Curve Reconstruction from Noisy Samples

Siu-Wing Cheng* Stefan Funke†‡ Mordecai Golin*
Piyush Kumar§ Sheung-Hung Poon* Edgar Ramos¶ ‡

August 20, 2003

Abstract

We present an algorithm to reconstruct a collection of disjoint smooth closed curves from 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 in the normal directions. Our reconstruction is faithful with probability approaching 1 as the sampling density increases.

1 Introduction

The combinatorial curve reconstruction problem has been extensively studied recently by computational geometers. The input consists of sample points on a collection of unknown disjoint smooth closed curves denoted by F. The problem calls for computing a set of polygonal curves that are provably *faithful*. That is, as the sampling density increases, the polygonal curves should converge to F.

Amenta et al. [2] obtained the first results in this problem. They proposed a *2D crust* algorithm whose output is provably faithful if the input satisfies the *ϵ-sampling* condition for any $\epsilon < 0.252$. For each point x on F, the *local feature size* $f(x)$ at x is defined as the distance from x to the medial axis of F. For $0 < \epsilon < 1$, a set S of samples is an ϵ-sampling of F if for any point $x \in F$, there exists a sample $s \in S$ such that $\|s - x\| \leq \epsilon \cdot f(x)$ [2]. The algorithm by Amenta et al. invokes the computation of a Voronoi diagram or Delaunay triangulation twice. Gold and Snoeyink [10] presented a simpler algorithm that invokes the computation of Voronoi diagram or Delaunay triangulation only once. Later, Dey and Kumar [3] proposed a *NN-crust*

*Department of Computer Science, HKUST, Hong Kong. Email:{scheng,golin,hung}@cs.ust.hk. Research of S.-W. Cheng and S.-H. Poon are partly supported by Research Grant Council, Hong Kong, China (project no. HKUST 6190/02E and HKUST 6169/03E). Research of M. Golin is partly supported by Research Grant Council, Hong Kong, China (project no. HKUST 6082/01E and HKUST 6206/02E).

†MPI, Saarbrücken, Germany. Email: funke@mpi-sb.mpg.de

‡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).

§Department of Computer Science, SUNY, USA. Email:piyush@ams.sunysb.edu. Research supported by National Science Foundation grant CCR-0098172.

¶Department of Computer Science, UIUC, USA. Email:eramosn@cs.uiuc.edu

algorithm for this problem. 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 on only one edge e, connect s to the closest sample among all samples u such that su makes an obtuse angle with e. The output curve is faithful for any $\epsilon \leq 1/3$ [3].

Dey, Mehlhorn, and Ramos [4] proposed a *conservative-crust* algorithm to handle curves with endpoints. Funke and Ramos [8] proposed an algorithm to handle curves that may have sharp corners and endpoints. Dey and Wenger [5, 6] also described algorithms and implementation for handling sharp corners. Giesen [9] discovered that the traveling salesperson tour through the samples is a faithful reconstruction, but this approach cannot handle more than one curve. Althaus and Mehlhorn [1] 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. For simplicity and notational convenience, we assume throughout this paper that $\min_{x \in F} f(x) = 1$ and F consists of a single smooth closed curve, although our algorithm works when F contains more than one curve.

In our model, a sample is generated by drawing a point from F followed by randomly perturbing the point in the normal direction. Let $L = \int_F \frac{1}{f(x)} dx$. The drawing of points from F follows the probability density function $\frac{1}{L \cdot f(x)}$. That is, the probability of drawing a point from a curve segment η is equal to $\int_\eta \frac{1}{f(x)} dx$ divided by L. 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. The distribution of each sample is independently identical. δ is the noise amplitude and we assume that $\delta \leq 1/(25\rho^2)$ where $\rho \geq 5$ is a constant chosen a priori by our algorithm. We assume throughout this paper that $\delta > 0$. We emphasize that the value of 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.

We prove that our algorithm returns a reconstruction which is faithful with probability at least $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$, where n is the number of input samples, ω is an arbitrary positive constant, and $f_{\max} = \max_{x \in F} f(x)$. 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 recognizing non-smooth features. We also expect that this clustering approach can be generalized to 3D for surface reconstruction problems.

The rest of the paper is organized as follows. Section 2 describes our algorithm. Section 3 introduces two decompositions of the space around F which is the main tool in our probabilistic argument. Sections 4 and 5 prove that our reconstruction is faithful with probability approaching 1. We conclude in Section 6.

2 Algorithm

We first highlight the key ideas. Our algorithm works by growing a disk neighborhood around each sample p until the samples inside the disk fit in a strip whose width is small relative to the radius of the disk. The final disk is the *coarse neighborhood* of p denoted by $coarse(p)$. $coarse(p)$ provides a first estimate of the curve locally and of its normal. A better estimation is possible. We shrink $coarse(p)$ by a certain factor. We take a slab bounded by two parallel tangent lines of the shrunken $coarse(p)$. We rotate the slab around p to minimize the spread of the samples inside along the direction of the slab. The final orientation of the slab provides a good normal estimation and it also allows us to estimate a *center point* close to F in place of p. Next, we decimate the center points as follows. We scan the center points in decreasing order

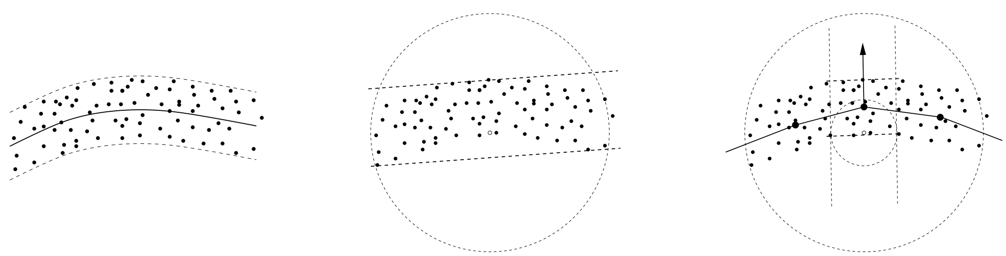


Figure 1: On the left, a smooth curve segment with a noise cloud. In the middle, a sufficiently large neighborhood identifies a strip with relatively large aspect ratio, which can provide preliminary point and normal estimates. On the right, concentrating on smaller neighborhoods, a better estimate of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

of the widths of their corresponding slabs. When we add the current center point p^* to the decimated set, we delete the other center points that are too close to p^*. Finally, we can run any reconstruction algorithm that is correct for a noise free sampling on the remaining center points. For example, the NN-Crust algorithm by Dey and Kumar [3].

We provide the details of the algorithm in the following. Let n be the total number of input samples. Let $\omega > 0$ and $\rho \geq 5$ be two predefined constants.

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

Coarse neighborhood: Let D be the disk that is centered at s and contains $\ln^{1+\omega} n$ samples. Let $initial(s)$ be the disk centered at s with radius $\sqrt{\text{radius}(D)}$. 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 $\frac{\text{radius}(coarse(s))}{\text{width}(strip(s))} \geq \rho$. The final disk $coarse(s)$ is the *coarse neighborhood* of s.

