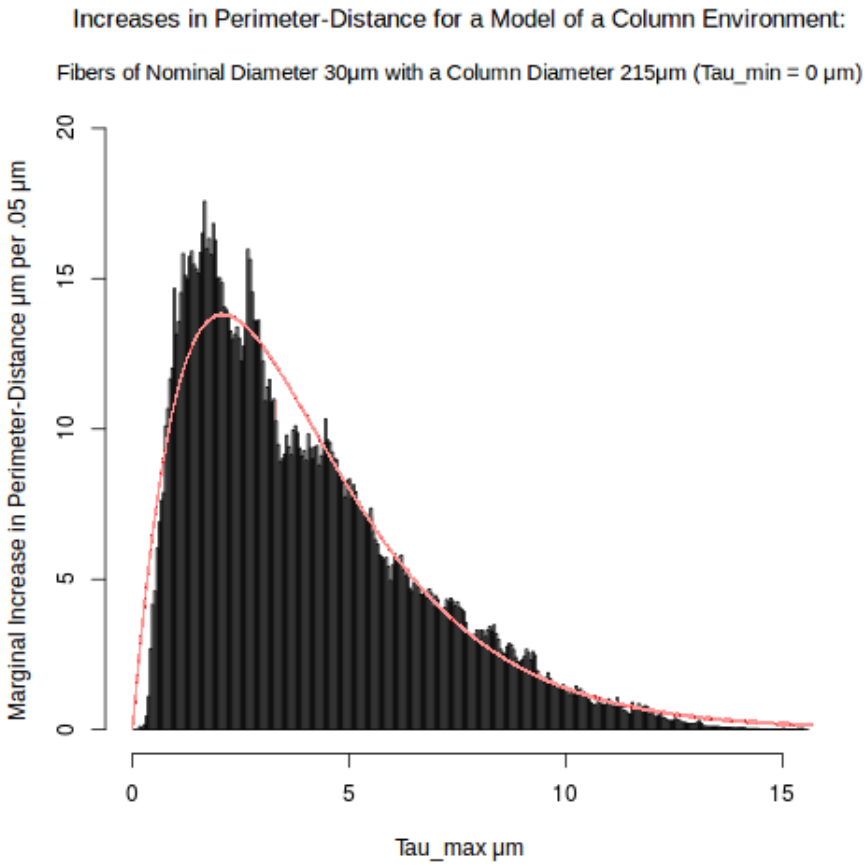
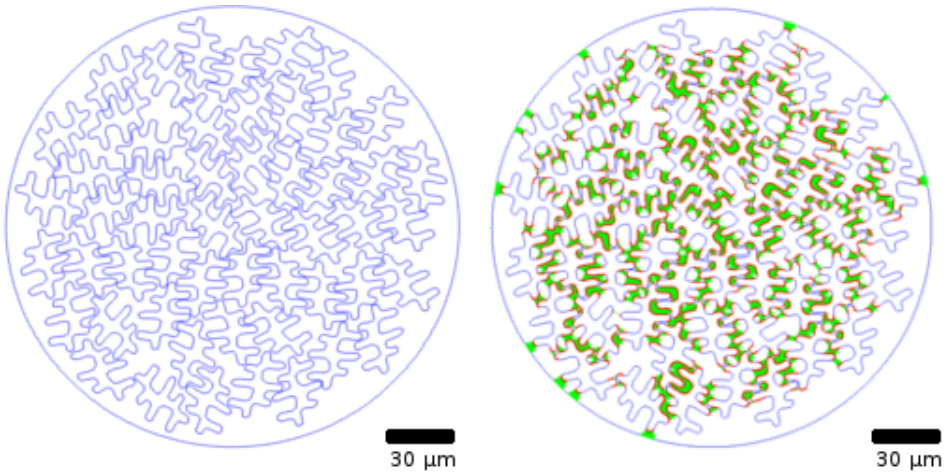


Figure 4.5: Chi-Squared Distribution and Perimeter-Distance: In this figure we see the marginal increases of perimeter-distance. We see that $\ell_{0,\tau_{max}}(\mathcal{D})$ increases very rapidly and then increases more slowly until $\tau_{max} = 15.6$. This figure also shows the maximum-likelihood fit for a chi-squared distribution using MASS::fitdistr from the statistical software package R. We can obtain an estimation of the greatest increase to perimeter-distance from the mode of the chi-squared distribution, $\tau_{max} = 2.09$.

