Curve Reconstruction from Noisy Samples

Siu-Wing Cheng*
Dept. Comp. Sci.
HKUST, Hong Kong
scheng@cs.ust.hk

Stefan Funke†
MPI
Saarbrücken, Germany
funke@mpi-sb.mpg.de

Mordecai Golin‡
Dept. Comp. Sci.
HKUST, Hong Kong
golin@cs.ust.hk

Piyush Kumar§
Dept. Comp. Sci.
SUNY, USA
piyush@ams.sunysb.edu

Sheung-Hung Poon*
Dept. Comp. Sci.
HKUST, Hong Kong
hung@cs.ust.hk

Edgar Ramos†
Dept. Comp. Sci.
UIUC, USA
eramosn@cs.uiuc.edu

ABSTRACT

We present an algorithm to reconstruct a collection of disjoint smooth closed curves from n noisy samples. Our noise model assumes that the samples are obtained by first drawing points on the curves according to a locally uniform distribution followed by a uniform perturbation of each point in the normal direction with a magnitude smaller than the minimum local feature size. The reconstruction is faithful with a probability that approaches 1 as n increases. We expect that our approach can lead to provable algorithms under less restrictive noise models and for handling non-smooth features.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Geometrical problems and computations*

General Terms

Algorithms

*Partly supported by Research Grant Council, Hong Kong, China (project no. HKUST 6190/02E).

†Partly supported by the IST Programme of the EU as a Shared-cost RTD (FET Open) Project under Contract No IST-2000-26473 (ECG - Effective Computational Geometry for Curves and Surfaces).

‡Partly supported by Research Grant Council, Hong Kong, China (project no. HKUST 6082/01E and HKUST 6206/02E).

§Supported by National Science Foundation grant CCR-0098172

Keywords

Sampling, Curve reconstruction, Probabilistic analysis

1. INTRODUCTION

The combinatorial curve reconstruction problem has been extensively studied recently by computational geometers. The input consists of a set of sample points on an unknown curve F. The problem calls for computing a polygonal curve that is provably *faithful*. That is, as the sampling density increases, each point on a polygonal curve should converge to F, the normal at each point on the polygonal curve should converge to the normal at the nearest point on F, and the polygonal curve is homeomorphic to F. Most of the results obtained can handle the case where F is a set of disjoint curves.

Amenta et al. [3] obtained the first results in this problem. They proposed a *2D crust* algorithm to reconstruct a set of disjoint smooth closed curves. They prove that the reconstruction is faithful if the input satisfies the ϵ-sampling condition for $\epsilon < 0.252$. For $0 < \epsilon < 1$, a set S of samples is an *ϵ-sampling* of F if for any point $x \in F$, there exists $s \in S$ such that $\|s - x\| \leq \epsilon \cdot f(x)$ [3]. The algorithm by Amenta et al. invokes the computation of a Voronoi diagram or Delaunay triangulation twice. Gold and Snoeyink [11] simplified the algorithm and their algorithm invokes the computation of Voronoi diagram or Delaunay triangulation only once.

Later, Dey and Kumar [4] proposed a different NN-crust algorithm for disjoint smooth closed curves. Since we will use the NN-crust algorithm, we briefly describe it. For each sample s in S, connect s to its nearest neighbor in S. Afterwards, if a sample s is incident only on edge e, add the shortest edge incident to s among all the edges that make an angle more than $\pi/2$ with e. Dey and Kumar [4] proved that the resulting set of edges form a faithful reconstruction for $\epsilon \leq 1/3$.

Dey, Mehlhorn, and Ramos [5] proposed *conservative-crust* to handle the case when F has endpoints. Funke and Ramos [9] proposed an algorithm for the case where F may have sharp corners and endpoints. Dey and Wenger [6, 7] also described algorithms and implementation for handling sharp corners. Giesen [10] discovered that the traveling salesperson (TSP) tour through the samples is a reconstruction of the underlying curve. Althaus and Mehlhorn [2] showed that such a traveling salesperson tour can be constructed in polynomial time.

Noise often arises in collecting the input samples. For example, when the input samples are obtained from 2D images by scanning. The noisy samples are typically classified into two types. The first type are samples that cluster around F but they generally do not lie on F. The second type are outliers that lie relatively far from F. No combinatorial algorithm is known so far that can compute a faithful reconstruction in the presence of noise. In this paper, we propose a method that can handle noise of the first type for a set of disjoint smooth closed curves. We assume that the input does not contain outliers. Proving a deterministic theorem seems difficult as arbitrary noisy samples can collaborate to form patterns to fool any reconstruction algorithm. Instead, we assume a particular model of noise distribution and prove that our reconstruction is faithful with probability approaching 1 as the number of samples increases.

In our model, a sample is generated by drawing a point from F followed by randomly perturbing the point in the normal direction. For each point x on F, we denote $f(x)$ as the local feature size at x, which is the distance from x to the medial axis of F. Suppose we let $L = \int_F \frac{1}{f(x)}dx$. The drawing of points from F follows the probability density function $\frac{1}{L\cdot f(x)}$. This in fact means, the probability of drawing a point from a curve segment η is equal to $\frac{1}{L} \cdot \int_\eta \frac{1}{f(x)}dx$. A point p drawn from F is then perturbed in the normal direction. The perturbation is uniformly distributed within an interval that has p as the midpoint, width 2δ, and aligns with the normal direction at p. δ is a constant no more than $\min_{x\in F} f(x)$. The distribution of each sample is independently identical. Throughout this paper, we assume that $\min_{x\in F} f(x) = 1$. Note that if $\delta > 1$, then the perturbed points from different parts of F will mix up at some place and it seems very difficult to estimate the unknown curve F around that neighborhood. We emphasize that the constant δ is unknown to our algorithm. Although the perturbation along the normal direction is restrictive, it isolates the effect of noise from the distribution of samples on F. This facilities an initial study of curve reconstruction in the presence of noise.

Our algorithm returns a reconstruction which is faithful with probability at least $1 - O(n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)}(f_{\max}^2 + \sqrt{\ln n}))$, where n is the number of noisy samples and $f_{\max}$ is the maximum local feature size. Therefore, the probability approaches 1 as n increases. Our algorithm works for noisy samples from a collection of disjoint smooth closed curves. But we will just assume a single unknown curve F for simplicity. The novelty of our algorithm is a method to cluster samples so that each cluster comes from a relatively flat portion of F. This allows us to estimate points that lie close to F. We believe that this clustering approach will also be useful for less restrictive noise models and recognizing non-smooth features. We also expect that this clustering approach will work for surface reconstruction in 3D.

The rest of the paper is organized as follows. Section 2 describes our algorithm. In Section 3, we introduce the β-decomposition of the space around F which is the main tool in our probabilistic argument. Sections 4, 5, and 6 prove that our reconstruction is faithful with probability approaching 1. Section 7 discusses the possibility of extending the algorithm to handle curves with non-smooth features.

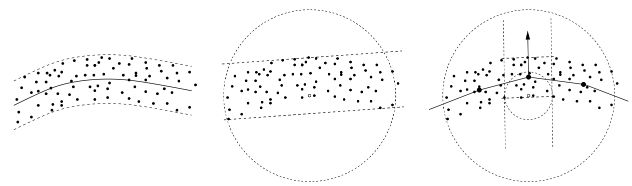

Figure 1: On the left, a smooth curve segment with a noise cloud. In the middle, a sufficiently large neighborhood identifies a strip with relatively large aspect ratio, which can provide preliminary point and normal estimates. On the right, concentrating on smaller neighborhoods, a better estimate of point and normal is possible.

2. ALGORITHM

We first highlight the key ideas in our algorithm. The algorithm works by growing a disk neighborhood around each sample p until a neighborhood coarse(p) is identified in which the samples it contains fit in a strip with small width relative to the neighborhood size. coarse(p) provides a first estimate of the curve locally and of its normal. A better estimation is possible by considering a smaller neighborhood refined(p). A possibility is to shrink coarse(p) by a certain factor and for each direction to consider all the samples in the slab bounded by the two lines parallel to the specified direction and tangent to the shrunken coarse(p); the spread of the points in the slab (along its direction) is minimized for a direction that provides a good normal estimation and it also provides a good point estimate s^*.[1] refined(p) is the optimal slab. Next, the sampling S is decimated: considering the samples in arbitrary order, put the current $p \in S$ into the decimated set S' and remove (decimate) all the points in a certain neighborhood that depends on refined(p) from consideration. Finally, we can use any reconstruction algorithm that is correct for a noise free sampling. For example, the NN-Crust from Dey and Kumar [4] in which each sample selects its nearest neighbor as its first reconstruction neighbor and its nearest neighbor in a small cone opposite to the first neighbor (the opposite halfplane in the original paper) as its second reconstruction neighbor.

We provide the details of our algorithm below. Let n be the total number of input samples.

Point Estimation: For each sample s, we estimate a point as follows.

Coarse neighborhood: Let initial(s) be the disk that is centered at s and contains $\ln^2 n$ samples. We initialize coarse(s) = initial(s) and compute an infinite strip strip(s) of minimum width that contains all samples inside coarse(s). We grow coarse(s) and maintain strip(s) until the ratio of the radius of coarse(s) to the width of strip(s) is greater than a predefined constant ρ. The final disk coarse(s) is the *coarse neighborhood* of s.