REFINED NEIGHBORHOOD: Let N_s be the upward direction perpendicular to $strip(s)$. The candidate neighborhood $candidate(s, \theta)$ is the slab that contains s in the middle, makes a signed acute angle θ with N_s, and has width $\min\{\sqrt{\text{radius}(initial(s))}, \text{radius}(coarse(s))/3\}$. The angle θ is positive (resp. negative) if it is on right (resp. left) of N_s. We enclose the samples in $candidate(s, \theta)$ by two parallel lines that are orthogonal to the direction of $candidate(s, \theta)$. These two lines form a rectangle $rectangle(s, \theta)$ with the boundary lines of $candidate(s, \theta)$. The width of $rectangle(s, \theta)$ is the width of $candidate(s, \theta)$. The height of $rectangle(s, \theta)$ is its length along the direction of $candidate(s, \theta)$. We vary θ within the range $[-\pi/1, \pi/10]$ to find an orientation that minimizes the height of $rectangle(s, \theta)$. Let θ^* be the minimizing angle. The *refined neighborhood* of s is $rectangle(s, \theta^*)$ and is denoted by $refined(s)$. We return the center point s^* of $refined(s)$.

PRUNING: We sort the center points s^* in decreasing order of $\text{width}(refined(s))$. Then we scan the sorted list and select a subset of center points: when we select the current center point s^*, we delete all center points u^* from the sorted list such that $\|s^* - u^*\| \leq \text{width}(refined(s))^{1/3}$.

OUTPUT: We run the NN-crust algorithm on the selected center points and return the output curve.

A few remarks are in order. Recall that $\min_{x \in F} f(x)$ is assume to be 1. For sufficiently large n (i.e., when the sampling is dense enough), the radius of D in the the step COARSE NEIGHBORHOOD is less than 1. So $1 > \sqrt{\text{radius}(D)} > \text{radius}(D)$, implying that $initial(s)$ contains D. Similarly, in the step REFINED NEIGHBORHOOD, $\sqrt{\text{radius}(initial(s))} > \text{radius}(initial(s))$. Clearly, $coarse(s)$ contains $initial(s)$. So the width of $candidate(s, \theta)$ and $refined(s)$ is at least $\text{radius}(initial(s))/3$ and at most $\sqrt{\text{radius}(initial(s))} < 1$. Therefore, in the step PRUNING, $\text{width}(refined(s))^{1/3} > \text{width}(refined(s))$. Thus, when we delete center points within a distance of $\text{width}(refined(s))^{1/3}$, we decimate center points outside $refined(s)$ too.

3 Decompositions

In this section, we introduce and analyze two decompositions of the space around F. They will be essential for the probabilistic 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. That is, $\tilde{x}$ is the projection of x onto F. (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 the points 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 and y on F, we use $F(x, y)$ to denote the curved segment traversed from x to y in clockwise direction. We use $|F(x, y)|$ to denote the length of $F(x, y)$.

We will use two types of decompositions, *β-partition* and *β-grid*. Let $0 < \beta < 1$ be a parameter. 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 $F(c_{i-1}, c_0)$, $|F(c_{i-1}, c_i)| = \beta^2 f(c_{i-1})$, and $|F(c_i, c_0)| \geq \beta^2 f(c_i)$. If c_i exists, it is the next cut-point and we continue. Otherwise, we have computed all the cut-points and we stop. The β-partition is the arrangement of F_δ^+, F_δ^-, and the normal segments at the cut-points. Figure 2 shows an example. We call each face of the β-partition a *β-slab*. The β-partition consists of a row of slabs stabbed by F.

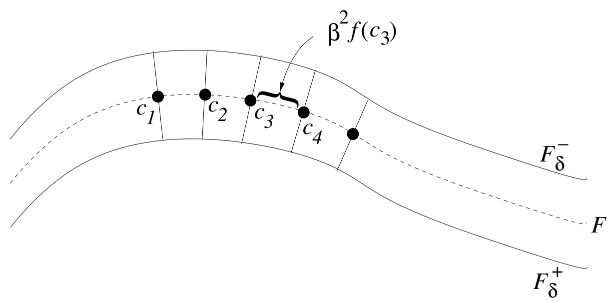


Figure 2: β-partition.

The cut-points for a β-grid are picked differently. 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 $F(c_{i-1}, c_0)$, $|F(c_{i-1}, c_i)| = \beta f(c_{i-1})$, and $|F(c_i, c_0)| \geq \beta f(c_i)$. If c_i exists, it is the next cut-point and we continue. Otherwise, we have computed all the cut-points and we stop. The β-grid is the arrangement of the following:

- The normal segments at the cut-points.
- F, F_δ^+, and F_δ^-.
- F_α^+ and F_α^- where $\alpha = i\beta\delta$ and i is an integer between 1 and $\lfloor 1/\beta \rfloor - 1$.

The β-grid has a grid structure. Figure 3 shows an example. We call each face of the β-grid a *β-cell*. There are $O(1/\beta)$ rows of cells "parallel to" F.

In Section 3.1, we prove several properties of F_α for any α. These properties will be used in Section 3.2 to bound the diameter of a β-cell. These properties will also be useful later in the paper. In Section 3.3, we analyze the probabilities of a β-slab and a β-cell containing certain numbers of samples. These probabilities are essential for the probabilistic analysis later.

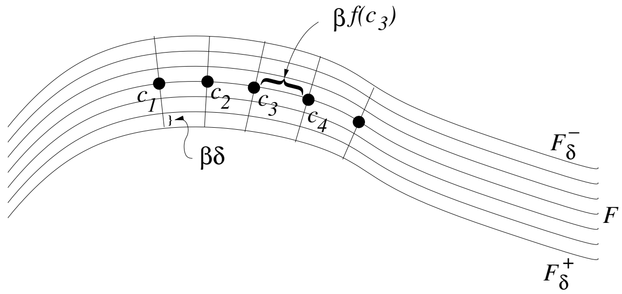


Figure 3: β-grid.

3.1 Properties of F_α

Lemma 3.1 *Any point p on F_α has two tangent disks with radii $f(\tilde{p}) - \alpha$ whose interior do not intersect F_α.*

Proof. Let M_α be the medial disk of F_α touching a point $p \in F_\alpha$. By the definition of F_α, there is a medial disk M of F touching $\tilde{p}$ such that M and M_α have the same center. Moreover, $\text{radius}(M_\alpha) = \text{radius}(M) - \alpha \geq f(\tilde{p}) - \alpha$. ◻

For each point p on F_α, define *cocone*(p, θ) as the double cone that has apex p and angle θ such that the normal of F_α at p is the symmetry axis of the double cone that lies outside it. The next lemma shows that F_α lies inside *cocone*(p, θ) for a small θ in a small neighborhood of p.

Lemma 3.2 *Let p be a point on F_α. Let D be a disk centered at p with radius at most $2(1 - \alpha)f(\tilde{p})$.*

(i) For any point $q \in F_\alpha \cap D$, the distance of q from the tangent at p is at most $\frac{\|p-q\|^2}{2(1-\alpha)f(\tilde{p})}$.

(ii) $F_\alpha \cap D \subseteq$ cocone$(p, 2\sin^{-1}\frac{\text{radius}(D)}{2(1-\alpha)f(\tilde{p})})$.

Proof. Assume that the tangent at p is horizontal. Consider (i). Refer to Figure 4(a). Let B 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 B. Let r be a point in $C \cap \partial B$. Let d be the distance of r from the tangent at p. By Lemma 3.1, d bounds the distance from q to the tangent at p. Observe that $\|p - q\| = \|p - r\| = 2(1 - \alpha)f(\tilde{p})\sin(\frac{\angle pxr}{2})$ and $d = \|p - r\| \cdot \sin(\frac{\angle pxr}{2})$. Thus, $d = 2(1 - \alpha)f(\tilde{p})\sin^2(\frac{\angle pxr}{2}) = \frac{\|p-q\|^2}{2(1-\alpha)f(\tilde{p})}$.

