CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

PROTEIN FOLDING: PLANAR CONFIGURATION SPACES OF DISC ARRANGEMENTS AND HINGED POLYGONS

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Applied Mathematics

by

Clinton Bowen

August 2014

The thesis of Clinton Bowen is approved:

Dr. Silvia Fernandez	Date
Dr. John Dye	Date
Dr. Csaba Tóth, Chair	Date

California State University, Northridge

Table of Contents

ABSTRACT

PROTEIN FOLDING: PLANAR CONFIGURATION SPACES OF DISC ARRANGEMENTS AND

HINGED POLYGONS

By

Clinton Bowen

Master of Science in Applied Mathematics

Chapter 1

Realizability Problems for Weighted Trees

In this chapter our goal is to prove Theorem **??**: It is NP-Hard to decide whether a given tree with positive vertex weights is the contact graph of a disk arrangements with specified radii. This chapter's approach to proving Theorem **??** introduces an ordered weighted tree T and perturbed ordered weight tree T_ε, the Hausdorff distance, and then prove the following lemma:

1.1 Hausdorff Distance

Let A and B be sets in the plane. The *directed Hausdorff distance* is:

$$d(A,B) = \sup_{a\in A} \inf_{b\in B} ||a-b|| \tag{1.1}$$

$d(A,B)$ finds the furthest point $a \in A$ from any point in B. *Hausdorff distance* is

$$D(A,B) = \max\{d(A,B), d(B,A)\} \tag{1.2}$$

In Figure 1.1, we have two sets X and Y and illustrate $d(X,Y)$ and $d(Y,X)$. From this, it is possible to calculate the Hausdorff distance between X and Y.

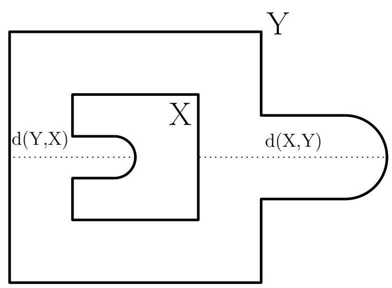

Figure 1.1: An illustrative example of $d(X,Y)$ and $d(Y,X)$ where X is the inner curve, and Y is the outer curve.

ε-approximation The weighted graph, G, is an ε-*approximation* of a polygon P if the Hausdorff distance between every realization of G as a contact graph of disks and a congruent copy of P is at most epsilon. A weighted graph G is said to be a $O(f(x))$-*approximation* of a polygon P if there is a positive constant M such that for all sufficiently large values of x the Hausdorff distance between every realization such realization of G as a contact graph of disks and a congruent copy of P is at $M \cdot |f(x)|$. A weighted graph G is said to be a *stable* if it has the property that for every two such realizations of G, the distance between the centers of the corresponding disks is at most ε after a suitable rigid transformation.

Suppose we have a unit disk U and we have a grid overlayed on the disk with side length δ. Let $S_1(\delta)$ be the the union of squares formed by the grid found completely in the interior of the disk U. Let $S_2(\delta)$ be the union of squares formed by the grid with some point of the boundary of the square contained in the interior of disk U. The Hausdorff distance of U and $S_1(\delta)$ is at most $H(S_1(\delta),U) = \sqrt{2}\delta$. Similarly, the Hausdorff distance of U and $S_2(\delta)$ is at most $H(S_2(\delta),U) = \sqrt{2}\delta$. Thus for any $\varepsilon > 0$ choose a δ such that $\sqrt{2}\delta \leq \varepsilon$. Similiaryly, the Hausdorff distance of U and $S_2(\delta)$ is at most $H(S_2(\delta),U) = \sqrt{2}\delta$.