Chapter 5

Channel-Area

In Chapter 4, for a domain $\mathcal{D}$, we showed how to identify the channel-walls via $\mathcal{S}_{\tau_{min},\tau_{max}}(\mathcal{D})$. In this chapter, we construct the channels, see Figure 5.1(b), using the line segments of $\mathcal{S}_{0,\tau_{min}}(\mathcal{D})$ and $\mathcal{S}_{0,\tau_{max}}(\mathcal{D})$.



(a) (b)

Figure 5.1: The Channels of the Domain: In (a), the model of the column environment is exhibited as the domain of fiber-polygons in the regular polygon approximating the inner column wall with a diameter of 215 μm where the polygon-fibers have a diameter of 30 μm. In (b), we have identified the channels of the column environment where $\tau_{min} = 1.5$ μm and $\tau_{max} = 4$ μm. In Figure 5.7, we show a larger image of (b).

5.1 Channel Distance

Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. In this section, we compute the area of the cross section through which macromolecules flow and may ultimately diffuse to some channel-wall. We start by identifying the interstices smaller than τ_{max}. We define these interstices in Definition 5.1.1 using the boundaries of $\mathcal{S}_{0,\tau_{max}}(\mathcal{D})$. From these interstices, we remove the interstices which are smaller than τ_{min}, using Definition 5.1.1. An example of the resulting channels of this process is shown in Figure 5.1(b). The area of the channels is defined to be channel-area, cf. Definition 5.1.8.

Definition 5.1.1. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $\tau_{max} > 0$. In order to form the interstices which are smaller than or equal to τ_{max}, we define for $S_i \in \mathcal{S}_{0,\tau_{max}}(\mathcal{D})$ the nearby points on the polygons in the domain, $D^-_{S_i}$, which are within τ_{max} of S_i by

$$R_{\tau_{max}}(S_i, \mathcal{D}) := \{x \in D^-_{S_i} : d(x, S_i) \leq \tau_{max}\}.$$

We can find $R_\tau(S_i, \mathcal{D})$ using a Minkowski sum by applying Equation 3.1,

$$R_{\tau_{max}}(S_i, \mathcal{D}) = D^-_{S_i} \cap (S_i + B_{\tau_{max}}) \tag{5.1}$$

where $B_{\tau_{max}}$ is a disk of radius τ_{max} centered at the origin. We define the set of maximal segments of $R_{\tau_{max}}(S_i, \mathcal{D})$ to be $\mathcal{R}_{\tau_{max}}(S_i, \mathcal{D})$.

Definition 5.1.2. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $\tau_{max} > 0$.

$$A_{0,\tau_{max}}(\mathcal{D}) = \bigcup_{S_i \in \mathcal{S}} \bigcup_{R_j \in \mathcal{R}_{\tau_{max}}(S_i, \mathcal{D})} CH(S_i, R_j).$$

Observe, the convex hull constructs triangles or quadrilaterals between line segments which are nearby to one another.

Example 5.1.3. Let $\mathcal{D}$ and $\mathcal{P}$ contain two triangles and $\tau > 0$. In Chapter 3, we discussed how to find $\mathcal{S}_{0,\tau}(\mathcal{D})$. For a given $S_i \in \mathcal{S}_{0,\tau}(\mathcal{D})$ we can use a Minkowski sum, according to

Equation 5.1, to find an $R_j \in R_\tau(S_i, \mathcal{D})$. The convex hull of S_i and R_j is contained in $A_{0,\tau}(\mathcal{D})$, by Definition 5.1.2. In Figure 5.2(a), we show how to use the Minkowski sum to find R_j. In Figure 5.2(b), we show the convex hull of S_i and R_j.

Example 5.1.4. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. The interstices smaller than τ_{max}, $A_{0,\tau_{max}}(\mathcal{D})$, consist of points, x, in the column environment where there is a line between polygons in the domain passing through x of length less than or equal to τ_{max}, see 5.3(a). So, the channels $A_{\tau_{min},\tau_{max}}(\mathcal{D})$, consist of points, x, in the column environment where there is a line between polygons in the domain passing through x of length less than or equal to τ_{max} but greater than τ_{min}, see 5.3(b).