Consider (ii). Refer to Figure 4(b). By (i), the distance between $F_\alpha \cap D$ and the tangent at p is bounded by $\frac{\text{radius}(D)^2}{2(1-\alpha)f(\tilde{p})}$. Let θ be the smallest angle such that *cocone*(p, θ) contains $F_\alpha \cap D$. Then $\sin\frac{\theta}{2} \leq \frac{\text{radius}(D)^2}{2(1-\alpha)f(\tilde{p})} \cdot \frac{1}{\text{radius}(D)} = \frac{\text{radius}(D)}{2(1-\alpha)f(\tilde{p})}$. ◻

The next lemma shows that the normal deviation is very small in a small neighborhood of any point in F_α.

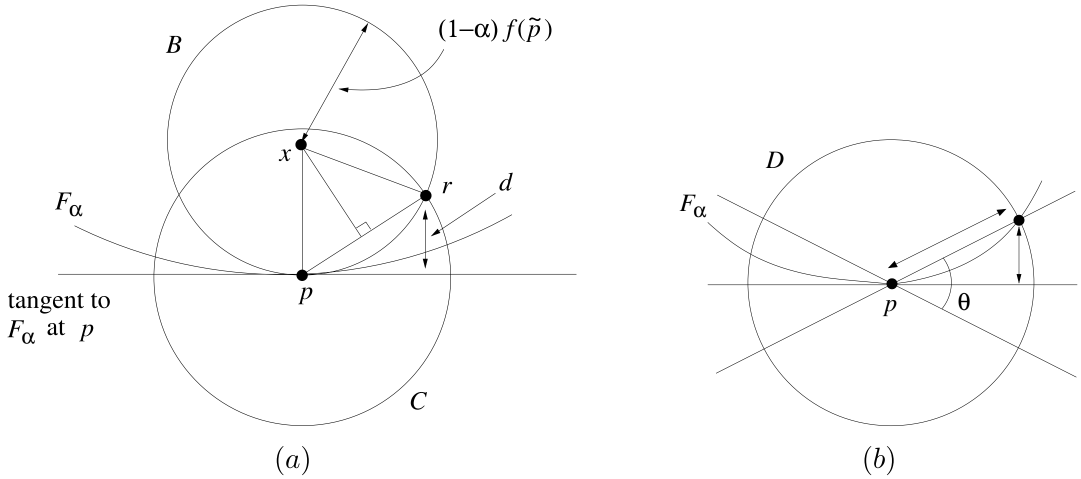


Figure 4: Illustration for Lemma 3.2.

Lemma 3.3 *Let p be a point on F_α. Let D be a disk centered at p with radius at most $\frac{(1-\alpha)f(\tilde{p})}{4}$. For any point $u \in F_\alpha \cap D$, the acute angle between the normals at p and u is at most* $2\sin^{-1}\frac{\|p-u\|}{(1-\alpha)f(\tilde{p})} \leq 2\sin^{-1}\frac{\text{radius}(D)}{(1-\alpha)f(\tilde{p})}$.

Proof. Take any point u on $F_\alpha \cap D$. Let ℓ be the tangent to F_α at u. Let ℓ' be the line that is perpendicular to ℓ and passes through u. Let C be the circle centered at p with radius $\|p-u\|$. Let A and B be the two tangent circles at p with radius $\frac{(1-\alpha)f(\tilde{p})}{2}$. Let x be the center of A. Without loss of generality, we assume that the tangent to F_α at p is horizontal, A is below B, u lies to the left of p, and the slope of ℓ is positive or infinite. (We ignore the case where the slope of ℓ is zero as there is nothing to prove then.) It follows that the slope of ℓ' is zero or negative. Refer to Figure 5.

By Lemma 3.1, u lies outside A and B. Let q be the intersection point between C and A on the left of p. Since $\|p-q\| = \|p-u\| \leq \frac{(1-\alpha)f(\tilde{p})}{4} = \text{radius}(A)/2$, q lies above x. Also, $\angle pxq = 2\sin^{-1}\frac{\|p-u\|}{(1-\alpha)f(\tilde{p})}$.

Suppose that ℓ' does not lie above x, see Figure 5(a). Since u lies above the support line of qx, the angle between ℓ' and the vertical is less than or equal to $\angle pxq = 2\sin^{-1}\frac{\|p-u\|}{(1-\alpha)f(\tilde{p})}$.

Suppose that ℓ' lies above x but not above p, see Figure 5(b). We show that this case is impossible. Let w the intersection point between A and ℓ' on the right of p. Note that p lies between u and w and $\angle upw > \pi/2$. If we grow a disk that lies below l and remains tangent to l at u, the disk will hit F_α at some point different from u when the disk passes through p or earlier. It follows that there is a medial disk M_u of F_α that touches u and lies below l. Observe that the center of M_u lies on the half of ℓ' on the right of u. Furthermore, the center of M_u lies on the line segment uw; otherwise, since $\angle upw > \pi/2$, M_u would contain p, contradiction. Thus, the distance from $\tilde{p}$ to the center of M_u is less than $\max\{\|p-u\|, \|p-w\|\} + \|p-\tilde{p}\| \leq 2 \cdot \text{radius}(A) + \alpha = (1-\alpha)f(\tilde{p}) + \alpha \leq f(\tilde{p})$. However, since the center of M_u is also a point on the medial axis of F, its distance from p should be at least $f(\tilde{p})$, contradiction.

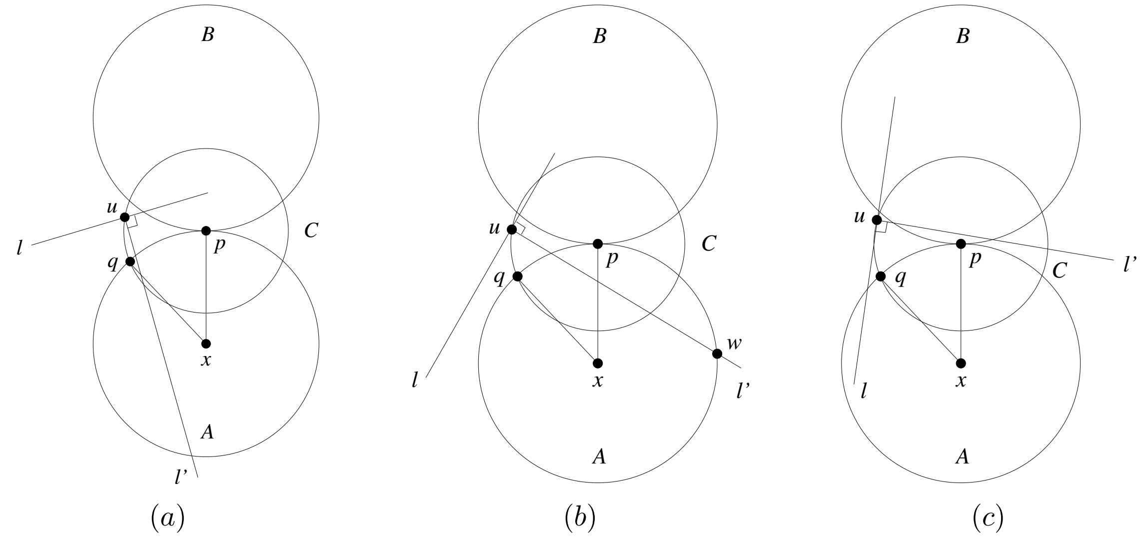


Figure 5: Illustration for Lemma 3.3.