Problem 1 (Appoximating Polygonal Shapes with Contact Graphs). For every $\varepsilon > 0$ and polygon P, there exists a contact graph $G = (V,E)$ such that the Hausdorff distance $d(P,G) < \varepsilon$

Lemma 1. *for ever $\varepsilon > 0$, there exists an ordered weighted tree T_ε such that every realization of T_ε as an* *ordered disk contact graph where the radii of the disks equal the vertex weights.*

Using Lemma 1, we prove Theorem **??** by extending the modified auxiliary construction in Chapter **??**.

We first cover the preliminary concepts of Hausdorff distance and the ordered weighted tree families of T and T_ε. We then continue with the proof of Lemma 1 and Theorem **??**.

1.2 Weight Trees T_k

In this section we describe a particular family of unit weight trees and corresponding contact graphs disk arrangements called *snowflakes*. Note that we regard snowflakes with unit weight as a weight of $\frac{1}{2}$. For $i \in \mathbb{N}$, the construction of the snowflake tree, T_i, is as follows:

- Let v_0 be a dvertex that has six paths attached to it: $p_1, p_2, \ldots, p_6$. Each path has i vertices.
- For every other path p_1, p_3, and p_5:
  - Each vertex on that path has two paths attached, one path on each side of p_k.
  - The number of vertices that lie on a path attached to the j^{th} vertex of p_k is $i-j$.

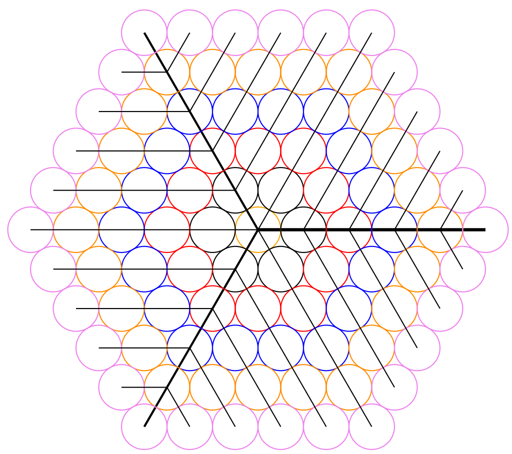

Figure 1.2: The same contact graph as in figure **??** overlayed with the a perfectly weighted snowflake tree.

A *perfectly weighted snowflake tree* is a snowflake tree with all vertices having weight $\frac{1}{2}$. A *perturbed snowflake tree* is a snowflake tree with all vertices having weight of 1 with the exception of v_0; in a perturbed snowflake tree, v_0 will have a weight of $\frac{1}{2}+\gamma$. For our analysis, all realizations of any snowflake, perfect or perturbed, shall have v_0 fixed at origin.

Perfectly Weighted Snowflake Tree. Consider the graph of the triangular lattice with unit distant edges:

$$\begin{aligned} V &= \left\{a \cdot (1,0) + b \cdot \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) : a, b \in \mathbb{Z}\right\} \\ E &= \{\{u,v\} : ||u-v|| = 1 \text{ and } u, v \in V\} \end{aligned}$$

The following graph, $G=(V,E)$ is said to be the *unit distance graph* of the triangular lattice. We can show that no two distinct edges of this graph are non-crossing. First suppose that there were two distinct edges that crossed, $\{u_1, v_1\}$ and $\{u_2, v_2\}$. With respect to u_1, there are 6 possible edges corresponding to it, with each edge $\frac{\pi}{3}$ radians away from the next. Neither edge crosses another; and so we have a contradiction that there are no edge crossings with $\{u_1, v_1\}$.

The perfectly weighted snowflake tree that is a subgraph over the *unit distance graph*, $G = (V, E)$, of the triangular lattice. To show this, for any S_i, fix $v_0 = 0\cdot\cdot(1,0) + 0\cdot\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = (0,0) \in V$ at origin. Next consider the six paths attached from origin. Fix each consecutive path $\frac{\pi}{3}$ radians away from the next such that the following points like on the corresponding paths: $(1,0) \in p_1, \left(\frac{1}{2}, \frac{\sqrt{2}}{3}\right) \in p_2, \left(-\frac{1}{2}\mathsf{p}_4, \frac{\sqrt{3}}{2}\right) \in p_3, (-1,0) \in p4, \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) \in p_5, \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) \in p_6$. For S_i, there are i vertices on each path.