Refined neighborhood: Let N_s be a direction perpendicular to the long side of strip(s). Let $\lambda_1 =$

[1]For technical reasons, this is done less elegantly in our provable algorithm.

$\ln n/\sqrt{n}$. The *refined neighborhood* refined(s) is the slab that contains s in the middle, parallel to N_s, and has width equal to

$$\max\{\sqrt{\lambda_1}\,\text{radius}(\text{coarse}(s)), \text{radius}(\text{initial}(s))\}.$$

Divide refined(s) into three subslabs of equal width. Let a (resp. b) be the sample in the leftmost (resp. rightmost) subslab furthest from s in the direction of refined(s). Let line(s) be the line through a and b. We rotate refined(s) in the clockwise and anti-clockwise direction and maintain line(s) . The range of the rotation is $[0, \pi/12]$. Within this range, we position refined(s) such that the angle between refined(s) and line(s) is maximized.

Center point: Compute the minimum rectangle R_s that aligns with refined(s) and contains all samples inside refined(s). We return the center point s^* of R_s.

Pruning: We select an arbitrary center point s^* and eliminate all center points whose distances from s are less than or equal to $\sqrt{\text{width}(\text{refined}(s))}$. We repeat until no center point is left. The final set of center points selected is the set of estimated points on which the NN-crust algorithm will be invoked.

3. β-DECOMPOSITION

We decompose a neighborhood of F into a collection of cells. The diameter of these cells and the probabilities of them being non-empty or containing a certain number of samples will be useful in the analysis of our algorithm.

For each point $x \in \mathbb{R}^2$ that does not lie on the medial axis of F, we use $\tilde{x}$ to denote the point on F closest to x. (We are not interested in points on the medial axis.) We call the bounded region enclosed by F the *inside* of F and the unbounded region the *outside* of F. For $0 < \alpha \leq \delta$, F_α^+ (resp. F_α^-) is the curve that passes through each point q inside (resp. outside) F such that $\|q - \tilde{q}\| = \alpha$. We use F_α to mean F_α^+ or F_α^- when it is unimportant to distinguish between inside and outside. The *normal segment* at a point $p \in F$ is the line segment consisting of points q on the normal of F at p such that $\|p - q\| \leq \delta$.

Given two points $x, y \in F$, we use $F(x, y)$ to denote the curved segment from x to y in clockwise direction around F. We use $|F(x, y)|$ to denote the length of $F(x, y)$. Let $0 < \beta < 1$ be a parameter. If $\delta \geq 1/\sqrt{\ln n}$, let $\kappa = \beta\delta$, otherwise let $\kappa = \beta$. We identify a set of *cut-points* on F as follows. We pick an arbitrary point c_0 on F as the first cut-point. Then for $i \geq 1$, we find the point c_i such that c_i lies in the interior of $F(c_{i-1}, c_0)$, $|F(c_{i-1}, c_i)| = \kappa f(c_{i-1})$, and $|F(c_i, c_0)| \geq \kappa f(c_i)$. If c_i exists, it is the next cut-point and we continue. Otherwise, we have computed all the cut-points and we stop.

A β-decomposition is the arrangement of the following curves and line segments:

1. The normal segments at the cut-points.

2. F, F_δ^+, and F_δ^-.

3. If $\delta \geq 1/\sqrt{\ln n}$, we also use F_α^+ and F_α^- where $\alpha = i\beta\delta$ and i is an integer between 1 and $\lfloor 1/\beta \rfloor - 1$.

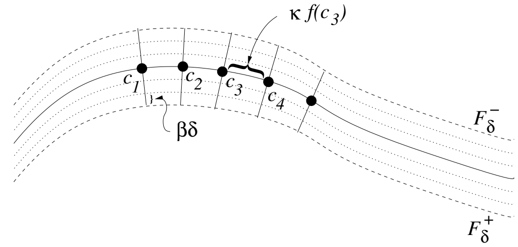

Figure 2: A β-decomposition.

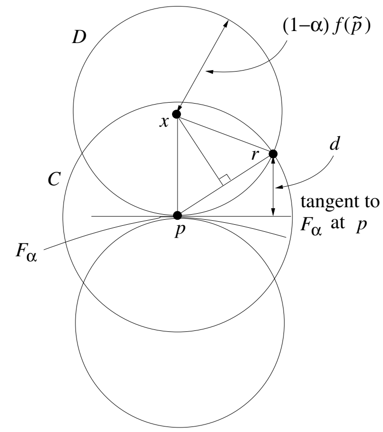

Figure 3: Illustration for Lemma 2.

If $\delta < 1/\sqrt{\ln n}$, the β-decomposition consists of two rows of cells on the two sides of F. Otherwise, there are $O(1/\beta)$ rows of cells.

3.1 Properties of F_α

Lemma 1 *Any point p on F_α has two tangent disks with radii $(1 - \alpha)f(\tilde{p})$ whose interior do not intersect F_α.*

Proof. Let M_α be the medial disk of F_α touching a point $p \in F_\alpha$. Let M be the medial disk of F touching $\tilde{p}$. By the definition of F_α, M and M_α have the same center and radius(M_α) = radius(M) $- \alpha \geq f(\tilde{p}) - \alpha$. Let D be a disk of radius $(1-\alpha)f(\tilde{p})$ that touches F_α at p. If $\text{int}(D) \cap F_\alpha \neq \emptyset$, then radius($M_\alpha$) $< (1 - \alpha)f(\tilde{p}) \leq f(\tilde{p}) - \alpha$. This is a contradiction. ∎

Lemma 2 *Let p be a point on F_α. For any point q on F_α, if $\|p - q\| < 2(1 - \alpha)f(\tilde{p})$, then the distance of q from the tangent at p is at most $\frac{\|p-q\|^2}{2(1-\alpha)f(\tilde{p})}$.*

Proof. Assume that the tangent at p is horizontal. Refer to Figure 3. Let D be the tangent disk at p that lies above p and has center x and radius $(1 - \alpha)f(\tilde{p})$. Let C be the circle centered at p with radius $\|p - q\|$. Since $\|p - q\| < 2(1 - \alpha)f(\tilde{p})$, C crosses D. Let r be a point in $C \cap \partial D$. Let d be the distance of r from the tangent at p. By Lemma 1, d bounds the distance from q to the tangent at p. Observe that $\|p - q\| = \|p - r\| = 2(1 - \alpha)f(\tilde{p})\sin(\angle pxr/2)$ and $d = \|p-r\|\cdot\sin(\angle pxr/2)$. Thus, $d = 2(1-\alpha)f(\tilde{p})\sin^2(\angle pxq/2) = \|p - q\|^2/(2(1 - \alpha)f(\tilde{p}))$. ∎

For each point p on F_α, define cocone(p, θ) to be the double cone that has apex p and angle θ such that the normal at p is the symmetry axis of cocone(p, θ) that lies outside cocone(p, θ).

Lemma 3 *Let p be a point on F_α. Let D be the disk with center p and radius $c \cdot f(\tilde{p})$ for some $c > 0$. Assume that $\delta \leq \min\{1/2, 1-2c\}$. Then $F_\alpha \cap D \subseteq \text{cocone}(p, \theta)$, where $\theta = \min\{2 \sin^{-1} c, \pi/6\}$.*

Proof. Since $\delta \leq 1 - 2c$, $c < 2(1-\delta) \leq 2(1-\alpha)$, Lemma 2 implies that the distance between $F_\alpha \cap D$ and the tangent at p is bounded by $c^2 f(\tilde{p})/(2(1-\delta))$. Let θ be the smallest angle such that cocone(p, θ) contains $F_\alpha \cap D$. Thus,

$$\sin \frac{\theta}{2} \leq \frac{c^2}{2(1-\delta)} \cdot \frac{1}{c} = \frac{c}{2(1-\delta)}.$$

Observe that $2(1-\delta) \geq 1$, so $c/(2(1-\delta)) \leq c$ which implies that $\theta \leq 2 \sin^{-1} c$. Also, since $\delta \leq 1-2c$, $c/(2(1-\delta)) < 1/4$ which implies that $\theta < \pi/6$. ◻

3.2 Diameter of a cell

In this section, we prove an upper bound and lower bound on the diameter of a β-cell. First, we need a technical lemma.

Lemma 4 *Let p be a point on F and let D be a disk centered at p with radius $c \cdot f(p)$ where $c \leq 1$. $F \cap D$ consists of one connected component and for any point $q \in F \cap D$, the normals at p and q make an angle at most $2 \sin^{-1}(c/2)$.*

Proof. If $F \cap D$ consists of more than one connected component, the medial axis of F contains some point inside D. This contradicts the fact that radius$(D) \leq f(x)$. From a result in [3], the angle between the normal at p and pq is at least $\pi/2 - \sin^{-1}(c/2)$. The same is true for the angle between the normal at q and pq. It follows that the angle between the normals at p and q is at most $2\sin^{-1}(c/2)$. ◻

Lemma 5 *Assume that $\beta \leq 1/4$. Let C be a β-cell. There is a constant $c_1 > 0$ such that for any point $t \in C$, if $\delta \geq 1/\sqrt{\ln n}$, the diameter of C is at most $c_1(\beta\delta f(\tilde{t}))$; otherwise, it is at most $c_1 \beta f(\tilde{t}) + 2\delta$.*

Proof. Suppose that the cell C lies between the normal segments at the cut-points c_i and c_{i+1}. For any point $t \in C$, the Lipschitz condition implies $f(\tilde{t}) \geq f(c_i) - \|c_i - \tilde{t}\| \geq (1-2\beta) f(c_i)$. Symmetrically, we have $f(\tilde{t}) \leq (1+2\beta) f(c_i)$. Assuming that $\beta \leq 1/4$, the above two inequalities imply that for any points $s, t \in C$

$$f(\tilde{t})/3 \leq f(\tilde{s}) \leq 3 f(\tilde{t}). \quad (1)$$

If $\delta < 1/\sqrt{\ln n}$, the diameter of C is clearly at most $2\beta f(c_i) + 2\delta$ which is at most $O(\beta f(\tilde{t})) + 2\delta$ by (1).