The remaining case is that ℓ' lies above p, see Figure 5(c). Since u lies outside B and the slope of ℓ' is zero or negative, ℓ' lies between p and the center of B. The situation is similar to the previous case where ℓ' lies between p and x. So a similar argument shows that this case is also impossible. ◻

3.2 Diameter of a β-cell

We need a technical lemma before proving an upper bound on the diameter of a β-cell.

Lemma 3.4 *Assume that $\beta \leq 1/4$. Let p and q be two points on F_α such that $|F(\tilde{p},\tilde{q})| \leq 2\beta f(\tilde{p})$. Then $\|p - q\| \leq \|\tilde{p} - \tilde{q}\| + 5\beta\delta$.*

Proof. Refer to Figure 6. Let r be the point $q - \tilde{q} + \tilde{p}$. Without loss of generality, assume that $\angle \tilde{p}pr \leq \angle \tilde{p}rp$. Lemma 3.3 implies that $\angle p\tilde{p}r \leq 2\sin^{-1} 2\beta$. Therefore, $\angle \tilde{p}rp \geq \pi/2 -$

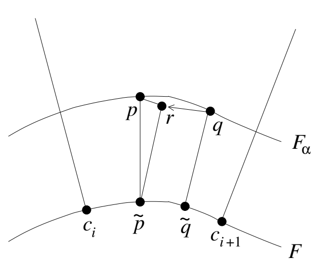


Figure 6: Illustration for Lemma 3.4.

$\sin^{-1} 2\beta$. By sine law, $\|p - r\| = \frac{\|p-\tilde{p}\| \cdot \sin \angle p\tilde{p}r}{\sin \angle \tilde{p}rp} \leq \frac{\delta \sin(2\sin^{-1} 2\beta)}{\cos(\sin^{-1} 2\beta)}$. Note that $\sin(2\sin^{-1} 2\beta) \leq 2\sin(\sin^{-1} 2\beta) = 4\beta$ and since $\beta \leq 1/4$, $\cos(\sin^{-1} 2\beta) \geq \cos(\sin^{-1}(1/2)) > 0.86$. So $\|p - r\| \leq$

$4\beta\delta/(0.86) < 5\beta\delta$. By triangle inequality, we get $\|p-q\| \le \|q-r\|+\|p-r\| = \|\tilde{p}-\tilde{q}\|+\|p-r\| < \|\tilde{p}-\tilde{q}\| + 5\beta\delta$. ◻

Lemma 3.5 *Assume that $\beta \le 1/4$ and $\delta < 1$. Let C be any β-cell that lies between the normal segments at the cut-points c_i and c_{i+1}. Then the diameter of C is at most $11\beta f(c_i)$.*

Proof. Let s and t be two points in C. Let p be the projection of s towards $\tilde{s}$ onto a side of C. Similarly, let q be the projection of t towards $\tilde{t}$ onto the same side of C. Note that $\tilde{p} = \tilde{s}$ and $\tilde{q} = \tilde{t}$. The triangle inequality and Lemma 3.4 imply that

$$\begin{aligned} \|s-t\| &\le \|p-q\| + \|p-s\| + \|q-t\| \\ &\le \|\tilde{p}-\tilde{q}\| + 5\beta\delta + \|p-s\| + \|q-t\|. \end{aligned}$$

Since $\|\tilde{p}-\tilde{q}\| = \|\tilde{s}-\tilde{t}\| \le 2\beta f(c_i)$ and both $\|p-s\|$ and $\|q-t\|$ are at most $2\beta\delta$, the diameter of C is at most $2\beta f(c_i) + 9\beta\delta \le 11\beta f(c_i)$. ◻

3.3 Number of samples in cells and slabs

We first need a lemma that estimates the probability of a sample point lying inside a β-cell and a β-slab.

Lemma 3.6 *Let $\lambda_k = \sqrt{\frac{k^2 \ln^{1+\omega} n}{n}}$ for some positive constant k. Let $r \ge 1$ be a parameter. Let C be a (λ_k/r)-slab or (λ_k/r)-cell. Let s be a sample. There exist constants κ_1 and κ_2 such that if n is so large that $\lambda_k \le 1/4$, then $\kappa_2\lambda_k^2/r^2 \le \Pr(s \in C) \le \kappa_1\lambda_k^2/r^2$.*

Proof. Recall that $L = \int_F \frac{1}{f(x)}dx$. Assume that C lies between the normal segments at the cut-points c_i and c_{i+1}. We use η to denote $F(c_i, c_{i+1})$ as a short hand. By our assumption on λ_k, for any point $x \in \eta$, if C is a λ_k-cell, then $\|x - c_i\| \le 2\lambda_k f(c_i)/r \le f(c_i)/2$; if C is a λ_k-slab, then $\|x - c_i\| \le 2\lambda_k^2 f(c_i)/r^2 \le f(c_i)/8$. The Lipschitz condition implies that $f(c_i)/2 \le f(x) \le 3f(c_i)/2$. If C is a λ_k-slab, then $\Pr(s \in C) = \Pr(\tilde{s} \text{ lies on } \eta)$, which is $\frac{1}{L}\cdot\int_\eta \frac{1}{f(x)}dx \in [\frac{2\lambda_k^2}{3Lr^2}, \frac{4\lambda_k^2}{Lr^2}]$. If C is λ_k-cell, then $\Pr(\tilde{s} \text{ lies on } \eta) = \frac{1}{L}\cdot\int_\eta \frac{1}{f(x)}dx \in [\frac{2\lambda_k}{3Lr}, \frac{4\lambda_k}{Lr}]$. Since $\Pr(s \in C \mid \tilde{s} \text{ lies on } \eta) \in [\frac{\lambda_k\delta}{2\delta r}, \frac{2\lambda_k\delta}{2\delta r}] = [\frac{\lambda_k}{2r}, \frac{\lambda_k}{r}]$, $\Pr(s \in C) \in [\frac{\lambda_k^2}{3Lr^2}, \frac{4\lambda_k^2}{Lr^2}]$. ◻

The following Chernoff bound [7] will be needed.

Lemma 3.7 *Let the random variables $X_1, X_2, \ldots, X_n$ be independent, with $0 \le X_i \le 1$ for each i. Let $S_n = \sum_{i=1}^n X_i$, and let $E(S_n)$ be the expected value of S_n. Then for any $\sigma > 0$, $\Pr(S_n \le (1-\sigma)E(S_n)) \le \exp(-\frac{\sigma^2 E(S_n)}{2})$, and $\Pr(S_n \ge (1+\sigma)E(S_n)) \le \exp(-\frac{\sigma^2 E(S_n)}{2(1+\sigma/3)})$.*

We are ready to analyze the probabilities of a β-slab and a β-cell containing certain numbers of samples.

Lemma 3.8 *Let $\lambda_k = \sqrt{\frac{k^2 \ln^{1+\omega} n}{n}}$ for some positive constant k. Let $r \ge 1$ be a parameter. Let C be a (λ_k/r)-slab or (λ_k/r)-cell. Let κ_1 and κ_2 be the constants in Lemma 3.6. Whenever n is so large that $\lambda_k \le 1/4$, the following hold.*

(i) C *is non-empty with probability at least* $1 - n^{-\Omega(\ln^\omega n/r^2)}$.

(ii) Assume that $r = 1$*. For any constant* $\kappa > \kappa_1 k^2$*, the number of samples in* C *is at most* $\kappa \ln^{1+\omega} n$ *with probability at least* $1 - n^{-\Omega(\ln^\omega n)}$.