We define the six paths from origin as follows:

$$\begin{aligned}
p_1 &= \left\{ a\cdot(1,0) = \vec{v}\,\middle|\, a \in \mathbb{R}^+ \right\} \\
p_2 &= \left\{ a\cdot\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \vec{v}\,\middle|\, a \in \mathbb{R}^+ \right\} \\
p_3 &= \left\{ -a\cdot(1,0) + a\cdot\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = a\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \vec{v}\,\middle|\, a \in \mathbb{R}^+ \right\} \\
p_4 &= \left\{ a\cdot(-1,0) = \vec{v}\,\middle|\, a \in \mathbb{R}^+ \right\} \\
p_5 &= \left\{ a\cdot\left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) = \vec{v}\,\middle|\, a \in \mathbb{R}^+ \right\} \\
p_6 &= \left\{ a\cdot(1,0) - a\cdot\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = a\cdot\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) \,\middle|\, a \in \mathbb{R}^+ \right\}
\end{aligned}$$

For S_i there exists i vertices on each path. We shall denote the i^{th} vertex on the j^{th} path as $v_{j,i}$. For each path defined above, the paths are defined as a set of vectors, $\vec{v} = a\cdot\vec{p}$ for some $a \in \mathbb{R}^+$ and $\vec{p} \in \mathbb{R}^2$. By setting $a = 1, 2, \ldots, i$, we obtain points that are contained in V. For $j = 1$, 3, 5 and $l = 3b \leq i$ where $b \in \mathbb{N}$, there exists two paths attached to each vertex $v_{j,l}$. We borrow the term *petiole* from botany to describe the two paths attached to $v_{j,l}$. In botany, the stalk that attaches to a stem of a plant is called a petiole; petioles usually have leaves attached to their ends. For S_i, each petiole attached to the k^{th} vertex of p_j, there are $i-k$ vertices. For each vertex v on a petiole, which is not in the paths p_1, p_3, or p_5, there are two *leafs* on either side of the vertex; each leaf is a vertex that has an edge with v. The one exception to the two leafs rule is on the first vertex of the petiole off of p_1, p_3, or p_5. In this exception, attach one leaf to the side of the vertex that is closest to center vertex v_0.

The triangular lattice is symmetric under rotation about v_0 by $\frac{\pi}{3}$ radians. For each vertex $v_{1,l}$ and $l = 3b \leq i$ where $b \in \mathbb{N}$, we place two petioles from it; the first petiole $\frac{\pi}{3}$ above p_1 at $v_{1,l}$ and $\frac{-\pi}{3}$ below p_1 at $v_{1,l}$ and call these petioles $p^+_{1,l}$ and $p^-_{1,l}$ respectively. With respect to $v_{1,l}$, one unit along $p^+_{1,l}$ is a point on the triangular lattice and similarly so on $p^-_{1,l}$. Continuing the walk along these paths, unit distance-by-unit distance, we obtain the next point corresponding point on the the triangular lattice up to $i-k$ distance away from $v_{1,l}$. Without loss of generality, for each each vertex v of the petiole which are not in p_1 have two associated leaf nodes v^+ and v^-; v^+ is placed $\frac{\pi}{3}$ and one unit above v and v^- is placed $\frac{-\pi}{3}$ and one unit below v. Thus all leaf nodes are in the triangular lattics. This shows that each of the $i-k$ vertices on $p^-_{1,l}$, $p^+_{1,l}$, and leafs are in V. By rotating all of the paths along p_1 by $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$, we obtain the the paths along p_3 and p_5 respectively, completing the construction.

In Figure 1.3, we have a set of unit radii disks arranged in a manner that outlines the perfectly weighted snowflake description above.

