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

We begin this chapter with the preliminary concepts to solving Problems **??** and **??**, and then prove related Theorem **??**. Recall the Unordered and Ordered Realizability Problem for a contact graph and corresponding theorem which state:

Problem ?? Given a planar graph with positive weighted vertices, is it a contact graph of some disk arrangement where the radii equal the vertex weights?

Problem ?? Given a planar graph with positive weighted vertices and a combinatorial embedding, is it a contact graph of some disk arrangement where the radii equal the vertex weights and the counter-clockwise order of neighbors of each disk is specified by the combinatorial embedding?

Theorem ?? It is NP-Hard to decide whether a given tree (resp., plane tree) with positive vertex weights is the contact graph (resp., contact graph) of a disk arrangements with specified radii.

The preliminary concepts are the Hausdorff distance, the Unit Disk Touching Graph Recongnition Problem, and the Perturbed Root with Unit Disk Leaves Touching Graph Recongnition Problem.

1.1 Unit Disk Touching Graph Recongnition Problem (UDTGRP)

The UDTGRP is to determine when given an instance of a weighted graph (V,E) where each vertex has unit weight can be realized as a disk touching graph (a disk arrangement).

In this section we describe a particular family of unit weight graphs and corresponding 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, S_i, is as follows:

- Let v_0 be a vertex 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 $i - j$.

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. This is said to be the canonical position under Hausdorff distance of the snowflake tree.

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

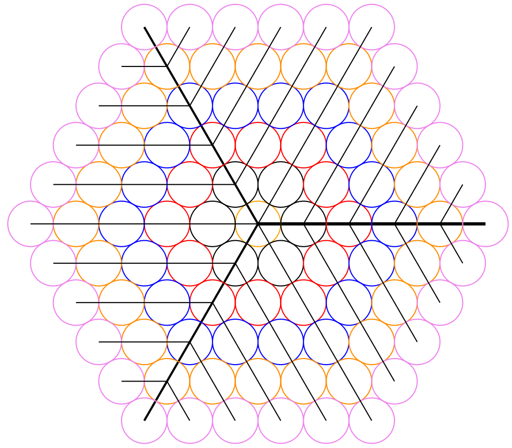

Figure 1.1: The same contact graph as in figure 1.2 overlayed with the a perfectly weighted snowflake tree.

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 $l = 1, ..., i$, there exists two paths attached to each vertex $v_{j,l}$. For S_i, each path attached to the k^{th} vertex of p_j, there are $i - k$ vertices. We will need to show that each of the $i - k$ vertices on each corresponding path are also in V.

The triangular lattice is symmetice under rotation about v_0 by $\frac{\pi}{3}$ radians. For each vertex $v_{1,l}$ for $l = 1, 2, ..., i - k$, we place two paths from it; the first path $\frac{\pi}{3}$ above p_1 at $v_{1,l}$ and $\frac{-\pi}{3}$ below p_1 at $v_{1,l}$ and call these paths $p^+_{1,l}$ 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 $v_{1,l}$. This shows that each of the $i-k$ vertices on $p^-_{1,l}$ and $p^+_{1,l}$ 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 1.2, we have a set of unit radius disks arranged in a manner that outlines regular, concentric hexagons.

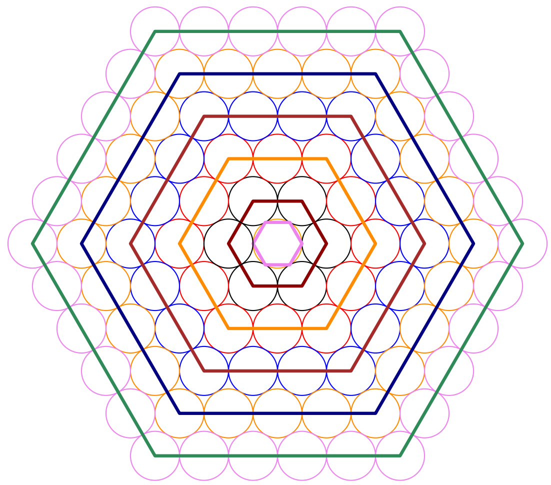

Figure 1.2: A contact graph that resembles the shape of concentric hexagons.

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 $d(P, G) < \varepsilon$

1.2 Perturbed Root with Unit Disk Leaves Touching Graph Recongnition Problem (PRUDTGRP)

The PRUDTGRP is to determine when given an instance of a weighted tree (V,E) where each vertex has unit weight with the exception of the root vertex having weight $\frac{1}{2} + \gamma$ where $\gamma > 0$ can be realized as a disk touching graph (a disk arrangement).

The perturbed snowflake follows the construction of the perfect snowflake with the exception of v_0 having weight $\frac{1}{2} + \gamma$ where $\gamma > 0$. 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 γ can modify the realization of a perfect snowflake S_i in the following ways:

1. **Modification of S_1.**

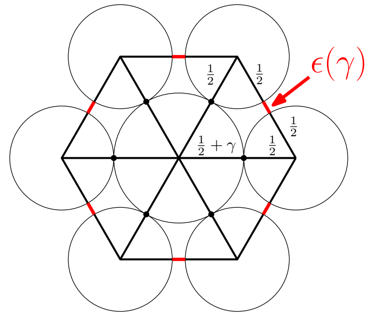

Figure 1.3: A canonical disk arrangement from a perturbed snowflake with 6 unit disks around a central disk with radius $\frac{1}{2}+\gamma$.