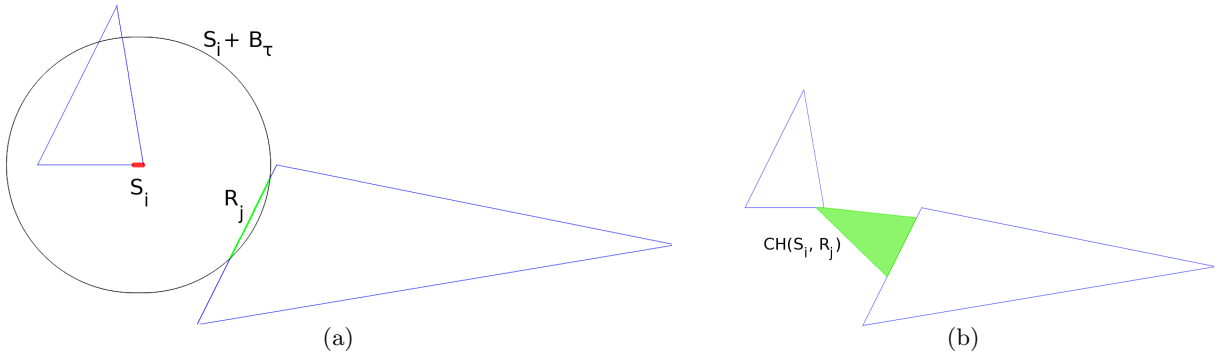


Figure 5.2: Identifying Channels: In these figures, we show a convex hull used to determine the channel between the triangles. In (a), we have constructed an interstice-wall on the left triangle, S_i; the boundary of $S_i + B_\tau$ that surrounds the interstice-wall on the left triangle, is used to identify R_j on the right triangle. In (b), we construct the convex hull of the interstice-wall, S_i on the left triangle with R_j on the right triangle.

Lemma 5.1.5. *Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. Then*

$$A_{0,\tau_{min}}(\mathcal{D}) \subseteq A_{0,\tau_{max}}(\mathcal{D}).$$

Proof. By Proposition 2.5.3 and Definition 1.2.4, it follows that $\mathcal{S}_{0,\tau_{min}}(\mathcal{D}) \subseteq \mathcal{S}_{0,\tau_{max}}(\mathcal{D})$. In particular, for each line segment in $S_i \in \mathcal{S}_{0,\tau_{min}}(\mathcal{D})$, S_i is a subset of a line segment in $\mathcal{S}_{0,\tau_{max}}(\mathcal{D})$, say S_j. By Definition 5.1.1, $\mathcal{R}_{\tau_{min}}(S_i, \mathcal{D}) \subseteq \mathcal{R}_{\tau_{max}}(S_j, \mathcal{D})$. Therefore, each