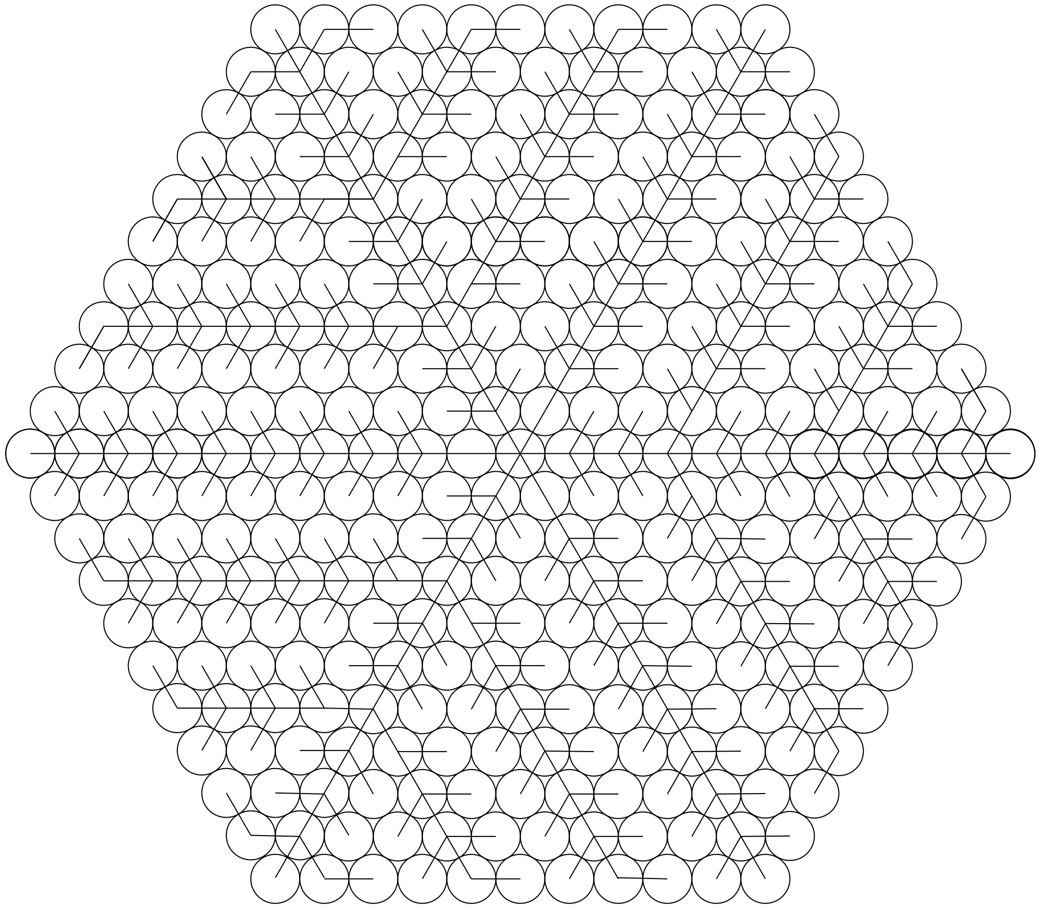

Figure 1.3: A contact graph that resembles the shape of concentric hexagons.

1.2.1 Perturbed Weighted Trees T_ε

A perturbed weighted tree T_ε is a weighted unit tree with unit weight on every vertex with the exception of the root vertex having weight $r+\varepsilon$ where $\varepsilon > 0$ can be realized as a disk touching graph (a disk arrangement) and r is the unit length.

The perturbed snowflake follows the construction of the perfect snowflake with the exception of v_0 having weight $r+\varepsilon$ where $\varepsilon > 0$. A perturbed snowflake realization has some distinct qualities from perfect snowflake realizations. The angular relationships between adjacent vertices may vary; the distance between adjacent and neighboring vertices may vary as well.