Figure 1.3 shows one realization of a perturbed S_1. For any realization of S_1, the perturbation γ can create gaps, $\varepsilon(\gamma)$ between adjacent disks that contact the root disk.

2. **Modification of disk placement corresponding to vertices of p_k.**

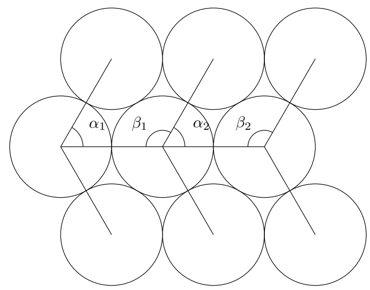

Figure 1.4: ?A?S

3. **Modification of disk placement along corresponding to the $p_{k,j}^{\mathbf{th}}$ vertex of S_i.** In Figure 1.4, we have a perturbed spine

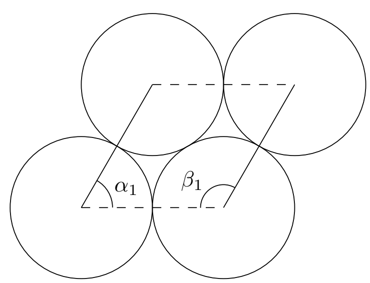

Figure 1.5: ?A?S

In Figure 1.5, we have four an arrangement of disks on the snowflake, off the spine and away from the central disk. We call this a vertebrae.

Lemma 1. *Given any realization of a perturbed snowflake of 7 weighted vertices, with the central vertex v_0 weighted $\frac{1}{2}+\gamma$ and the others weighted $\frac{1}{2}$, the total additional distance between all vertices is $6\varepsilon(\gamma)$ compared to a perfect snowflake of 7 unit weight vertices.*

Proof. Consider a canonical disk arrangement of a perturbed snowflake of 7 weighted vertices (see Figure 1.3). The side length of the sides formed between the center of the central disk and two adjacent disks around the central disk is $1+\gamma$. Let the distance between the two adjacent disks be $1+\varepsilon(\gamma)$. There are a total of $6\varepsilon(\gamma)$ between adjacent centers of disks. The total perimeter of the hexagon formed about the centers of the disks in contact with the central disk is $6+6\varepsilon(\gamma)$. Note that 1) the total perimeter of the hexagon formed on a perfect snowflake of 7 weighted vertices is 6 and 2) the canonical disk arrangement can be transformed to any other disk arrangement corresponding to the perturbed snowflake of 7 weighted vertices by pushing the the ring of disks around the central disk together such that all adjacent disks are in contact with each other with the exception of the disks at the end. □

1.3 On the Decidability of Problem ??

Proof. Consider a $k \times (\sqrt{3}k)$ rectangle section of a triangular lattice, and place disks of radius 1 at each grid point as in Fig. ?????. The contact graph of these disks contains 2-cycles. Consider the spanning tree T of the contact graph indicated in Fig. ????. The tree T decomposes into paths of collinear edges: T contains two paths along the two main diagonals, each containing $2k-1$ vertices; all other paths have an endpoint on a main diagonal. We now modify the disk arrangement to ensure that its contact graph is T. The disks along the main diagonal do not change. We reduce the radii of all other disks by a factor of $1-k^{-3}$ (as a result, they lose contact with other disks), and then successively translate them parallel in the direction of the shortest path in T to the main diagonal until the contact with the adjacent disk is reestablished. The Hausdorff distance between the union of these disks and the initial $k \times (\sqrt{3}k)$ rectangle is clearly less than 1. However, the contact tree T with these radii no longer has a unique realization (small perturbations are possible). To show stability, we argue by induction on the hop distance from the central disk. There are $O(i)$ disks at i hops from the central disk, most one which have radius $(1-k^{-3})\frac{1}{2}$. Since all radii are 1 or $(1-k^{-3})\frac{1}{2}$,the six neighbors of the central disk can differ from the regular hexagon by at most $O(k)$. Similarly, the disks at i hops from the center be off from the triangular grid pattern by $O(i2^{k-3})$, for $i = 1, 2, \ldots, k$.

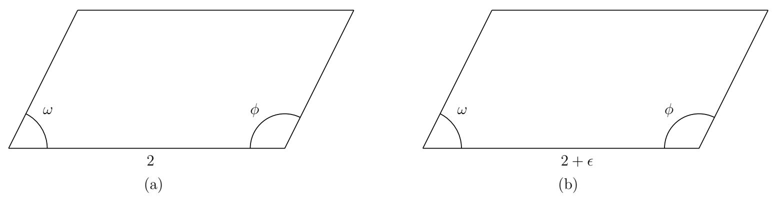

Figure 1.6

□

1.4 Hausdorff Distance

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

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

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

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

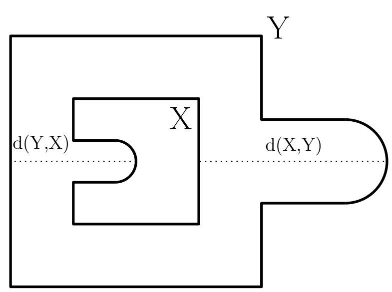

Figure 1.7: An illustrative example of $d(X,Y)$ and $d(Y,X)$ where X is the inner curve, and Y is the outer curve.

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.