(iii) Assume that $r = 1$*. For any constant* $\kappa < \kappa_2 k^2$*, the number of samples in* C *is at least* $\kappa \ln^{1+\omega} n$ *with probability at least* $1 - n^{-\Omega(\ln^\omega n)}$.

Proof. Let $X_i (i = 1, \ldots, n)$ be a random binomial variable taking value 1 if the sample point s_i is inside C, and value 0 otherwise. Let $S_n = \sum_{i=1}^n X_i$. Then $E(S_n) = \sum_{i=1}^n E(X_i) = n \cdot \Pr(s_i \in C)$. This implies that

$$E(S_n) \le \frac{\kappa_1 n \lambda_k^2}{r^2} = \frac{\kappa_1 k^2 \ln^{1+\omega} n}{r^2}, \qquad E(S_n) \ge \frac{\kappa_2 n \lambda_k^2}{r^2} = \frac{\kappa_2 k^2 \ln^{1+\omega} n}{r^2}.$$

By Lemma 3.7,

$$\begin{aligned} \Pr(S_n \le 0) &= \Pr(S_n \le (1-1)E(S_n)) \\ &\le \exp(-\frac{E(S_n)}{2}) \\ &\le \exp(-\Omega(\frac{\ln^{1+\omega} n}{r^2})). \end{aligned}$$

Consider (ii). Let $\sigma = \frac{\kappa}{\kappa_1 k^2} - 1 > 0$. Since $r = 1$, we have

$$\kappa \ln^{1+\omega} n = \kappa_1 n \lambda_k^2 (1+\sigma) \ge (1+\sigma) E(S_n).$$

By Lemma 3.7,

$$\begin{aligned} \Pr(S_n > \kappa \ln^{1+\omega} n) &\le \Pr(S_n > (1+\sigma)E(S_n)) \\ &\le \exp(-\frac{\sigma^2 E(S_n)}{2 + 2\sigma/3}) \\ &= \exp(-\Omega(\ln^{1+\omega} n)). \end{aligned}$$

Consider (iii). Let $\sigma = 1 - \frac{\kappa}{\kappa_2 k^2} > 0$. Since $r = 1$, we have

$$\kappa \ln^{1+\omega} n = \kappa_2 n \lambda_k^2 (1-\sigma) \le (1-\sigma) E(S_n).$$

By Lemma 3.7,

$$\begin{aligned} \Pr(S_n < \kappa \ln^{1+\omega} n) &\le \Pr(S_n < (1-\sigma)E(S_n)) \\ &\le \exp(-\frac{\sigma^2 E(S_n)}{2}) \\ &= \exp(-\Omega(\ln^{1+\omega} n)). \end{aligned}$$

□

4 Coarse neighborhood

In this section, we bound the radii of *initial*(s) and *coarse*(s) for each sample s. Then we show that *strip*(s) provides a rough estimate of the slope of the tangent to F at $\tilde{s}$. Recall that $\lambda_k = \sqrt{\frac{k^2 \ln^{1+\omega} n}{n}}$.

4.1 Radius of *initial*(s)

We first need a technical lemma that bounds the distance between two normal segments from below.

Lemma 4.1 *Assume that $\delta \leq 1/8$ and $\lambda_k \leq 1/4$. Let c_i and c_{i+1} be two consecutive cut-points of a λ_k-partition. For any point on the normal segment at c_{i+1} (resp. c_i), its distance from the support line of the normal segment at c_i (resp. c_{i+1}) is at least $|F(c_i, c_{i+1})|/6$.*

Proof. Assume that the normal at c_i is vertical. Take any two points $p, q \in F_\alpha$ such that $\tilde{p} = c_i$ and $\tilde{q} = c_{i+1}$. We first bound the distance from q to the support line of the normal segment at c_i. The same approach also works for the distance from p to the support line of the normal segment at c_{i+1}.

Let r be the orthogonal projection of q onto the tangent to F_α at p. Observe that the distance of q from the support line of the normal segment at c_i is $\|p - r\|$. We are to prove that $\|p - r\| \geq |F(c_i, c_{i+1})|/6$. For any point $x \in F_\alpha(p, q)$, we use θ_x to denote the non-obtuse angle between the normals at $\tilde{x}$ and c_i. By Lemma 3.3, we have $\theta_x \leq 2 \sin^{-1} \frac{|F(c_i,c_{i+1})|}{f(c_i)}$. By our assumption on λ_k, $\frac{|F(c_i,c_{i+1})|}{f(c_i)} \leq 2\lambda_k^2 \leq 1/8$. It follows that $\sin^{-1} \frac{|F(c_i,c_{i+1})|}{f(c_i)} < \frac{2|F(c_i,c_{i+1})|}{f(c_i)}$. Therefore,

$$\begin{aligned} \theta_x &\leq \frac{4|F(c_i, c_{i+1})|}{f(c_i)} &(1) \\ &\leq 8\lambda_k^2. &(2) \end{aligned}$$

This implies that $F_\alpha(p, q)$ is monotone along the tangent to F_α at p; otherwise, there is a point $x \in F_\alpha(p, q)$ such that $\theta_x = \pi/2 > 8\lambda_k^2$, contradiction. It follows that $F(c_i, c_{i+1})$ is also monotone along the tangent to F at c_i. Refer to Figure 7. Assume that the tangents at p and c_i are horizontal, p lies below c_i, and q lies to the right of p. Let r' be the orthogonal projection of c_{i+1} onto the tangent to F at c_i. The monotonicity of $F(c_i, c_{i+1})$ implies that

$$\|c_i - r'\| = \int_{F(c_i,c_{i+1})} \cos\theta_x \, dx \overset{(2)}{\geq} |F(c_i, c_{i+1})| \cdot \cos(8\lambda_k^2) > 0.8|F(c_i, c_{i+1})|,$$

as $\cos(8\lambda_k^2) \geq \cos(0.5) > 0.8$. Let d be the horizontal distance between r and r'. Observe that $d = \|c_{i+1} - q\| \cdot \sin\theta_q \leq \delta\theta_q$, which is at most $4\delta|F(c_i, c_{i+1})|$ by (1). We conclude that

$$\begin{aligned} \|p - r\| &\geq \|c_i - r'\| - d \\ &\geq (0.8 - 4\delta)|F(c_i, c_{i+1})| \\ &\overset{\delta \leq 1/8}{>} \frac{|F(c_i, c_{i+1})|}{4}. \end{aligned}$$

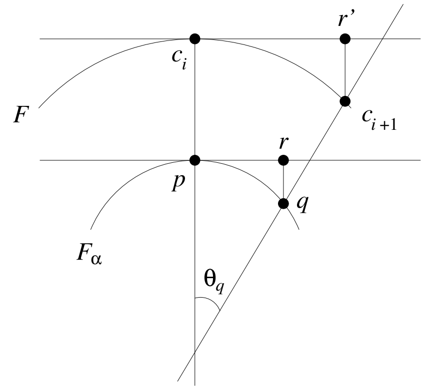


Figure 7: Illustration for Lemma 4.1.

This proves the first part of the lemma.