In general, the perturbation ε can modify the realization of a perfect snowflake S_i in the following ways:

Modification of S_1.

Given a instance of a perturbed snowflake with v_0 having weight $\frac{1}{2}+\varepsilon$ where $\varepsilon > 0$, vertices neighboring v_0 each have a range of placement on the plane when realizaed as a disk arrangement. Figure **??** shows a realization of S_1 and illustrates one such example of possible gaps, ε, that could be created between adjacent disks of S_1 in a perfect snowflake.

Note that (1) the adjacent disks in a perfect snowflake may or may not be adjacent in a given perturbed snowflake of S_1 and (2) $S_1 \subseteq S_i$ for any $i \in \mathbb{N}$. Given a snowflake in arbirary position with n unit segments per arm, the arms of the snowflake has a maximal length of n, end to end, if in canonical position; otherwise, the arm will have an end to end length less than n (See Figure 1.4). The arm of a snowflake in arbitrary position corresponds to a compression and shift of vertices.

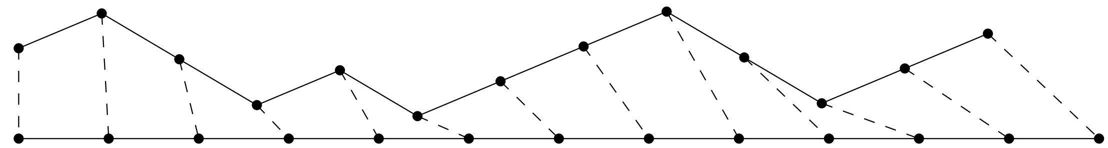

Figure 1.4: The polyline at the bottom represents a snowflake arm in canonical position. The polyline above represents a snowflake arm in non-canonical position.

We will show for any $\varepsilon > 0$ and arbitrary position of vertices, the placement of vertices is close to canonical position. In order to show this, we show the components of a perturbed snowflake in arbitrary position are close to canonical position. The argument comprises of three parts: (1) Showing that the pertubation of S_1 is small, (2) show that the displacement along the arms for all S_i for $i \geq 1$ is small, and (3) show that the displacement along the petioles is small.

Displacement on S_1 is small. Suppose we are given an ordered weighted tree T_ε such that the corresponding disk arrangement is a perturbed S_1. In a perfect snowflake of S_1 the six disks around the central disk kiss each other. The angle formed from the center of the central disk to the centers of any two adjacent disks is $\frac{\pi}{3}$. The side lengths of the equalateral triangle formed by the centers of three adjacent disks, one of which is the central disk, is $2r$. For a perturbed S_1 the the central disk is weighted $r+\varepsilon$. This can yield a change of angular displacement from $\frac{\pi}{3}$ to $\frac{\pi}{3}+2\chi$. To find the bounds of how large or small χ can be, we show the trigonometric relation of the half angle of the triangle corresponding to three adjacent disks (See Figure 1.5):

$$\begin{aligned}
\sin\left(\frac{\pi}{6}-\chi\right) &= \frac{1}{2r+\varepsilon} \\
\sin\frac{\pi}{6}\cos\chi &= \frac{1}{2r+\varepsilon}+\cos\frac{\pi}{6}\sin\chi \\
&\iff \\
\frac{1}{2} &\geq \frac{1}{2}\cos\chi \\
&= \frac{1}{2r+\varepsilon}+\frac{\sqrt{3}}{2}\sin\chi \\
&\geq \frac{1}{2r+\varepsilon}+\frac{\sqrt{3}}{2}\left(\chi-\frac{\chi^3}{6}\right) \\
&\iff \\
\frac{1}{2}-\frac{1}{2r+\varepsilon} &\geq \frac{\sqrt{3}}{2}\left(\chi-\frac{\chi^3}{6}\right) \qquad \text{if } \chi < 1 \\
\frac{2r+\varepsilon-2}{2(2+\varepsilon)} &\geq \frac{5\sqrt{3}}{12}\chi \\
& \quad \text{for } r = 1 \\
\frac{3\varepsilon}{5\sqrt{3}} = \frac{12}{5\sqrt{3}}\frac{\varepsilon}{4} &\geq \chi
\end{aligned}$$

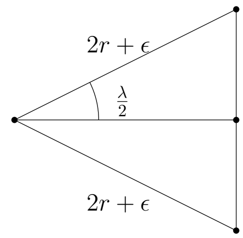

Figure 1.5: This figure depicts a triangle corresponding to the center of the central disk and two adjacent disks.