Suppose that $\delta \geq 1/\sqrt{\ln n}$. Refer to Figure 4. Let p be the projection of s onto a latitudinal side of C along the direction normal F to $\tilde{s}$. Similarly, let q be the projection of t onto the same latitudinal side. Let r be the point $q - \tilde{t} + \tilde{s}$. Without loss of generality, assume that $\angle \tilde{s}pr \geq \angle \tilde{s}rp$. By sine law,

$$\|p - r\| = \|r - \tilde{s}\| \cdot \sin \angle p\tilde{s}r / \sin \angle \tilde{s}pr.$$

By Lemma 4, the angle between the normals at $\tilde{s}$ and c_i

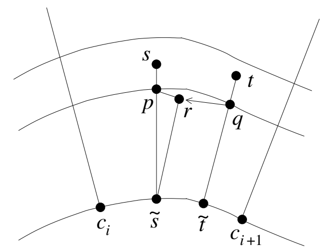

Figure 4: Illustration for Lemma 5.

is at most $2\sin^{-1}(\beta)$. The same holds for the normals at $\tilde{t}$ and c_i. It follows that the angle between the normals at $\tilde{s}$ and $\tilde{t}$ is at most $4\sin^{-1}(\beta) \leq 8\beta$ as $\beta \leq 1/4$. Therefore, $\angle \tilde{s}pr \geq \pi/2 - 4\beta$. So we get $\|p - r\| \leq \delta \sin(8\beta)/\cos(4\beta) \leq 8\beta\delta/\cos(4\beta)$ which is less than $16\beta\delta$ as $\cos(4\beta) \geq 1/2$. By triangle inequality, we get $\|p - q\| \leq \|p - r\| + \|q - r\| = \|p - r\| + \|\tilde{s} - \tilde{t}\| \leq 16\beta\delta + 2\beta\delta f(c_i)$. Hence,

$$\begin{aligned} \|s - t\| &\leq \|p - s\| + \|p - q\| + \|q - t\| \\ &\leq \beta\delta + +16\beta\delta + 2\beta f(c_i) + \beta\delta \\ &\overset{(1)}{=} O(\beta\delta f(\tilde{t})). \end{aligned}$$

◻

Lemma 6 *Assume that $\delta < 1/16$ and $\beta \leq 1/4$. Let C be a β-cell. There is a constant $c_2 > 0$ such that for any point $t \in C$, if $\delta \geq 1/\sqrt{\ln n}$, the diameter of a β-cell C is at least $c_2 \beta\delta f(\tilde{t})$; otherwise, it is at least $c_2 \beta f(\tilde{t})$.*

Proof. We only deal with the case where $\delta \geq 1/\sqrt{\ln n}$. Similar analysis works for the case where $\delta < 1/\sqrt{\ln n}$. Let $F_\alpha(p, q)$ be one side of C. Let $l = |F_\alpha(p, q)|$ and let $\tilde{l} = |F(\tilde{p}, \tilde{q})|$. Observe that

$$\beta\delta f(\tilde{p}) \leq \tilde{l} \leq 2\beta\delta f(\tilde{p}). \quad (2)$$

For any point $s \in F_\alpha(p, q)$, we use θ_s to denote the acute angle between the normals at s and p. We lower bound the diameter of C by lower bounding l.

By Lemma 4, we have

$$\theta_s \leq 2 \sin^{-1} \beta\delta \leq 4\beta\delta \leq 1, \quad (3)$$

as $\beta \leq 1/4$. $F_\alpha(p, q)$ is monotone with respect to the tangent at p. Otherwise, there is a point $s \in F_\alpha(p, q)$ such that $\theta_s = \pi/2$, contradiction. Refer to Figure 5. Assume that the tangents at p and $\tilde{p}$ are horizontal. Let r and r' be the vertical projections of q and $\tilde{q}$ onto the tangents at p and $\tilde{p}$, respectively. Let s (resp. s') be the intersection between the normal at q and the tangent at p (resp, $\tilde{p}$). The monotonicity of $F_\alpha(p, q)$ implies the following.

$$\|p - r\| \geq \int_{F_\alpha(p,q)} \cos \theta_s \, ds \geq l \cdot \cos(4\beta\delta) \geq l/2, \quad (4)$$

as $\beta \leq 1/4$. Similarly, we get

$$\|\tilde{p} - r'\| \geq \tilde{l}/2. \quad (5)$$

We lower bound l by lower bounding $\|p-r\|$. By equation (1) in the proof of Lemma 5, it suffices to show that $\|p-r\| = \Omega(\beta\delta f(\tilde{p}))$.

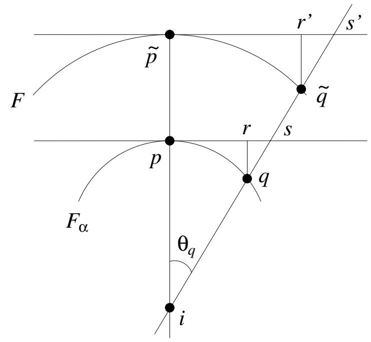

Figure 5: Illustration for Lemma 6.

If $l \geq \tilde{l}$, then $\|p-r\| \geq \tilde{l}/2 \geq \beta\delta f(\tilde{p})/2$. Assume that $l < \tilde{l}$. So $F \neq F_\alpha$. By (3), $\tan\theta_q \leq 2\theta_q \leq 8\beta$. We have $\|p-r\| = \|p-s\| - \|r-s\| = \|p-s\| - \|q-r\| \cdot \tan\theta_q$. Thus,

$$\|p-r\| \geq \|p-s\| - \|q-r\| \cdot 8\beta \tag{6}$$

Starting with Lemma 2, we get $\|q-r\| \leq l^2/(2(1-\alpha)f(\tilde{p})) < \tilde{l}^2/(2(1-\delta)f(\tilde{p}))$ which is $O(\beta^2\delta^2 f(\tilde{p}))$ by (2).

Consider the similar triangles ips and $i\tilde{p}s'$. We have $\|\tilde{p}-i\| = \|\tilde{p}-s'\|/\tan\theta_q \geq \|\tilde{p}-r'\|/8\beta\delta$ which is at least $f(\tilde{p})/16$ by (2) and (5). This implies that $\|p-i\|/\|\tilde{p}-i\| = 1 - \|p-\tilde{p}\|/\|\tilde{p}-i\| \geq 1 - 16\delta/f(\tilde{p}) \geq 1-16\delta$. Thus, $\|p-s\| = \|\tilde{p}-s'\| \cdot \|p-i\|/\|\tilde{p}-i\| = \Omega(\|\tilde{p}-r'\|)$ which is $\Omega(\beta\delta f(\tilde{p}))$ by (2) and (5).

Substituting the bounds for $\|p-s\|$ and $\|q-r\|$ into (6), we get $\|p-r\| \geq \Omega(\beta\delta f(\tilde{p})) - O(\beta^2\delta^2 f(\tilde{p})) = \Omega(\beta\delta f(\tilde{p}))$. ◻

3.3 Number of samples in a cell

The following lemma estimates the probability of a sample point lying inside a cell.

Lemma 7 *Let C be a cell in a β-decomposition. There exists constants c_3 and c_4 such that if $\delta \geq 1/\sqrt{\ln n}$, the probability that a randomly picked sample falls inside C is at least $c_4\beta^2\delta$ and at most $c_3\beta^2\delta$; otherwise, the probability is between $c_4\beta$ and $c_3\beta$.*

Proof. Assume that C is bounded by normal segments at cut-points c_i and c_{i+1}. We use η to denote $F(c_i, c_{i+1})$ as a short hand. For any point $s \in \eta$, $\|s-c_i\| \leq |\eta| \leq 2\beta f(c_i)$. By the Lipschitz property, we have

$$f(c_i) - \|s-c_i\| \leq f(s) \leq f(c_i) + \|s-c_i\|$$

Assume that $L = \int_F \frac{1}{f(s)} ds$. Let p be a randomly picked sample.

If $\delta < 1/\sqrt{\ln n}$, the probability that $p \in C$ is equal to the probability that $\tilde{p}$ lies on η, which is

$$\frac{1}{L} \cdot \int_\eta \frac{1}{f(s)} ds \in \left[\frac{\beta}{(1+2\beta)L},\ \frac{2\beta}{(1-2\beta)L}\right].$$

If $\delta \geq 1/\sqrt{\ln n}$, the probability that $\tilde{p}$ lies on η is

$$\frac{1}{L} \cdot \int_\eta \frac{1}{f(s)} ds \in \left[\frac{\beta\delta}{(1+2\beta\delta)L},\ \frac{2\beta\delta}{(1-2\beta\delta)L}\right].$$

The probability that p lies inside C on the condition that $\tilde{p}$ lies on η is $\beta\delta/\delta = \beta$. We conclude that the probability that p lies inside C is equal to

$$\frac{1}{L} \cdot \int_\eta \frac{\beta}{f(s)} ds \in \left[\frac{\beta^2\delta}{(1+2\beta)L},\ \frac{2\beta^2\delta}{(1-2\beta)L}\right].$$

◻

The following Chernoff bound [8] will be needed.