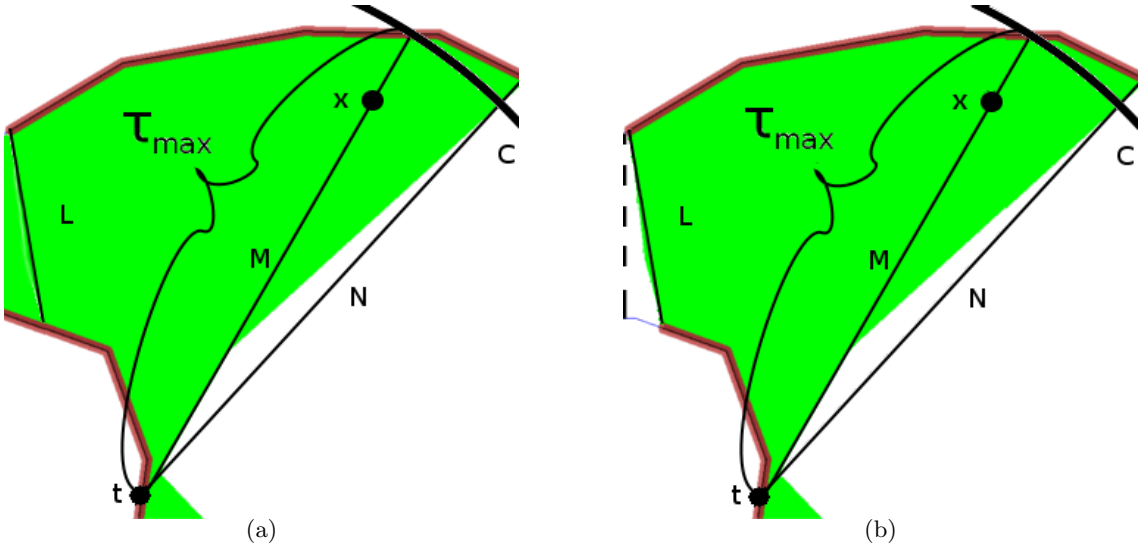


Figure 5.3: The Boundary of the Channels: In (a), we shade $A_{0,\tau_{max}}(\mathcal{D})$. In (b), we shade the channel, $A_{\tau_{min},\tau_{max}}(\mathcal{D})$. The boundaries of the channels are not naively connected to channel-walls; notice, the line segments L and N have endpoints which are endpoints of maximal line segments of the channel-walls in (b), however, L and N are not boundaries of the channels. τ_{max} is the length of the line segment M, and thus x is included in the interstice in (b). In the upper right corner of the figures, the curve C is part of the circle of radius τ_{max} centered at the point t. There is an unshaded region bounded by the line segment N and the shaded channel; i.e., this region is not part of the channel. Any point y in this region is not part of the channel because the length of any line segment passing though y with endpoints on the respective channel-walls is greater than τ_{max}. We see that the line segment N starts at the center of the circle, C, and extends beyond the circle, i.e., the length of N is greater than τ_{max}. Since N starts and ends on opposite channel-walls and is of length greater than τ_{max} there must be some point in N which is not part of the channel.

convex hull in $A_{0,\tau_{min}}(\mathcal{D})$ is a subset of a convex hull of $A_{0,\tau_{max}}(\mathcal{D})$, hence the containment.

□

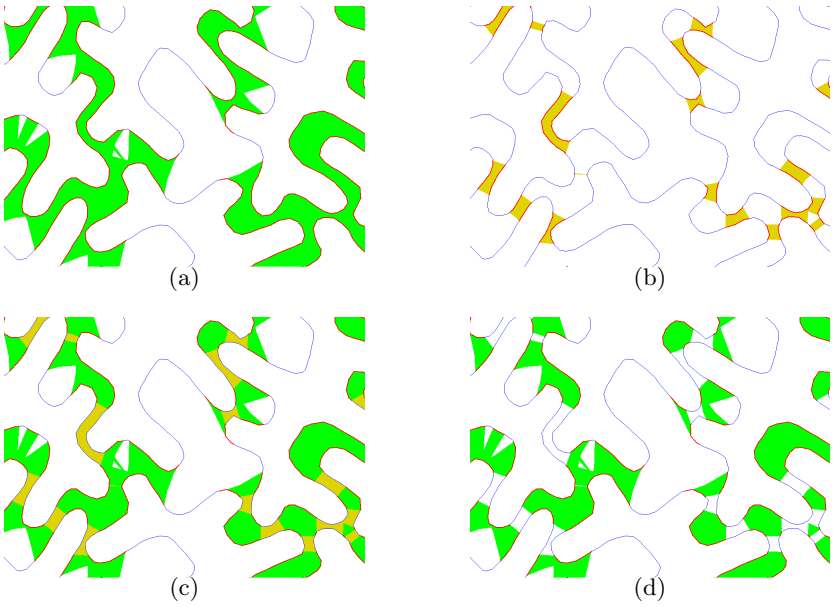


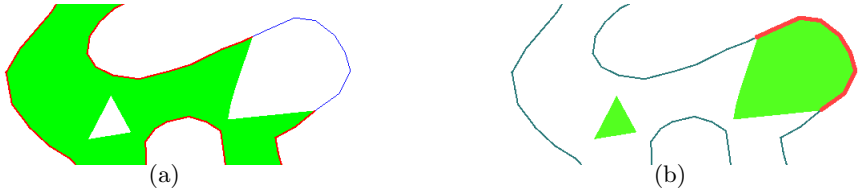
Figure 5.4: Building Channel Distance: In (a), the interstices of size less than τ_{max}, $A_{0,\tau_{max}}(\mathcal{D})$, are shaded. In (b), the interstices of size less than τ_{min}, $A_{0,\tau_{min}}(\mathcal{D})$, are shaded. In (c), we show $A_{0,\tau_{min}}(\mathcal{D}) \subseteq A_{0,\tau_{max}}(\mathcal{D})$. In (d), the channels, $A_{\tau_{min},\tau_{max}}(\mathcal{D}) = A_{0,\tau_{max}}(\mathcal{D}) \setminus A_{0,\tau_{min}}(\mathcal{D})$, are shown, from which channel-area can be computed.