For any $\varepsilon > 0$, the bounds for angular displacement formed at the center of the central disk and two adjacent disks is:

$$\frac{\pi}{3} - \frac{6\varepsilon}{5\sqrt{3}} \leq \frac{\pi}{3} - 2\chi = \lambda_{\min} \leq \lambda \leq \lambda_{\max} = \frac{\pi}{3} + 2\chi \leq \frac{\pi}{3} + \frac{6\varepsilon}{5\sqrt{3}}$$

Displacement on the arms is small. To show that the angluar displacement along the arm is small, we extend the angular argument on the perturbed S_1 and by induction, show that it is small for all i.

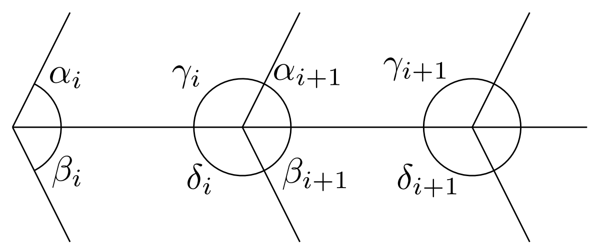

Figure 1.6: An arm depicted at the i^{th} and $(i+1)^{\text{st}}$ vertex.

Denote the angles on the concave side of the i^{th} vertex as α_i and β_i and the convex side of the $(i+1)^{\text{st}}$ vertex as γ_i and δ_i respectively (see Figure 1.6 for reference).

For any vertex, the sum of angles about the vertex is 2π, e.g.:

$$\gamma_i + \delta_i + \alpha_{i+1} + \beta_{i+1} = 2\pi$$

Suppose we numbered the the disks about the central disk 1 through 6. Without loss of generality, the angles α_0 and β_0 correspond to the angles formed between the central angle, disks i and $i+1$ and disks $i+1$ and $i+2$ respectively, for $i = 1,2,3$. The bounds for α_0 and β_0 are the same as λ in the earlier argument, i.e.:

$$\begin{array}{rcl} \frac{\pi}{3} - \frac{6\varepsilon}{5\sqrt{3}} & \leq \quad \alpha_0 \quad \leq & \frac{\pi}{3} + \frac{6\varepsilon}{5\sqrt{3}} \\ \frac{\pi}{3} - \frac{6\varepsilon}{5\sqrt{3}} & \leq \quad \beta_0 \quad \leq & \frac{\pi}{3} + \frac{6\varepsilon}{5\sqrt{3}} \end{array}$$

We know that $\alpha_0 + \beta_0 \leq \frac{2\pi}{3} + \frac{12\varepsilon}{5\sqrt{3}}$. We also know that in canonical position:

$$\begin{array}{rcl} \pi & = & \alpha_0 + \gamma_0 \\ \pi & = & \beta_0 + \delta_0 \end{array}$$

Together, we have the following result:

$$\begin{array}{rcl} rcl2\pi & = & \alpha_0+\gamma_0+\beta_0+\delta_0 \\ 2\pi & = & \alpha_0+\gamma_0+(2\pi-\alpha_1-\beta_1) \\ \alpha_1+\beta_1 & = & \alpha_0+\gamma_0 \\ & \leq & \frac{2\pi}{3}+\frac{12\varepsilon}{5\sqrt{3}} \end{array}$$

Displacement on the petioles is small.

Bibliography