Lemma 8 *Let the random variables $X_1, X_2, \ldots, X_n$ be independent, with $0 \leq X_k \leq 1$ for each k. Let $S_n = \sum X_k$, and let $E(S_n)$ be the expected value of S_n. Then for any $\sigma > 0$, $\Pr(S_n \leq (1-\sigma)E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2})$, and $\Pr(S_n \geq (1+\sigma)E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2(1+\sigma/3)})$.*

Lemma 9 *Let $\lambda_k = k\ln n/\sqrt{n}$. Assume that $\delta \geq 1/\sqrt{\ln n}$. Let C be a cell in a β-decomposition. Let c_3 and c_4 be the constants in Lemma 7.*

(i) If $\beta = \lambda_k/r$ for some $k, r > 0$, then C is non-empty with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/r^2)}$.

(ii) If $\beta = \lambda_k/r$ for some $k, r > 0$, then for any constant $c > c_3k^2/r^2$, the number of samples in C is at most $c\ln^2 n$ with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/r^2)}$.

(iii) If $\beta = \lambda_k/r$ for some $k, r > 0$, then for any constant $c < c_4k^2/r^2$, the number of samples in C is at least $c\ln^{3/2} n$ with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/r^2)}$.

Proof. Let C be a (λ_k/r)-cell. Let $X_i (i = 1, \ldots, n)$ be a random binomial variable taking value 1 if the sample point p_i is inside C, and value 0 otherwise. Let $S_n = \sum_{i=1}^n X_i$. Then $E(S_n) = \sum_{i=1}^n E(X_i) = n \cdot \Pr(p_i \text{ lies inside } C)$. This implies $E(S_n) \leq c_3 n\lambda_k^2\delta/r^2$ and $E(S_n) \geq c_4 n\lambda_k^2\delta/r^2 \geq c_4k^2\ln^{3/2} n/r^2$.

By Lemma 8, $\Pr(S_n \leq 0) = \Pr(S_n \leq (1-1)E(S_n)) \leq \exp(-\frac{E(S_n)}{2}) = \exp(-\Omega(\frac{\ln^{3/2} n}{r^2})) = n^{-\Omega(\sqrt{\ln n}/r^2)}$. This proves (i). Consider (ii). Let $\sigma = (cr^2/(c_3k^2)) - 1 > 0$. Then $c\ln^2 n = c_3 n\lambda_k^2(1+\sigma)/r^2 \geq (1+\sigma)E(S_n)$. By Lemma 8, $\Pr(S_n > c\ln^2 n) \leq \Pr(S_n > (1+\sigma)E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2+2\sigma/3}) = \exp(-\Omega(\frac{\ln^{3/2} n}{r^2})) = n^{-\Omega(\sqrt{\ln n}/r^2)}$. Consider (iii). Let $\sigma = 1 - (cr^2/(c_4k^2)) > 0$. Then $c\ln^{3/2} n = c_4k^2\ln^{3/2} n(1-\sigma)/r^2 \leq (1-\sigma)E(S_n)$. By Lemma 8, $\Pr(S_n < c\ln^{3/2} n) \leq \Pr(S_n < (1-\sigma)E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2}) = \exp(-\Omega(\frac{\ln^{3/2} n}{r^2})) = n^{-\Omega(\sqrt{\ln n}/r^2)}$.

◻

Lemma 10 *Let $\lambda_k = k\ln n/\sqrt{n}$. Assume that $\delta < 1/\sqrt{\ln n}$. Let C be a cell in a β-decomposition. Let c_3 and c_4 be the constants in Lemma 7.*

(i) If $\beta = \lambda_k^2$ for some $k > 0$, then C is non-empty with probability at least $1 - n^{-\Omega(\ln n)}$.

(ii) If $\beta = \lambda_k^2$ for some $k > 0$, then for any constant $c > c_3k^2$, the number of samples in C is at most $c\ln^2 n$ with probability at least $1 - n^{-\Omega(\ln n)}$.

(iii) If $\beta = \lambda_k^2$ for some $k > 0$, then for any constant $c < c_4 k^2$, the number of samples in C is at least $c \ln^2 n$ with probability at least $1 - n^{-\Omega(\ln n)}$.

Proof. Let C be a λ_k-cell. Let $X_i (i = 1, \ldots, n)$ be a random binomial variable taking value 1 if the sample point p_i is inside C, and value 0 otherwise. Let $S_n = \sum_{i=1}^{n} X_i$. Then $E(S_n) = \sum_{i=1}^{n} E(X_i) = n \cdot \Pr(p_i \text{ lies inside } C)$. This implies $c_4 n \lambda_k^2 \leq E(S_n) \leq c_3 n \lambda_k^2$.

Consider (i). By Lemma 8, $\Pr(S_n \leq 0) = \Pr(S_n \leq (1-1)E(S_n)) \leq \exp(-\frac{E(S_n)}{2}) = \exp(-\Omega(\ln^2 n)) = n^{-\Omega(\ln n)}$. Consider (ii). Let $\sigma = (c/(c_3 k^2)) - 1 > 0$. Then $c \ln^2 n = c_3 n \lambda_k^2 (1 + \sigma) \geq (1 + \sigma) E(S_n)$. By Lemma 8, $\Pr(S_n > c \ln^2 n) \leq \Pr(S_n > (1 + \sigma)E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2 + 2\sigma/3}) = \exp(-\Omega(\ln^2 n)) = n^{-\Omega(\ln n)}$. Consider (iii). Let $\sigma = 1 - (c/(c_4 k^2)) > 0$. Then $c \ln^2 n = c_4 n \lambda_k^2 (1 - \sigma) \leq (1 - \sigma) E(S_n)$. By Lemma 8, $\Pr(S_n < c \ln^2 n) \leq \Pr(S_n < (1 - \sigma) E(S_n)) \leq \exp(-\frac{\sigma^2 E(S_n)}{2}) = \exp(-\Omega(\ln^2 n)) = n^{-\Omega(\ln n)}$. ◻

4. COARSE NEIGHBORHOOD

In this section, we prove an upper bound and a lower bound on the radius of coarse(s) for each sample s. Recall that $\lambda_k = k \ln n / \sqrt{n}$.

Lemma 11 *Assume $\delta \leq 1/(6\rho + 2)$. Let s be a sample. There exists n_0 such that when $n > n_0$,* radius(coarse(s)) $\leq 3\rho\delta f(\tilde{s}) +$ radius(initial(s)) *and the following hold with probability at least $1 - n^{-\Omega(\sqrt{\ln n})}\sqrt{\ln n}$.*

- *If $\delta \geq 1/\sqrt{\ln n}$,* radius(initial($s$)) $= O(\lambda_1 \delta f(\tilde{s}) \sqrt{\ln n})$;
- *If $\delta < 1/\sqrt{\ln n}$,* radius(initial(s)) $= O(\lambda_1^2 f(\tilde{s})) + 2\delta$.

Proof. If $\delta < 1/\sqrt{\ln n}$, by Lemma 10(iii), initial(s) does not contain the cell C containing s in some λ_k^2-decomposition with probability at least $1 - n^{-\Omega(\ln n)}$. So radius(initial(s)) is at most the diameter of C, which is at most $c_1 \lambda_k^2 f(\tilde{s}) + 2\delta = O(\lambda_1^2 f(\tilde{s})) + 2\delta$. (The big-Oh constant absorbs the constant k.)

If $\delta \geq 1/\sqrt{\ln n}$, by Lemma 9(iii), initial(s) does not contain more than $\sqrt{\ln n}$ cells in some λ_k-decomposition with probability at least $1 - n^{-\Omega(\sqrt{\ln n})}\sqrt{\ln n}$. So radius(initial($s$)) is at most the sum of diameters of $\sqrt{\ln n}$ contiguous cells, including the one containing s. By the Lipschitz condition, the local feature sizes differ by constant factors within these cells. It follows from Lemma 5 that radius(initial(s)) $\leq c' \lambda_k \delta f(\tilde{s}) \sqrt{\ln n}$ for some constant c', which is $O(\lambda_1 \delta f(\tilde{s}) \sqrt{\ln n})$. (The big-Oh constant absorbs the constant k.)

We use K to denote $c_1 \lambda_k^2 f(\tilde{s}) + 2\delta$ or $c' \lambda_k f(\tilde{s}) \sqrt{\ln n}$. Let n_0 be the constant such that $K \leq 1/2$ whenever $n > n_0$.