Let d_p be the distance from p to the support line of the normal segment at c_{i+1}. We can use the same approach to lower bound d_p. The only difference is that for any point $x \in F_\alpha(p,q)$, the non-obtuse angle ϕ_x between the normals at $\tilde{x}$ and c_{i+1} satisfies

$$\phi_x \leq 2\,\sin^{-1}\frac{|F(c_i,c_{i+1})|}{f(c_{i+1})}.$$

Note that the denominator is $f(c_{i+1})$ instead of $f(c_i)$ in (1). Nevertheless, by the Lipschitz condition, $f(c_{i+1}) \geq f(c_i) - \|c_i - c_{i+1}\| \geq f(c_i) - |F(c_i,c_{i+1})| \geq (1-2\lambda_k^2)f(c_i)$, which is at least $7f(c_i)/8$ as $2\lambda_k^2 \leq 1/8$. Therefore,

$$\phi_x \leq 2\,\sin^{-1}\frac{8|F(c_i,c_{i+1})|}{7f(c_i)} \leq \frac{16}{7}\cdot\frac{2|F(c_i,c_{i+1})|}{f(c_i)} < \frac{5|F(c_i,c_{i+1})|}{f(c_i)} < 10\lambda_k^2.$$

Observe that $\phi_x \leq 10\lambda_k^2 < \pi/2$. So $F_\alpha(p,q)$ and $F(c_i,c_{i+1})$ are monotone along the tangents to F_α at q and F at c_{i+1}, respectively. Also, $\cos\phi_x \geq \cos(10\lambda_k^2) \geq \cos(0.625) > 0.8$. Hence, the previous calculation yields

$$\begin{aligned} d_p &\geq (0.8-5\delta)|F(c_i,c_{i+1})| \\ &\overset{\delta\leq 1/8}{>} \frac{|F(c_i,c_{i+1})|}{6}. \end{aligned}$$

□

We are ready to bound the radius of *initial*(s).

Lemma 4.2 *Let h be a constant less than $\sqrt{\frac{1}{3\kappa_1}}$ and let m be a constant greater than $\sqrt{\frac{2}{\kappa_2}}$, where κ_1 and κ_2 are the constants in Lemma 3.6. Let $\psi_h = \lambda_h/3$ and $\psi_m = \sqrt{11\lambda_m}$. Let s be a sample. If $\delta \leq 1/8$, $\lambda_h \leq 1/4$, and $\lambda_m \leq 1/4$, then*

$$\psi_h\sqrt{f(\tilde{s})} \leq \text{radius}(\textit{initial}(s)) \leq \psi_m\sqrt{f(\tilde{s})}$$

with probability at least $1 - O(n^{-\Omega(\ln^\omega n)})$.

Proof. Let D be the disk centered at s that contains $\ln^{1+\omega}$ samples. We first prove the upper bound. Take a λ_m-grid such that s lies on the normal segment at the cut-point c_0. Let C be the λ_m-cell between the normal segments at c_0 and c_1 that contains s. By Lemma 3.8(iii), C contains at least $2\ln^{1+\omega} n$ samples with probability at least $1 - n^{-\Omega(\ln^\omega n)}$. Since D contains $\ln^{1+\omega} n$ samples, radius(D) is less than the diameter of C with probability at least $1-n^{-\Omega(\ln^\omega n)}$. By Lemma 3.5, radius($D$) $\leq 11\lambda_m f(c_0) = 11\lambda_m f(\tilde{s})$. It follows that radius(*initial*(s)) $= \sqrt{\text{radius}(D)} \leq \sqrt{11\lambda_m f(\tilde{s})}$.

Next, we prove the lower bound. Take a λ_h-partition such that s lies on the normal segment at the cut-point c_0. Consider the cut-points c_j for $-1 \leq j \leq 1$. (We use c_{-1} to denote the last cut-point picked.) We have $\|c_{-1} - c_0\| \leq |F(c_{-1}, c_0)| \leq 2\lambda_h^2 f(c_{-1}) < 0.125 f(c_{-1})$ as $\lambda_h \leq 1/4$. The Lipschitz condition implies that

$$f(c_{-1}) \geq f(c_0)/1.125 > 0.8 f(c_0). \tag{3}$$

Let d_{-1} and d_1 be the distances from s to the support lines of the normal segments at c_{-1} and c_1, respectively. By Lemma 4.1,

$$d_{-1} \geq \frac{|F(c_{-1}, c_0)|}{6} \geq \frac{\lambda_h^2 f(c_{-1})}{6} \overset{(3)}{>} \frac{\lambda_h^2 f(c_0)}{8},$$

$$d_1 > \frac{|F(c_0, c_1)|}{6} \geq \frac{\lambda_h^2 f(c_0)}{6}.$$

By Lemma 3.8(ii), the λ_h-slabs between c_{-1} and c_0 and between c_0 and c_1 contain at most $\ln^{1+\omega} n/3$ points with probability at least $1 - O(n^{-\Omega(\ln^\omega n)})$. Hence, for D to contain $\ln^{1+\omega} n$ points, radius(D) $> \max\{d_{-1}, d_1\} \geq \lambda_h^2 f(c_0)/6$. Note that $f(\tilde{s}) = f(c_0)$ as $\tilde{s} = c_0$ by construction. It follows that radius(*initial*(s)) $= \sqrt{\text{radius}(D)} > \lambda_h \sqrt{f(\tilde{s})}/3$. ◻

4.2 Radius of *coarse*(s)

In this section, we prove an upper bound and a lower bound on the radius of *coarse*(s).

Lemma 4.3 *Assume $\rho \geq 4$ and $\delta \leq 1/(25\rho^2)$. Let m be the constant and ψ_m be the parameter in Lemma 4.2. Let s be a sample. If $\lambda_m \leq 1/(396\rho^2)$, then*

$$\text{radius}(\mathit{coarse}(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$$

with probability at least $1 - O(n^{-\Omega(\ln^\omega n)})$.

Proof. 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 $5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$. By Lemma 4.2, $\psi_m\sqrt{f(\tilde{s})} \geq$ radius(*initial*(s)), so D contains *initial*(s) with probability at least $1 - O(n^{\Omega(\ln^\omega n)})$. We are to show that *coarse*(s) cannot grow beyond D. First, since $\lambda_m \leq 1/(396\rho^2)$,

$$\psi_m = \sqrt{11\lambda_m} \leq 1/(6\rho) \leq 1/24.$$

Observe that both s_1 and s_2 lie inside D. Since $5\rho\delta \le 1/(5\rho) \le 1/20$ and $\psi_m \le 1/24$, radius$(D) < (1-\delta)f(\tilde{s})$. Thus, the distance between any two points in $D \cap F_\delta^+$ is at most $2(1-\delta)f(\tilde{s})$. By Lemma 3.2(i), the maximum distance between $D \cap F_\delta^+$ and the tangent to F_δ^+ at s_1 is at most $\frac{(5\rho\delta+\psi_m\sqrt{f(\tilde{s})})^2}{2(1-\delta)f(\tilde{s})} \le \frac{(5\rho\delta\sqrt{f(\tilde{s})}+\psi_m\sqrt{f(\tilde{s})})^2}{2(1-\delta)f(\tilde{s})}$ as $f(\tilde{s}) \ge 1$. Thus, this distance is upper bounded by $\frac{(5\rho\delta+\psi_m)^2}{2(1-\delta)}$ which is less than $0.51(5\rho\delta+\psi_m)^2$ as $\delta \le 1/(25\rho^2)$. 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 $2\delta + 1.1(5\rho\delta+\psi_m)^2 = 2\delta + 1.1(5\rho)^2\delta^2 + 2.2(5\rho)\psi_m\delta + 1.1\psi_m^2$. Since $\delta \le 1/(25\rho^2)$ and $\psi_m \le 1/(6\rho)$, we have $1.1(5\rho)^2\delta^2 \le 1.1\delta$, $2.2(5\rho)\psi_m\delta < 1.84\delta$, and $1.1\psi_m^2 < \psi_m/\rho$. We conclude that the strip width is no more than $2\delta + 1.1\delta + 1.84\delta + \psi_m/\rho < 5\delta + \psi_m/\rho \le \text{radius}(D)/\rho$. This shows that *coarse*(s) cannot grow beyond D. ◻

Next, we bound radius(*coarse*(s)) from below. We use $f_{\max}$ to denote $\max_{x\in F} f(x)$.