Definition 5.1.6. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. We define

$$A_{\tau_{min},\tau_{max}}(\mathcal{D}) := A_{0,\tau_{max}}(\mathcal{D}) \setminus A_{0,\tau_{min}}(\mathcal{D})$$

where the right hand side is a set difference. We define the set of maximal polygons of $A_{\tau_{min},\tau_{max}}(\mathcal{D})$ to be $\mathcal{A}_{\tau_{min},\tau_{max}}(\mathcal{D})$.

Example 5.1.7. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. We see that $A_{0,\tau_{min}}(\mathcal{D}) \subseteq A_{0,\tau_{max}}(\mathcal{D})$ in Figure 5.4(a) with those of smaller size in Figure 5.4(b). Channels are formed by computing $A_{0,\tau_{min}}(\mathcal{D})$ and removing $A_{0,\tau_{max}}(\mathcal{D})$ as in Figure 5.4. The interstices with sizes smaller than τ_{min}, $A_{0,\tau_{min}}(\mathcal{D})$, may contain holes,



(a) (b)

Figure 5.5: Channel Holes: As shown in Figure 5.3, points in the interstices are not included in $A_{0,\tau_{min}}(\mathcal{D})$ when no line segment can passes though the point and connect to two polygons in the domain with length less than τ_{min}. Hence, in (a) the unshaded regions consists of a collection of points which are not too close to the fiber-polygons. In (b), we show $A_{\tau_{min},\tau_{max}}(\mathcal{D})$ where $A_{0,\tau_{max}}(\mathcal{D})$ (not shown) contains all the interstices of the domain. Therefore, all of the unshaded regions from (a) are included in $A_{\tau_{min},\tau_{max}}(\mathcal{D})$.

see Figure 5.5(a). When $A_{0,\tau_{min}}(\mathcal{D})$ is removed from $A_{0,\tau_{max}}(\mathcal{D})$ the holes form floating regions included in the channels, $A_{\tau_{min},\tau_{max}}(\mathcal{D})$, as seen in Figure 5.5(b).

Definition 5.1.8. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. We define *channel-area* of $\mathcal{D}$ to be

$$a_{\tau_{min},\tau_{max}}(\mathcal{D}) := \sum_{A_i \in \mathcal{A}_{\tau_{min},\tau_{max}}(\mathcal{D})} \text{area}(A_i).$$

Remark 5.1.9. By Lemma 5.1.5,

$$a_{\tau_{min},\tau_{max}}(\mathcal{D}) = \left(\sum_{A_i \in \mathcal{A}_{0,\tau_{max}}(\mathcal{D})} \text{area}(A_i) \right) - \left(\sum_{A_i \in \mathcal{A}_{0,\tau_{max}}(\mathcal{D})} \text{area}(A_i) \right).$$

Example 5.1.10. Let $\mathcal{P}$ and C define a domain $\mathcal{D}$ as in Definition 1.2.2. Let $0 < \tau_{min} < \tau_{max}$. The interstice-walls, $\mathcal{S}_{0,\tau_{max}}(\mathcal{D})$ and $\mathcal{S}_{0,\tau_{min}}(\mathcal{D})$, found when computing perimeter-distance are used to find the boundaries of the channels. This establishes a connection between perimeter-distance and channel-area. The interstice-walls are shown in Figure 5.6.

In this chapter, we showed how to use the channel-walls of the model to find the channels which represents the cross sectional flow regions of the column environment which contain macromolecules with may diffuse to the fibers. We showed how the computation of channel-area can be performed in a similar manner as perimeter distance, cf. Remark

5.1.9 and Proposition 3.3.1. We found a simple geometric interpretation of the channels, Example 5.1.4.