Let s_1 and s_2 be points on F_δ^+ and F_δ^- such that $\tilde{s_1} = \tilde{s_2} = \tilde{s}$. Let D be the disk centered at s with radius $3\rho\delta f(\tilde{s}) +$ radius(initial(s)). So both s_1 and s_2 lie inside D. Since $\delta \leq 1/(6\rho + 2)$ and $K \leq 1/2$, $3\rho\delta + K < (1 - \delta)$. Thus, the distance between any two points in $D \cap F_\delta^+$ is less than $2(1 - \delta) f(\tilde{s})$. By Lemma 2, the maximum distance between $D \cap F_\delta^+$ and the tangent at s_1 is at most $(3\rho\delta + K)^2 f(\tilde{s})/(2(1 - \delta)) \leq (3\rho\delta + K) f(\tilde{s})/(2(1 - \delta))$. The same is also true for $D \cap F_\delta^-$. It follows that the samples inside D lie inside a strip of width at most $2\delta + (3\rho\delta + K) f(\tilde{s})/(1 - \delta)$. This is at

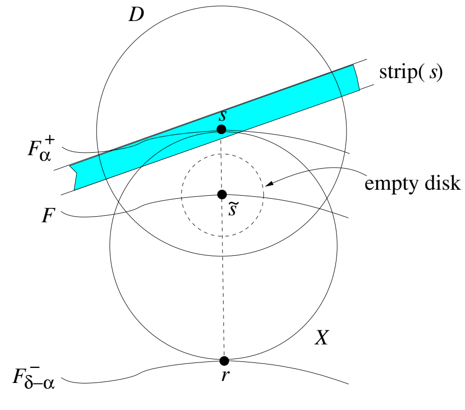

Figure 6: For the proof of Lemma 12.

most $3\delta f(\tilde{s})$ as $\delta \leq 1/(6\rho + 2)$ and $K \leq 1/2$. Thus, coarse(s) cannot grow beyond D. ◻

Lemma 12 *Let $f_{\max}$ be the maximum local feature size. Let s be a sample. There exists n_0 such that whenever $n > n_0$,* radius(coarse(s)) $\geq \sqrt{\rho}\delta$ *and one of the following holds with probability at least $1 - O(f_{\max}^2 n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)})$.*

- *If $\delta \geq 1/\sqrt{\ln n}$,* radius(initial($s$)) $= \Omega(\lambda_1 \delta \sqrt{f(\tilde{s})})$.
- *If $\delta < 1/\sqrt{\ln n}$,* radius(initial(s)) $= \Omega(\lambda_1^2 f(\tilde{s}))$.

Proof. Let D be the disk that has center s and radius radius(coarse(s))$/\sqrt{\rho}$. The width of strip(s) is less than or equal to radius(coarse(s))$/\rho =$ radius(D)$/\sqrt{\rho}$. Refer to Figure 6. Assume that $s \in F_\alpha^+$. Let r be the point on $F_{\delta-\alpha}^-$ such that $\tilde{r} = \tilde{s}$. There is a disk X with radius $\delta/2$ that is tangent to F_α^+ at s and $F_{\delta-\alpha}^-$ at r. If radius(coarse(s)) $< \sqrt{\rho}\delta$, then radius(D) $< \delta$ and so $D \cap X$ contains a disk with radius radius(D)/2. Thus, $(D \cap X) -$ strip(s) contains a disk Y with radius $((\frac{1}{4} - \frac{1}{2\sqrt{\rho}})$radius($D$)). Note that Y must be empty. We seek a lower bound on radius(Y) in order to apply Lemma 9(i) or Lemma 10(i) to conclude that Y is empty with small probability. This will complete the proof. Let $R =$ radius(initial(s)).

Consider the case where $\delta \geq 1/\sqrt{\ln n}$. Assume that $R < \lambda_1 \delta f(\tilde{s})$. Let c be some constant to be determined later. Take some λ_k-decomposition such that each cell has at most $c \ln^2 n$ samples with probability at least $1 - n^{-\Omega(\ln n)}$. Let C_i be any cell intersected by initial(s). Assume that C_i projects onto $F(c_i, c_{i+1})$. We divide C_i into boxes with normal segments at distance $\lambda_k \delta$ apart. By adapting the proof in Lemma 9(ii), one can show that each box contains at most $c \ln^2 n / f(c_i)$ points with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/f(c_i))} \geq 1 - n^{-\Omega(\sqrt{\ln n}/f_{\max})}$. By the Lipschitz condition, $c \ln^2 n / f(c_i) \leq c \ln^2 n / (\mu f(\tilde{s}))$ for some $0 < \mu < 1$. Each box has diameter $\Theta(\lambda_k \delta)$ and bounded aspect ratio. Assume that the box containing s and others at no more than $\mu f(\tilde{s})/c$ boxes away all contain at most $c \ln^2 n / (\mu f(\tilde{s}))$ points. This event happens with probability at least $1 - O(f_{\max}^2 n^{-\Omega(\sqrt{\ln n}/f_{\max})})$. If initial($s$) does not contain any box completely, then initial(s) intersects at most K boxes for some constant K. Therefore, if we choose c such that $cK/(\mu f_{\max}) < 1$, then in order that initial(s) contains $\ln^2 n$ points, initial(s) must contain some box. Since initial(s) contains some box, all boxes intersected by initial(s) are contained by some constant factor expansion of initial(s). Since initial(s) contains $\ln^2 n$ points, initial(s) intersects at least $\mu f(\tilde{s})/c$ boxes, including the one containing s. A packing argument implies that $R = \Omega(\lambda_1 \delta \sqrt{f(\tilde{s})}) = \Omega(\lambda_1 \delta)$. Since radius$(Y) = \Omega(R)$, Y contains a cell in some $(\lambda_l / f_{\max})$-decomposition. By Lemma 9(i), Y is non-empty with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)}$.

Consider the case where $\delta < 1/\sqrt{\ln n}$. Take some λ_k^2-decomposition such that each cell contains at most $\ln^2 n/3$ points with probability at least $1 - n^{-\Omega(\ln n)}$. Assume that $\tilde{s}$ lies on $F(c_i, c_{i+1})$. If initial(s) intersects normal segments at two cut points or more, then clearly $R = \Omega(|F(c_i, c_{i+1})|) = \Omega(\lambda_k^2 f(\tilde{s})) = \Omega(\lambda_1^2 f(\tilde{s}))$, If initial$(s)$ intersects normal segments at one cut point or less, then initial(s) intersects at most two cells. It follows that initial(s) contains less than $2\ln^2/3$ points with probability at least $1 - O(n^{-\Omega(\ln n)})$. So $R = \Omega(\lambda_1^2 f(\tilde{s}))$ with probability at least $1 - O(n^{-\Omega(\ln n)})$. Given this lower bound on R, Y contains a cell in some λ_r^2-decomposition and so Y is non-empty with probability at least $1 - n^{-\Omega(\ln n)}$. ◻

5. CRUDE ANGLE ESTIMATES

Lemma 13 *Assume that $\delta \le 1/(40\rho^{3/2})$. Let s be a sample. Let γ be the angle between the normal at $\tilde{s}$ and the normal of the long side of* strip(s). *There exists n_0 such that whenever $n > n_0$, $\gamma \le \sin^{-1}(5/\sqrt{\rho})$ with probability at least $1 - n^{-\Omega(\sqrt{\ln n}/f_{\max})}$.*

Proof. Assume that the tangent at s is horizontal and $s \in F_\alpha$. Let ℓ be the lower boundary of strip(s). Let w be the width of strip(s). Let x be the intersection between ℓ and the tangent at s. Let R denote radius(coarse(s)). Without loss of generality, assume that the slope of ℓ is positive. Let γ be the acute angle between ℓ and the horizontal. Refer to Figure 7

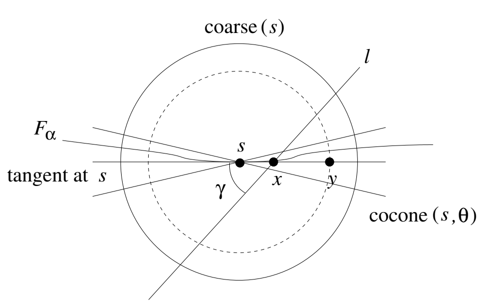

Figure 7: Illustration for Lemma 13.

Consider the case where $\|s - x\| \ge R/16$. The width of strip(s) is equal to $\|s - x\| \sin\gamma$. It follows that $\sin\gamma \le 16/\rho$. Consider the case where $\|s - x\| < R/16$. Suppose that $\sin\gamma > 5/\sqrt{\rho}$. Let y be the point on the tangent at s such that $\|s - y\| = 15R/16$ and x lies between s and y. By Lemma 3, F_α lies inside cocone(s, θ) where $\theta \le 2\sin^{-1}(R/f(\tilde{s}))$. By Lemma 11, for sufficiently large n, $\theta \le 2\sin^{-1}(4\rho\delta) \le 16\rho\delta$ which is at most $2/(5\sqrt{\rho})$ as $\delta \le 1/(40\rho^{3/2})$.

In Figure 7, the length of the dashed arc above cocone(s, θ) and below ℓ is at least $\|x - y\| \cdot \gamma - 15R\theta/16 \ge 7R\gamma/8 - 3R/(8\sqrt{\rho}) > 4R/\sqrt{\rho}$. This implies that we can find a point $r \in F_\alpha \cap$ coarse(s) such that the disk D_r centered at r with radius $\min\{4R/\sqrt{\rho}, R/32\}$ lies inside coarse(s) and below ℓ.

If $\delta \ge 1/\sqrt{\ln n}$, then Lemma 12 implies that radius$(D_r) = \Omega(\lambda_1 \delta \sqrt{f(\tilde{s})})$. By Lemma 5 and the Lipschitz condition, D_r contains a cell C in some $(\lambda_k/\sqrt{f_{\max}})$-decomposition. But then C must be empty as C lies outside strip(s). This occurs with probability at most $n^{-\Omega(\sqrt{\ln n}/f_{\max})}$.