Lemma 4.4 *Assume that $\delta \le 1/8$ and $\rho \ge 4$. Let h be the constant in Lemma 4.2. Let s be a sample. If $\lambda_h \le 1/32$, then*

$$\text{radius}(\mathit{coarse}(s)) \ge \max\{2\sqrt{\rho}\delta, \text{radius}(\mathit{initial}(s))\}$$

with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.

Proof. Since *coarse*(s) is grown from *initial*(s), radius(*coarse*(s)) $\ge$ radius(*initial*(s)). We are to prove that radius(*coarse*(s)) $\ge 2\sqrt{\rho}\delta$. Let D be the disk that has center s and radius radius(*coarse*(s))$/\sqrt{\rho}$. Let X be the disk centered at $\tilde{s}$ with radius δ. Note that $s \in X$ and X is tangent to F_δ^+ and F_δ^-. Since $\delta \le 1/8$ and $f(\tilde{s}) \ge 1$, $f(\tilde{s}) - \delta > \delta$ and so Lemma 3.1 implies that X lies inside the finite region bounded by F_δ^+ and F_δ^-.

Suppose that radius(*coarse*(s)) $< 2\sqrt{\rho}\delta$. Then radius$(D) < 2\delta$. If D contains X, X is a disk inside $D \cap X$ with radius at least radius$(D)/2$. If D does not contain X, then since $s \in X$, $D \cap X$ contains a disk with radius radius$(D)/2$. The width of *strip*(s) is less than or equal to radius(*coarse*(s))$/\rho$ = radius$(D)/\sqrt{\rho}$. Thus, $(D \cap X) - \mathit{strip}(s)$ contains a disk Y such that

$$\text{radius}(Y) \ge (\frac{1}{4} - \frac{1}{4\sqrt{\rho}}) \cdot \text{radius}(D) \ge \frac{\text{radius}(D)}{8}.$$

Note that Y is empty and Y lies inside the finite region bounded by F_δ^+ and F_δ^-. Take a point $p \in Y$. Since $p \in Y \subseteq D$ and radius$(D) < 2\delta$, $\|\tilde{p} - \tilde{s}\| \le \|p - \tilde{p}\| + \|s - \tilde{s}\| + \|p - s\| \le 4\delta \le 1/2$ as $\delta \le 1/8$. The Lipschitz condition implies that $f(\tilde{p}) \le 3f(\tilde{s})/2$. Observe that radius$(D)$ = radius(*coarse*(s))$/\sqrt{\rho} \ge$ radius(*initial*(s))$/\sqrt{\rho}$. Thus, Lemma 4.2 implies that radius$(Y) \ge$ radius$(D)/8 \ge \lambda_h\sqrt{f(\tilde{s})}/(24\sqrt{\rho}) > \lambda_h\sqrt{f(\tilde{p})}/(30\sqrt{\rho})$ with probability at least $1 - O(n^{-\Omega(\ln^\omega n)})$. Let $\beta = \lambda_h/(330\sqrt{\rho f_{\max}})$. Then radius$(Y) > 11\beta f(\tilde{p})$. By Lemma 3.5, Y contains a β-cell. By Lemma 3.8(i), this β-cell is empty with probability at most $n^{-\Omega(\ln^\omega n/f_{\max})}$. This implies that radius(*coarse*(s)) $< 2\sqrt{\rho}\delta$ occurs with probability at most $O(n^{-\Omega(\ln^\omega n/f_{\max})})$. ◻

4.3 Rough tangent estimate: *strip*(s)

In this section, we prove that the slope of *strip*(s) is a rough estimate of the slope of the tangent at $\tilde{s}$. We first prove a technical lemma about various properties of *coarse*(s) and F_α inside *coarse*(s). Although the lemma contains a long list of properties, their proofs are short.

Lemma 4.5 *Assume $\rho \geq 5$ and $\delta \leq 1/(25\rho^2)$. Let m be the constant and ψ_m be the parameter in Lemma 4.2. Let s be a sample. If $2\sqrt{\rho}\delta \leq \mathrm{radius}(coarse(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$ and $\psi_m \leq 1/100$, then for any F_α and for any point $x \in F_\alpha \cap coarse(s)$, the following hold:*

(i) $5\rho\delta + \psi_m \leq 0.05$, $\frac{5\rho\delta+\psi_m}{2(1-\delta)} \leq 0.03$, *and* $\frac{5\rho\delta+\psi_m+2\delta}{2(1-\delta)} \leq 0.03$,

(ii) *$F_\alpha \cap coarse(s)$ consists of one connected component,*

(iii) *the angle between the normals at s and x is at most* $2\sin^{-1}\frac{5\rho\delta+\psi_m+2\delta}{(1-\delta)} \leq 2\sin^{-1}(0.06)$,

(iv) $x \in cocone(s_1, 2\sin^{-1}\frac{5\rho\delta+\psi_m+2\delta}{2(1-\delta)}) \subseteq cocone(s_1, 2\sin^{-1}(0.03))$ *where s_1 is the point on F_α such that $\tilde{s_1} = \tilde{s}$.*

(v) $0.9f(\tilde{s}) < f(\tilde{x}) < 1.1f(\tilde{s})$,

(vi) *if x lies on the boundary of $coarse(s)$, the distance between s and the orthogonal projection of x onto the tangent at s is at least $0.8 \cdot \mathrm{radius}(coarse(s))$, and*

(vii) *for any $y \in F_\alpha \cap coarse(s)$, the acute angle between xy and the tangent at x is at most* $\sin^{-1}(6\rho\delta + 1.2\psi_m)) \leq \sin^{-1}(0.06)$.

Proof. A straightforward calculation shows (i).

If $F_\alpha \cap coarse(s)$ consists of more than one connected component, the medial axis of F_α intersects the interior of $coarse(s)$. Since F and F_α have the same medial axis, the distance from $\tilde{s}$ to the medial axis is at most $2\,\mathrm{radius}(coarse(s)) \leq 2(5\rho\delta + \psi_m\sqrt{f(\tilde{s})}) \leq 2(5\rho\delta + \psi_m)f(\tilde{s}) < f(\tilde{s})$ by (i), contradiction. This proves (ii).

Let s_1 be the point on F_α such that $\tilde{s_1} = \tilde{s}$. The distance $\|s_1 - x\| \leq \|s - x\| + \|s - s_1\| \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})} + 2\delta \leq (5\rho\delta + \psi_m + 2\delta)f(\tilde{s})$. By Lemma 3.3, the angle between the normals at s_1 and x is at most $2\sin^{-1}\frac{\|s_1-x\|}{(1-\delta)f(\tilde{s})} \leq 2\sin^{-1}\frac{5\rho\delta+\psi_m+2\delta}{(1-\delta)} \leq 2\sin^{-1}(0.06)$ by (i). This proves (iii).

By Lemma 3.2(ii), $x \in cocone(s_1, 2\sin^{-1}\frac{\|s_1-x\|}{2(1-\delta)f(\tilde{s})}) \subseteq cocone(s_1, 2\sin^{-1}(0.03))$. This proves (iv).