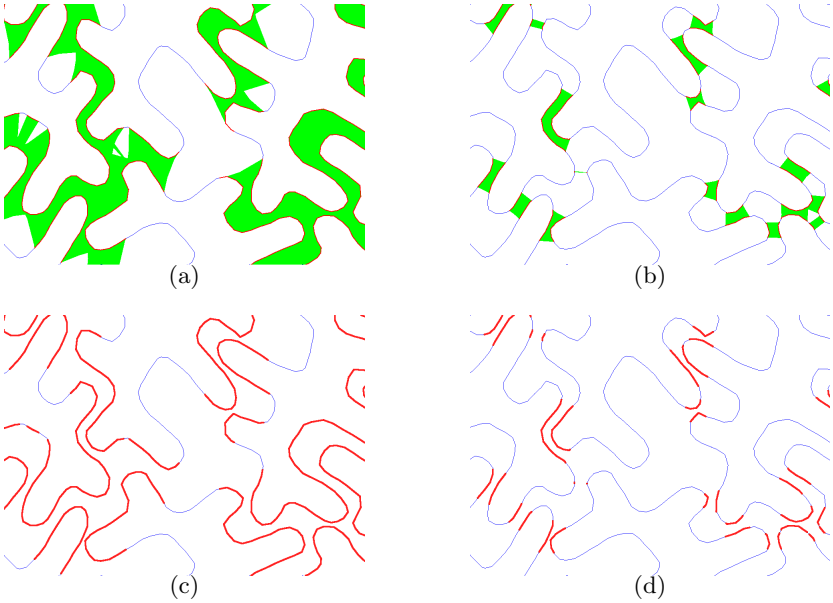


Figure 5.6: Channel Distance: We show a detailed image of the visualizations of perimeter and channel-area calculations performed on many polygons where $\tau_{min} = 0$. In (a), we shade the interstices of size smaller than τ_{max}, $\mathcal{A}_{0,\tau_{max}}(\mathcal{D})$. In (c), we show the collection of interstice-walls $\mathcal{S}_{0,\tau_{max}}(\mathcal{D})$. In (b), we shade the interstices of size smaller than τ_{min}, $\mathcal{A}_{0,\tau_{min}}(\mathcal{D})$. In (d), we show the collection of interstice-walls $\mathcal{S}_{0,\tau_{min}}(\mathcal{D})$.