If $\delta < 1/\sqrt{\ln n}$, then Lemma 12 implies that radius$(D_r) \ge \max\{\Omega(\lambda_1^2 f(\tilde{s})), 4\sqrt{\rho}\delta/\sqrt{\rho}\} \ge \Omega(\lambda_1^2 f(\tilde{s})) + 2\delta$. By Lemma 5 and the Lipschitz condition, D_r contains a cell C in some λ_k^2-decomposition. But then C must be empty as C lies outside strip(s). This occurs with probability at most $n^{-\Omega(\ln n)}$. ◻

Lemma 14 *For any two points p and s on F_α, if $\|p - s\| \le \lambda f(\tilde{s})$ for some $\lambda > 0$, then $\|\tilde{p} - \tilde{s}\| \le \frac{2}{2 - 3\delta - \lambda} \cdot \|p - s\|$.*

Proof. Triangle inequality implies that $\|\tilde{p} - \tilde{s}\| \le \|s - \tilde{s}\| + \|p - s\| \le (\alpha + \lambda) f(\tilde{s})$. Assume that $\|\tilde{p} - \tilde{s}\| = \mu f(\tilde{s})$ for some $\mu \le \alpha + \lambda$. Assume that the normal at $\tilde{s}$ is vertical and $\tilde{p}$ lies below $\tilde{s}$. Let x be the center of the disk that touches F at $\tilde{s}$, lies below $\tilde{s}$, and has radius $f(\tilde{s})$.

Then $\sin(\angle \tilde{p} x \tilde{s}/2) \le \mu/2$. So the horizontal distance between $\tilde{p}$ and $\tilde{s}$ is at least $\|\tilde{p} - \tilde{s}\| \cos(\angle \tilde{p} x \tilde{s}/2) \ge (1 - \mu/2) \cdot \|\tilde{p} - \tilde{s}\|$. By Lemma 4, the angle between the normals at $\tilde{p}$ and $\tilde{s}$ is at most $2\sin^{-1}(\mu/2)$. So the horizontal distance between p and s is at least $(1 - \mu/2) \cdot \|\tilde{p} - \tilde{s}\| - \alpha \sin(2\sin^{-1}(\mu/2)) \ge (1 - \mu/2) \cdot \|\tilde{p} - \tilde{s}\| - \alpha\mu = \|\tilde{p} - \tilde{s}\| \cdot (1 - \mu/2 - \alpha/f(\tilde{s}))$. Since $\mu \le \alpha + \lambda$ and $f(\tilde{s}) \ge 1$, we get $\|p - s\| \ge \|\tilde{p} - \tilde{s}\| \cdot (1 - 3\alpha/2 - \lambda/2)$. ◻

To guarantee that we will rotate and align refined(s) approximately well with the normal at $\tilde{s}$, we must show that any other orientation of refined(s) will yield a significant error, i.e., a significantly smaller angle between refined(s) and line(s). This depends on the directions of the normals at the intersections between refined(s) and F_δ as we rotate refined(s). The following lemma shows that if a line ℓ through s makes an angle θ with the normal at $\tilde{s}$, then for any point $q \in F_\alpha$ that ℓ passes through, the angle between ℓ and $\tilde{q}$ cannot be much smaller or larger.

Lemma 15 *Assume that $\delta \le 1/80$. Let ℓ be a line through a sample s. Let $\theta \le \pi/4$ be the angle that ℓ makes with the normal at $\tilde{s}$. Suppose that ℓ intersects F_α at a point q. Then the angle between ℓ and the normal at $\tilde{q}$ is at least $\theta/2$ and at most $3\theta/2$.*

Proof. Assume that $s \in F_{\alpha_0}^+$, s lies above $\tilde{s}$, and the tangent at $\tilde{s}$ is horizontal. Let r be the point on F_α such that $\tilde{r} = \tilde{s}$. Without loss of generality, assume that r lies below $\tilde{s}$ since this maximizes $\|q - r\|$ which in turn maximizes the angle between the normals at $\tilde{q}$ and $\tilde{s}$. So $F_\alpha = F_\alpha^-$. Let D be the disk that touches F_α^- at r, has center x and radius $(1 - \alpha) f(\tilde{s})$, and lies below r. Let γ denote $\angle qxr$. Without loss of generality, assume that q lies below r and let d denote the vertical distance between q and r. Refer to Figure 8.

We have $(1 - \alpha - d)\tan\gamma = (\alpha_0 + \alpha + d)\tan\theta$. Since $\theta \le \pi/4$, it can be shown that $d \le \alpha_0 + \alpha$. It follows that $\tan\gamma \le$

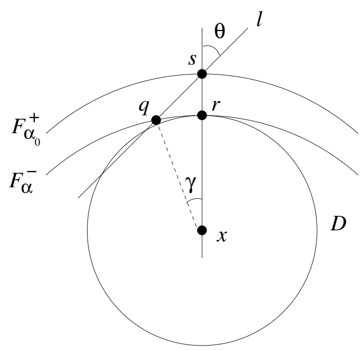

Figure 8: Illustration for Lemma 15.

$\frac{2(\alpha_0+\alpha)}{1-\alpha_0-2\alpha} \cdot \tan\theta \leq \frac{4\delta}{1-3\delta} \cdot \tan\theta$. Note that $\|q - r\| \leq (1 - \alpha) f(\tilde{s}) \tan\gamma < f(\tilde{s})\tan\gamma \leq \frac{4\delta}{1-3\delta} f(\tilde{s}) \tan\theta$. Since $\theta \leq \pi/4$, $\tan\theta \leq 2\theta$. Also, since $\delta \leq 1/80$, $4/(1-3\delta) < 5$. So $\|q-r\| < 10\delta\theta f(\tilde{s})$. By Lemma 14, $\|\tilde{q} - \tilde{r}\| \leq \frac{20\delta\theta}{2-3\delta-10\delta\theta} \cdot f(\tilde{s})$, which is at most $20\delta\theta f(\tilde{s}) \leq \theta f(\tilde{s})/4$. By Lemma 4, the angle between the normals at $\tilde{q}$ and $\tilde{s}$ is at most $2\sin^{-1}(\theta/8) \leq \theta/2$. Hence, the angle between ℓ and the normal at $\tilde{q}$ is at least $\theta/2$ and at most $3\theta/2$. ◻

6. GUARANTEES

In this section, we prove that the reconstruction returned by our algorithm is faithful with high probability. Recall that we rotate refined(s) in clockwise and anti-clockwise directions to identify the normal at $\tilde{s}$. The range of rotation is $[0, \pi/12]$. We first show that the rotation will align refined(s) approximately well with the normal at $\tilde{s}$.

Lemma 16 *Assume that $\delta \leq 1/(40\rho^{3/2})$. Let s be a sample. There exists n_0 such that whenever $n > n_0$, the angle between* refined(s) *and the normal at $\tilde{s}$ is bounded by a function $F(s,n)$ with probability at least $1 - O(n^{-\Omega(\sqrt{\ln n}/f^2_{\max})}(f^2_{\max} + \sqrt{\ln n}))$, where $F(s,n)$ tends to zero as $n \to \infty$.*

Proof. Take a particular orientation of refined(s) within the rotation range $[0, \pi/12]$. For convenience, we rotate the plane such that refined(s) is vertical. Assume that F^-_δ lies below F. It can be verified that each boundary line of refined(s) intersects F^-_δ exactly once inside coarse(s).

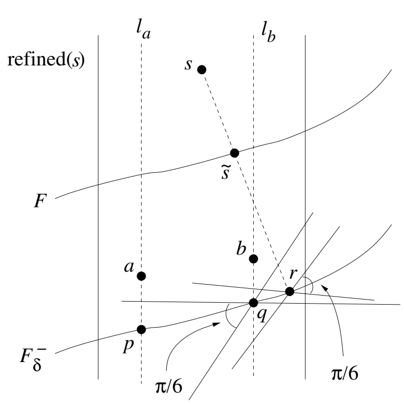

Figure 9: Illustration for Lemma 16.

Refer to Figure 9. Recall that we divide refined(s) into three subslabs of equal width. Let a and b be the lowest samples in the leftmost and rightmost subslabs respectively. Let ℓ_a (resp. ℓ_b) be the vertical line that passes through a (resp. b). Let p (resp. q) be the intersection between ℓ_a (resp. ℓ_b) and F^-_δ inside coarse(s). Assume that p is to the left of q.

Let θ be the angle between the vertical and the normal at $\tilde{s}$. Let γ be the angle between the vertical and the normal at $\tilde{q}$. We first bound θ and γ. By Lemma 13, the angle between the normal at $\tilde{s}$ and the initial refined(s) is at most $\sin^{-1}(5/\sqrt{\rho})$. By a proper choice of ρ, this is bounded by $\pi/12$. Since the range of rotation is $[0, \pi/12]$, we conclude that $\theta \in [0, \pi/6]$. By Lemma 15, $\theta/2 \leq \gamma \leq 3\theta/2 \leq \pi/4$.