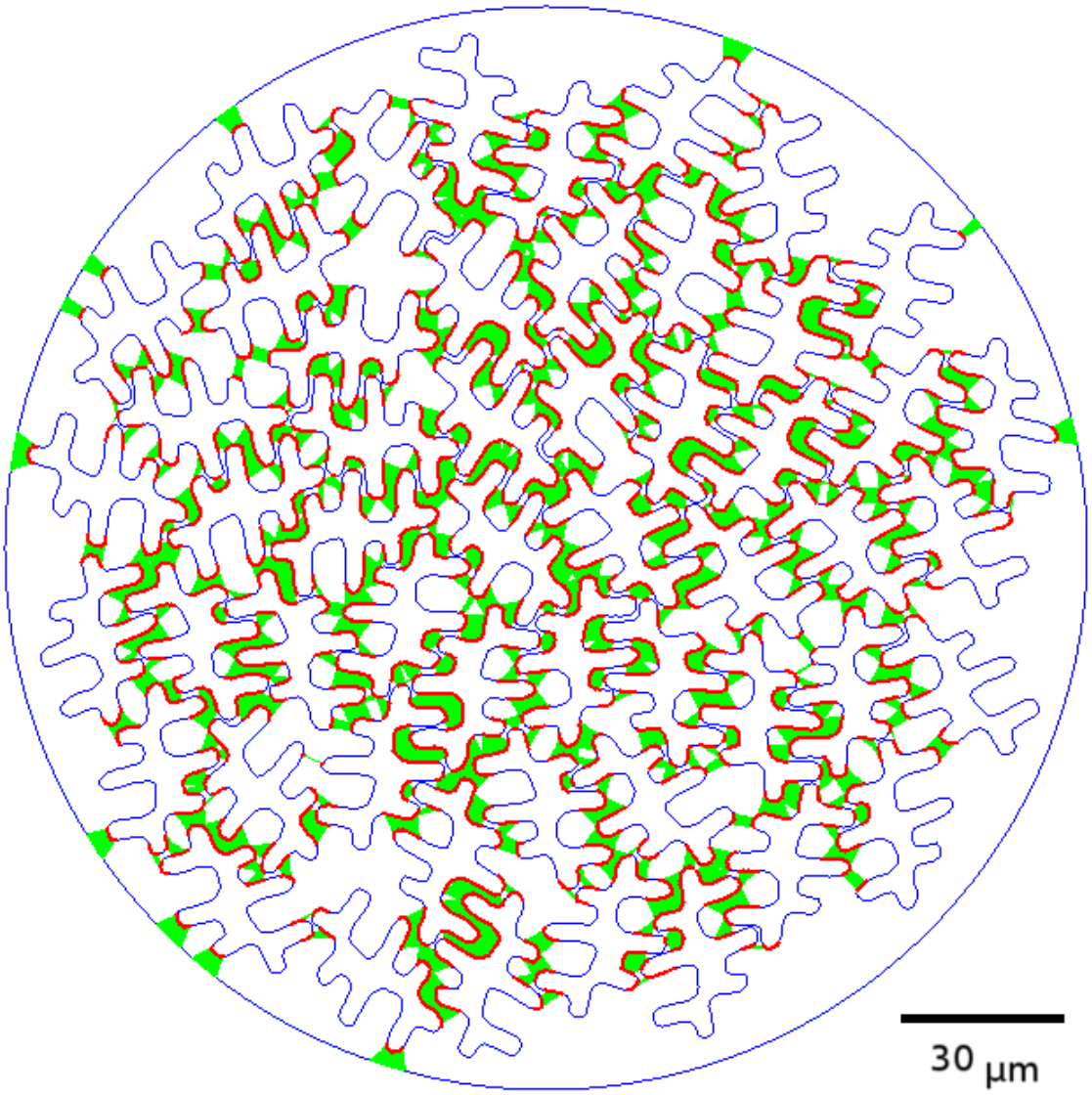


Figure 5.7: Channel Distance on the Domain: We show a larger image of Figure 5.1 where the model is of a column environment with an inner diameter of 215 μm where the polygon-fibers have a diameter of 30 μm. The shaded channels represent the cross sectional area through which macromolecules flow and may diffuse to a fiber boundary.

Chapter 6

Conclusion and Future Work

In this chapter, we explain the ways in which our calculations can be applied to purification processes using high performance liquid chromatography. We also explain some areas for future work with perimeter-distance and channel-area.

6.1 Conclusion

The efficiency of purification processes using high performance liquid chromatography are determined by the positions of the fibers in the column environment. We identified the channel-walls which represents the boundaries of fibers in the cross section of the column environment where macromolecules may diffuse. In Section 2.4, we define the total length of the boundaries of the fiber-polygons to be perimeter-distance. In Chapter 4, we showed how to compute perimeter-distance with an implementation in C++ with CGAL.

In Chapter 5, we identified the interstices in our model, the channels, which represent the regions of the cross section of the column through which macromolecules contained in a fluid travels before it may diffuse to a boundary of a fiber. In Chapter 5, we show the geometrical connection between the channel-walls and the channels of our model. We also defined, channel-area which gives an approximation of the flow regions of a cross section of the column environment through which macromolecules pass in a fluid before diffusing to

the fibers.

For a given instance of a column environment $\mathcal{D}$ and a molecule with empirically determined τ_{min} and τ_{max}, we have created software that approximates the useful boundaries of the fibers of the column environment. Perimeter-distance gives an approximation of the length of the useful boundaries of the fibers.

6.2 Future Work

The example of the column environment, from [7] and shown in Figure 1.2, used in the previous chapters was of a smaller diameter than the sizes of column diameters currently in use. We would like to obtain examples of column environments which contain as many fibers as commonly used column environments and are of full size. These computations may provide more relevant data for the purification process.

In the cross-section of a column environment, the total area of all the interstices divided by the area of the column is called the interstitial fraction. Current research shows that for the fibers in the column environment the optimal interstitial fraction for purification processes in columns of various diameter sizes is approximately 63%, see [13]. We can calculate perimeter-distance on domains $\mathcal{D}$ which approximate the column environments for which experimental data of the efficiency of the separation in column environment has already been found. We can then use a statistical software to test if there is a significant correlation between perimeter-distance calculations and the existing column environment efficiency data. This would validate the ability of perimeter-distance to predict the efficiency of the separation process in the column environment.

Computation of perimeter-distance can be applied to column environments with different fiber specifications, including size and shape. There appears to be limited experimental results where other shaped fibers are used in monolithic columns. Perimeter-distance calculation can perform calculations on examples with available alternative fiber shapes. These calculations may help determine which alternative fiber shapes merit further

research and experimentation.