Second, we lower bound the vertical distance between p and q. Let t be the intersection between the tangent to F^-_δ at q and ℓ_a. The vertical distance between t and q is at least (width(refined(s))/3) $\cdot \tan\gamma \geq$ (width(refined(s))/3) $\cdot \tan(\theta/2) \geq$ width(refined(s))$\cdot(\theta/6)$. It remains to bound the distance of p from the tangent at q. By Lemma 3, p lies inside cocone$(q, \pi/6)$. So $\|p - q\| \leq$ width(refined(s))$/\cos(\gamma + \pi/6)$. Since $\gamma \leq \pi/4$, we have $\|p-q\| = O($width(refined(s))$)$. By Lemma 2, the distance from p to the tangent to F^-_δ at q is at most $O(\|p - q\|^2) = O($width(refined(s))$^2)$. So $\|p - t\| = O($width(refined(s))$^2/\cos\gamma) = O($width(refined(s))$^2)$ as $\gamma \leq \pi/4$. In all, the vertical distance between p and q is at least width(refined(s)) $\cdot (\theta/6) - O($width(refined(s))$^2)$.

Third, the horizontal distance between a and b is at most width(refined(s)).

Fourth, we claim that $\|q-r\| = O(\delta\theta)$ and $f(\tilde{q}) = O(f(\tilde{s}))$. Let r be the point on F^-_δ such that $\tilde{r} = \tilde{s}$. It can be verified that q lies inside cocone$(r, \pi/6)$. Since $\|r - s\| \leq 2\delta$ and the angle between rs and the vertical is θ, it can be shown using the sine law that $\|q - r\| \leq 2\delta\sin(\theta)/\cos(\theta + \pi/6) \leq 2\delta\theta/\cos(\pi/6) < 3\delta\theta$. So by Lemma 14, $\|\tilde{q} - \tilde{r}\| = O(\|q - r\|) = O(\delta)$. Then the Lipschitz condition implies that $f(\tilde{q}) \leq f(\tilde{r}) + O(\delta) = O(f(\tilde{s}))$.

Fifth, we lower bound the vertical distance between a and b and bound the angle θ. We first consider the case where $\delta \geq 1/\sqrt{\ln n}$. Let C be the cell containing q in some $(\lambda_k/f_{\max})$-decomposition. By Lemma 5, C has diameter $O(\lambda_k\delta)$. Lemma 11 and Lemma 12 imply that for sufficiently large n, width(refined(s)) $= \sqrt{\lambda_1}$ radius(coarse(s)) $= \Omega(\sqrt{\lambda_1}\delta)$. So we can choose a small enough k such that C lies inside the rightmost subslab. Since b is the lowest point in the rightmost subslab, if $\|b-q\|$ is greater than the diameter of C, then C must be empty. By Lemma 9(i), C is empty with probability at most $n^{-\Omega(\sqrt{\ln n}/f^2_{\max})}$. Thus, $\|b - q\| = O(\lambda_k\delta)$. Similarly, one can show that $\|a-p\| = O(\lambda_1\delta)$. The vertical distance between a and b is at least the vertical distance between p and q minus the absolute value of $\|a-p\| - \|b-q\|$. This implies that the vertical distance between a and b is at least width(refined(s))$\cdot(\theta/6) - O($width(refined(s))$^2) - O(\lambda_1\delta)$. Finally, let ψ denote the angle between ab and the horizontal. From the previous discussion, $\tan\psi$ is at least $\theta/6 - O($width(refined(s))$) - O(\lambda_1\delta/$width(refined(s))$)$. The angle between the vertical and the line through ab is at least $\pi/2 - \psi$. This bound is maximum when ψ is minimum, i.e., $\theta = O($width(refined(s)) $+ \lambda_1\delta/$width(refined(s))$)$. Since width(refined(s)) $= \Omega(\sqrt{\lambda_1}\delta)$, $O(\lambda_1\delta/$width(refined(s))$) = O(\sqrt{\lambda_1})$. The upper bound on $O($width(refined(s))$)$ follows from Lemma 11.

The analysis of the fifth step for the case where $\delta <$

$1/\sqrt{\ln n}$ is similar. We consider the cell C containing q in some λ_k^2-decomposition. The difference is that we need to use $\Omega(\lambda_1^2 f(\tilde{s}))$ instead of $\Omega(\sqrt{\lambda_1}\delta)$ as the lower bound for width(refined(s)), since δ can be very small. If we straightforwardly bound $\|b-q\|$ and $\|a-p\|$ by the diameter of C which is $O(\lambda_k^2 f(\tilde{s}))$. Then we get $\|b-q\|/\text{width}(\text{refined}(s)) = O(1)$, which does not tend to zero. Instead, observe that the angle between the normals at q and r is $O(\|q-r\|) = O(\delta\theta)$. Thus, the normal at q makes an angle $\theta + O(\delta\theta)$ with the vertical. Since C is roughly perpendicular to the normal at q, the upper bound for $\|b-q\|$ can be improved to $O(\lambda_k^2 f(\tilde{s})(1+\delta)\theta)$. We choose k small enough such that $\|b-q\|/\text{width}(\text{refined}(s))$ is at most $\theta/20$. One can bound $\|a-p\|$ similarly. The normal at p makes an angle $\theta + O(\delta\theta) + O(\text{width}(\text{refined}(s))$ with the vertical. So $\|a-p\| = O(\lambda_k^2 f(\tilde{s})\theta + \lambda_k^2 f(\tilde{s}) \cdot \text{width}(\text{refined}(s)))$. Like before, we can choose k such that $\|a-p\|/\text{width}(\text{refined}(s))$ is $\theta/20 + O(\lambda_k^2 f(\tilde{s}))$. In all, we can conclude as before that $\tan\psi$ is at least $\theta/6 - O(\text{width}(\text{refined}(s))) - \theta/10 - O(\lambda_k^2 f(\tilde{s}))$. This yields $\theta = O(\text{width}(\text{refined}(s)) + \lambda_k^2 f(\tilde{s}))$, which tends to zero as $n \to \infty$. ◻

Let R_s be the minimum rectangle that aligns with the final refined(s) and contains all samples inside refined(s). Our algorithm works with the center point of R_s. We show that the center point of R_s indeed converges to $\tilde{s}$ as n increases.

Lemma 17 *Assume that $\delta \le 1/(40\rho^{3/2})$. Let s be a sample. There exists n_0 such that whenever $n > n_0$, the distance between the center point of R_s and $\tilde{s}$ is bounded by a function $G(s,n)$ with probability at least $1 - O(n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)}(f_{\max}^2 + \sqrt{\ln n}))$, where $G(s,n)/\sqrt{\text{width}(\text{refined}(s))}$ tends to zero as $n \to \infty$.*

Proof. Assume that s lies on F_α^+ and the normal at $\tilde{s}$ is vertical. Let θ be the angle between the final refined(s) and the vertical. Let r_d (resp. r_u) be the ray that shoots downward (resp. upward) from s and makes an angle θ with the vertical. Let x and y be the points on F_δ^+ and F hit by r_u and r_d respectively. Let z be the point on F_δ^- hit by r_d. Refer to Figure 10.

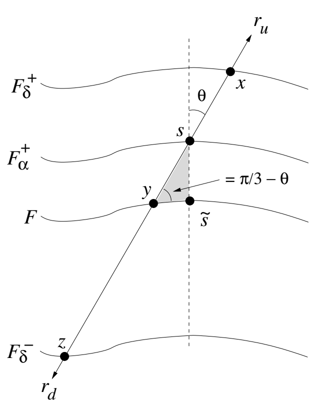

Figure 10: For the proof of Lemma 17.

First, we calculate the location of the midpoint of xz. By Lemma 3, y lies inside cocone($\tilde{s}, \pi/6$). By applying sine law to the shaded triangle in Figure 10, $\|\tilde{s} - y\| \le \frac{\alpha \sin\theta}{\sin(5\pi/12 - \theta)}$. By Lemma 16, $5\pi/12 - \theta > \pi/3$ with probability at least $1 - O(n^{-\Omega(\sqrt{\ln n})})$ whenever $n \ge n_0$. So $\|\tilde{s} - y\| < 2\alpha\sin\theta \le 2\alpha\theta$ with probability at least $1 - O(f_{\max}^2 n^{-\Omega(\sqrt{\ln n}/f_{\max})})$ for sufficiently large n. By triangle inequality, $\|s-\tilde{s}\| - \|\tilde{s}-y\| \le \|s-y\| \le \|s-\tilde{s}\| + \|\tilde{s}-y\|$ which yields $\alpha - 2\alpha\theta \le \|s-y\| \le \alpha + 2\alpha\theta$. We can use a similar argument to show that $(\delta-\alpha) - 2(\delta-\alpha)\theta \le \|s-x\| \le (\delta-\alpha) + 2(\delta-\alpha)\theta$. Combining them yields $\delta - 2\delta\theta \le \|x-y\| \le \delta + 2\delta\theta$. Similarly, we can show that $\delta - 2\delta\theta \le \|y-z\| \le \delta + 2\delta\theta$. Combining the last two inequalities yields $2\delta - 4\delta\theta \le \|x-z\| \le 2\delta + 4\delta\theta$. So the distance of the midpoint of xz from x is at most $\delta + 2\delta\theta$. Thus, the distance of the midpoint of xz from y is at most $\delta + 2\delta\theta - \|x-y\| \le 4\delta\theta$.

Second, we calculate the location of the center point of R_s. Although the center point of R_s lies on the line containing xz, it may not coincide with the midpoint of xz. The top side of R_s may lie above or below x. Within refined(s), Lemma 3 implies that F_δ^+ lies inside cocone($x, \pi/6$) around x. Therefore, if the top side of $R(s)$ is above x, its distance from x is $O(\text{width}(\text{refined}(s)))$. Suppose that the top side of $R(s)$ is below x. We consider the case where $\delta \ge 1/\sqrt{\ln n}$. The analysis for the case where $\delta < 1/\sqrt{\ln n}$ is similar. As in the proof of Lemma 16, one can show that $f(\tilde{x}) = O(f(\tilde{s}))$. Let C be the cell containing x in some $(\lambda_k/f_{\max})$-decomposition. By Lemma 5, C has diameter $O(\lambda_k\delta)$. Lemma 11 and Lemma 12 imply that for sufficiently large n, $\text{width}(\text{refined}(s)) = \sqrt{\lambda_1}\,\text{radius}(\text{coarse}(s)) = \Theta(\sqrt{\lambda_1}\delta)$. So we can choose a small enough k such that C lies inside refined(s). If the top side of R_s lies below C, then C must be empty. By Lemma 9(i), C is empty with probability at most $n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)}$. Thus, the actual position of the top side of R_s may move the center point of R_s away from the midpoint of xz. However, the displacement is $O(\text{width}(\text{refined}(s)) + \lambda_k\delta)$. The actual position of the bottom side of R_s has the same effect.

In all, the distance between the center point of R_s and y is bounded by $4\delta\theta + O(\text{width}(\text{refined}(s)) + \lambda_k\delta)$. Since $\|\tilde{s}-y\| \le 2\alpha\theta$, we conclude that the distance between the center point of R_s and $\tilde{s}$ is bounded by $6\delta\theta + O(\text{width}(\text{refined}(s)) + \lambda_k\delta)$. The proof of Lemma 16 reveals that θ is bounded by $O(\text{width}(\text{refined}(s)) + \sqrt{\lambda_1})$. Since $\text{width}(\text{refined}(s)) = \Theta(\sqrt{\lambda_1}\delta)$, $\delta\theta/\sqrt{\text{width}(\text{refined}(s))} = O(\delta\sqrt{\text{width}(\text{refined}(s))} + \lambda_1^{1/4}\sqrt{\delta})$, which tends to zero as $n \to \infty$. Similarly, the division of $O(\text{width}(\text{refined}(s)) + \lambda_k\delta)$ by $\sqrt{\text{width}(\text{refined}(s))}$ tends to zero as $n \to \infty$. ◻

Theorem 1 *Assume that $\delta \le 1/(40\rho^{3/2})$. There exists n_0 such that given $n > n_0$ noisy samples from a smooth curve F, our algorithm computes a reconstruction that is faithful with probability at least $1 - O(n^{-\Omega(\sqrt{\ln n}/f_{\max}^2)}(f_{\max}^2 + \sqrt{\ln n}))$.*

Proof. By Lemma 17, the center points converge to F with probability at least $1 - O(f_{\max}^2 n^{-\Omega(\sqrt{\ln n}/f_{\max})})$. After the pruning step, for each surviving center point s^*, the distance between s^* and the nearest surviving center points is $\Theta(\sqrt{\text{width}(\text{refined}(s))})$. NN-crust will connect s^* to a neighboring center point r^*. The Lipschitz condition implies that $f(\tilde{r}) = \Theta(f(\tilde{s}))$. Examining the proof of Lemma 17 reveals that $G(r,n) = \Theta(G(s,n)$ when $f(\tilde{r}) = \Theta(f(\tilde{s}))$. So the distance of r^* from $\tilde{r}$ is $O(G(s,n))$. Thus, r^*s^* makes an angle $O(G(s,n)/\sqrt{\text{width}(\text{refined}(s))})$ with the tangent to F at $\tilde{s}$. Moreover, within a distance of $O(\sqrt{\text{width}(\text{refined}(s))})$

F may turn an angle $O(\sqrt{\text{width}(\text{refined}(s))})$. In all, the normal of r^*s^* makes an angle of $O(\sqrt{\text{width}(\text{refined}(s))} + G(s,n)/\sqrt{\text{width}(\text{refined}(s))})$ with the normal at any point $q \in F$ between $\tilde{r}$ and $\tilde{s}$. Since both $\sqrt{\text{width}(\text{refined}(s))}$ and $G(s,n)/\sqrt{\text{width}(\text{refined}(s))})$ tend to zero as $n \to \infty$, the normalwise convergence follows. Pointwise convergence and the pruning step guarantee that s^* is connected to r^* only when $\tilde{r}$ is adjacent to $\tilde{s}$ on F. Thus, the reconstructed curve is homeomorphic to F. ⊒

A straightforward implementation gives a running time of $O(n^2 \log n)$. We believe that the running time can be improved.

7. DISCUSSION

We expect that the approach will also work for handling curves with features: the sampled "curve" consists of a collection of simple curve segments that may only share endpoints, thus forming features like corners, branchings and terminals. Some previous works have already considered terminal and corner points. Allowing branchings extends this to the most general problem. Furthermore, we aim to handle features in the presence of noise. A motivation for allowing branchings is that if we consider surfaces in 3-d with features like sharp edges and corners, then these form a curve graph (in 3-d) with corners, branchings and terminals. The output reconstruction is expected to identify the features as part of the reconstruction. As in previous works, the definition of local feature size is modified to avoid a zero local feature size in corners and branchings points, by pruning the medial axis near the features. The shape fitting can be done by finding a branching of k slabs – the Minkowski sum of a disk and k rays originating from a common apex (see figure) – with smallest width that contains the points. Almost brute force algorithms for these fitting problems run in polynomial time. Linear time approximation algorithms seem possible by adapting recent work on k-line centers [1].

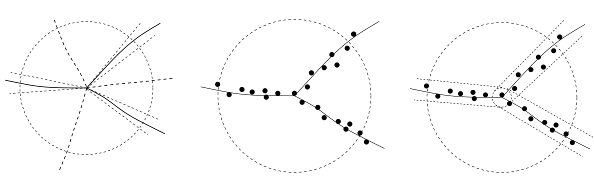

Figure 11: Degree 3 branching, Noisy sampling and Fitting.

We also need a Modified NN-Crust that works correctly for a noise free locally uniform sampling from a curve with features. Such a variant is possible if we assume that for each feature in the curve, the sampling should include a sample s which is identified and provided with a k cones corresponding to the incident curve branches. This is the case for us, since this is information is obtained from the feature fitting step. In the Modified NN-Crust, each feature sample s selects the nearest neighbor in each of its cones, then each non-feature sample s that was not selected by a feature sample proceeds as in the NN-Crust, and each non-feature s that was selected by a feature sample s' selects the nearest neighbor in a cone opposite to s'

To guarantee that the original curve is reconstructed, a very restricted (locally uniform) sampling condition is needed: as it has been pointed out before, the sampling can "simulate" non-existing features and "destroy" real ones. So, a witness guarantee as in [5] is desirable. Beyond this, we also use uniformity of the sampling to assure that the type of the neighborhood can be determined locally. To avoid this, the steps of neighborhood identification and global reconstruction should be interconnected. For example, though at a small scale, a neighborhood may seem to contain a terminal, it may be that this is not the case and that this is only realized when a global consistent reconstruction is not possible under this assumption. Appropriate rules for the interaction between feature fitting and reconstruction need to be explored.

An integration of fitting and reconstruction is also necessary to avoid our current assumption of dense noise. In a different direction, it seems possible to handle outliers if the algorithm uses shape fitting with outliers.

8. REFERENCES

[1] P. K. Agarwal, C. M. Procopiuc and K. Varadarajan. Approximation algorithms for k-line center. *Proc. 10th Annu. European Sympos. Alg.,* 2002.

[2] E. Althaus and K. Mehlhorn. Traveling salesman based curve reconstruction in polynomial time. *SIAM J. Comput.*, 31 (2001), 27–66.

[3] N. Amenta, M. Bern and D. Eppstein. The crust and the beta-skeleton: combinatorial curve reconstruction, *Graphical Models and Image Processing*, 60 (1998), 125–135.

[4] T. K. Dey and P. Kumar. A simple provable algorithm for curve reconstruction. *Proc. 10th. Annu. ACM-SIAM Sympos. Discrete Alg.*, 1999, 893–894.

[5] T. K. Dey, K. Mehlhorn and E. Ramos. Curve reconstruction: connecting dots with good reason. *Comput. Geom. Theory & Appl.*, 15 (2000), 229–244.

[6] T. K. Dey and R. Wenger. Reconstructing curves with sharp corners. *Comput. Geom. Theory & Appl.*, 19 (2001), 89–99.

[7] T. K. Dey and R. Wenger. Fast reconstruction of curves with sharp corners. *Comput. Geom. Theory & Appl.*, 12 (2002), 353–400.

[8] M. Habib, C. McDiarmid, J.Ramirez-Alfonsin and B.Reed. Probabilistic Methods for Algorithmic Discrete Mathematics. Springer Verlag Belin Heidelberg, 1998, 198-200.

[9] S. Funke and E. A. Ramos. Reconstructing a collection of curves with corners and endpoints. *Proc. 12th Annu. ACM-SIAM Sympos. Discrete Alg.*, 2001, 344–353.

[10] J. Giesen. Curve reconstruction, the Traveling Salesman Problem and Menger's Theorem on length. *Discrete & Comp. Geom.*, 24 (2000), 577–603.

[11] C. Gold and J. Snoeyink. A one-step crust and skeleton extraction algorithm. *Algorithmica*, 30 (2001), 144–163.