The example of the column environment in Figure 1.2 was created with assistance of human decisions. It would be advantageous to develop a program that simulates filling the column environment with fiber-polygons. Perimeter-distance could be used to determine if the results of the simulation were consistent or if significant variations among repeated constructions of column environments occur in the formation of channel-walls.

Perimeter-distance may also be used to create new fiber shapes. By using a genetic algorithm, it is possible to optimize the fiber shape for the separation of molecules in the column environment during the purification process. This could be realized by allowing the algorithm to make incremental changes to polygons in each generation. Each variation of the polygon would fill a respective example of a column environment. Perimeter-distance can then identify the fiber, or fibers, among each generation which would be best for the column environment. By preforming this process repeatedly optimized fiber shapes for the column environment could be found.

Extensive use of perimeter-distance in any of these applications would make it more advantageous to develop parallelized code for perimeter-distance calculations. The implementation of perimeter-distance used the exact number type CORE in CGAL. Currently, programs using CORE are not able to be parallelized. However, a parallelization compatible version is in development [6].

To reduce the total runtime of an application using perimeter-distance, one could also pursue an alternative to the heuristic for reducing the number of intersections computed. A bucket sort or quad-tree could be used, as described in [15]. Recall that the complexity, in tests, for the heuristic for reducing the number of intersections between line segments and the Minkowski sum of a disk and line segment was $O(n^2)$ where n is the number of polygons. A bucket sort may be implemented on polygons in the domain which results in a lower complexity, see [17]. A quadtree, a commonly used spatial indexing structure, would be a beneficial option since the paths in the tree determine the distance between the objects. Furthermore, with a few modifications, either of these approaches may be applied

to the edges of the polygons which could significantly reduce unnecessary calculations.

The major limitation in our development of the program to compute channel-area is the use of memory. Memory usage increases prohibitively when computing channel-area via the aggregate union of polygons. It is possible that the excessive memory consumption is due to coefficient swell associated with computing resultants, due to the algebraic computation in CORE. One possible way to reduce the amount of memory used is to use intersections instead of unions. Channel-area can then be computed by using the inclusion-exclusion principal.

Bibliography

[1] Qt. www.qt.io.

[2] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.

[3] CORE, Core Library. http://cs.nyu.edu/exact/core_pages/.

[4] PALMETTO CLUSTER, The Clemson Univeristy Palmetto Cluster. http://citi.clemson.edu/palmetto/.

[5] E. J. Borowski. *Collins dictionary of mathematics*. Collins, London, 2007.

[6] M. A. Burr. personal communication.

[7] C. L. Cox. personal communication.

[8] L. Zhilin D. Min and C. Xiaoyong. Extended hausdorff distance for spatial objects in gis. *International Journal of Geographical Information Science*, 21(4):459–475, 2007.

[9] M. de Berg, O. Cheong, M. van Kreveld, and M Overmars. *Computational geometry algorithms and applications*. Springer, Berlin, 2008.

[10] H. Edelsbrunner. Computing the extreme distances between two convex polygons. *J. Algorithms*, 6(2):213–224, June 1985.

[11] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850–863, 1993.

[12] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[13] K. M. Randunu, S. Dimartino, and R. K. Marcus. Dynamic evaluation of polypropylene capillary-channeled fibers as a stationary phase in high-performance liquid chromatography. *J. Sep. Science*, (35):32703280, 2012.

[14] K. M. Randunu and R. K. Marcus. Initial evaluation of protein throughput and yield characteristics on nylon 6 capillary-channeled polymer (c-cp) fiber stationary phases by frontal analysis. *Biotechnol Prog*, (29(5)):1222–9, 2013.

[15] R. L. Rivest C. Stein T. H. Cormen, C. E. Leiserson. *Introduction to algorithms*. MIT Press, Cambridge, Mass, 2001.

[16] Z. Wang and R. K. Marcus. Determination of pore size distributions in capillary-channeledpolymer fiber stationary phases by inverse size-exclusionchromatography and implications for fast protein separations. *Journal of Chromatography A*, (1351):8289, 2014.

[17] M. Weiss. *Data structures and algorithm analysis in C*. Benjamin/Cummings Pub. Co, Redwood City, Calif, 1994.