The distance $\|\tilde{s} - \tilde{x}\| \leq \|s - \tilde{s}\| + \|s - x\| + \|x - \tilde{x}\| \leq (5\rho\delta + \psi_m + 4\delta)f(\tilde{s}) < 0.1f(\tilde{s})$. Then the Lipschitz condition implies (v).

Consider (vi). Refer to Figure 8. Assume that the tangent at s is horizontal. By sine law, $\sin\angle sxs_1 = \frac{\|s-s_1\|\cdot\sin\angle ss_1x}{\|s-x\|} \leq \frac{2\delta}{\mathrm{radius}(coarse(s))}$ as $\|s - s_1\| \leq 2\delta$ and $\|s - x\| = \mathrm{radius}(coarse(s))$. Since $\mathrm{radius}(coarse(s)) \geq 2\sqrt{\rho}\delta$ and $\rho \geq 5$, we have $\angle sxs_1 \leq \sin^{-1}\frac{1}{\sqrt{\rho}} < \sin^{-1}(0.5)$. By (iv), $\angle s_1sx \geq \pi - \angle sxs_1 - (\pi/2 + \sin^{-1}(0.03)) > \pi/2 - \sin^{-1}(0.5) - \sin^{-1}(0.03)$. Thus, the horizontal distance between s and x is equal to $\|s-x\|\cdot\sin\angle s_1sx \geq \|s-x\|\cdot\cos(\sin^{-1}(0.5)+\sin^{-1}(0.03)) > 0.8 \cdot \|s - x\|$.

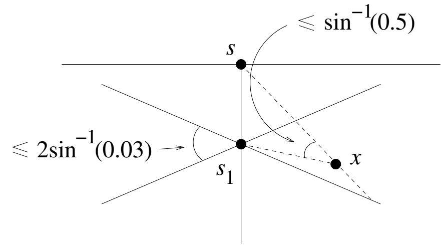


Figure 8:

Consider (vii). Since $y \in F_\alpha \cap coarse(s)$, $\|x-y\| \leq 2\,\text{radius}(coarse(s)) \leq 2(5\rho\delta+\psi_m)f(\tilde{s}) < 0.1f(\tilde{s})$ by (i). So Lemma 3.2(ii) applies and the acute angle between xy and the tangent at x is at most $\sin^{-1}\frac{\|x-y\|}{2(1-\delta)f(\tilde{x})} \leq \sin^{-1}\frac{(5\rho\delta+\psi_m)f(\tilde{s})}{(1-\delta)f(\tilde{x})}$. Since $f(\tilde{x}) \geq 0.9f(\tilde{s})$ by (v) and $\delta \leq 1/(25\rho^2)$, the acute angle is less than $\sin^{-1}(1.2(5\rho\delta+\psi_m))$, which is less than $\sin^{-1}(0.06)$ by (i). ◻

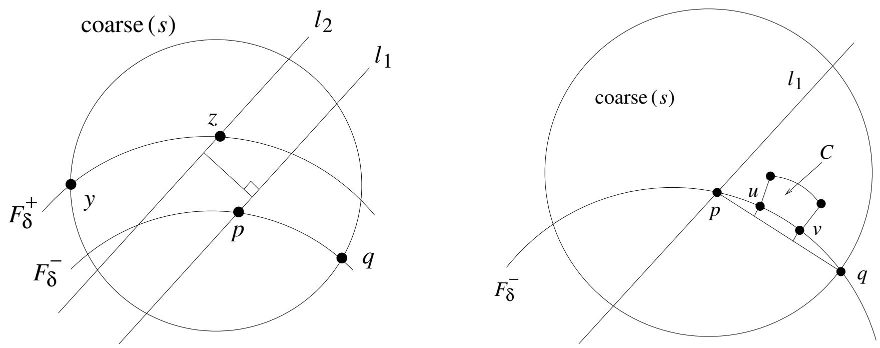
We are ready to analyze the slope of *strip*(s). We highlight the key ideas before giving the proof. Let $\mathcal{B}$ be the region between F_δ^+ and F_δ^- inside *coarse*(s). If *strip*(s) makes a large angle with the tangent at $\tilde{s}$, *strip*(s) would cut through $\mathcal{B}$ in the middle. In this case, if $\mathcal{B} \cap strip(s)$ is narrow, there would be a lot of areas in $\mathcal{B}$ outside *strip*(s). But these areas must be empty which occur with low probability. Otherwise, if $\mathcal{B} \cap strip(s)$ is wide, we show that *strip*(s) can be rotated to reduce its width further, contradiction. We give the detailed proof below.

Lemma 4.6 *Assume that $\rho \geq 5$ and $\delta \leq 1/(25\rho^2)$. Let m be the constant and ψ_m be the parameter in Lemma 4.2. Let s be a sample. For sufficiently large n, the acute angle between the tangent at $\tilde{s}$ and the direction of strip(s) is at most* $3\sin^{-1}\frac{5\rho\delta+\psi_m+2\delta}{(1-\delta)}+\sin^{-1}(6\rho\delta+1.2\psi_m) \leq 4\sin^{-1}(0.06)$ *with probability at least* $1-O(n^{-\Omega(\ln^\omega n/f_{\max})})$.

Proof. Let ℓ_1 and ℓ_2 be the lower and upper bounding lines of *strip*(s). Without loss of generality, we assume that the normal at $\tilde{s}$ is vertical, the slope of *strip*(s) is non-negative, $F_\delta^- \cap coarse(s)$ lies below $F_\delta^+ \cap coarse(s)$, and $\psi_m \leq 1/100$ for sufficiently large n. Let h and m be the constants and ψ_h and ψ_m be the parameters in Lemma 4.2. We first assume that $\max\{2\sqrt{\rho}\delta, \psi_h\sqrt{f(\tilde{s})}\} \leq \text{radius}(coarse(s)) \leq 5\rho\delta+\psi_m\sqrt{f(\tilde{s})}$ and take the probability of its occurrence into consideration later. As a short hand, we use η_1 to denote $\frac{5\rho\delta+\psi_m+2\delta}{(1-\delta)}$ and η_2 to denote $6\rho\delta+1.2\psi_m$.

Observe that both ℓ_1 and ℓ_2 must intersect the space that lies between F_δ^+ and F_δ^- inside *coarse*(s). Otherwise, we can squeeze *strip*(s) and reduce its width, contradiction. If ℓ_1 intersects $F_\alpha \cap coarse(s)$ twice for some α, then ℓ_1 is parallel to the tangent at some point on $F_\alpha \cap coarse(s)$. By Lemma 4.5(iii), the direction of *strip*(s) makes an angle at most $2\sin^{-1}\eta_1$ with the horizontal and we are done. Similarly, we are done if ℓ_2 intersects $F_\alpha \cap coarse(s)$ twice for some α. The remaining case is that both ℓ_1 and ℓ_2 intersect $F_\alpha \cap coarse(s)$ for any α at most once. Suppose that the acute angle between the direction of *strip*(s) and the horizontal is more than $3\sin^{-1}\eta_1+\sin^{-1}\eta_2$. We show that this occurs with probability $O(n^{-\Omega(\ln^\omega n/f_{\max})})$.

Let q be the right intersection point between F_δ^- and the boundary of *coarse*(s). If ℓ_1 intersects $F_\delta^- \cap coarse(s)$, let p denote the intersection point; otherwise, let p denote the leftmost



(a) (b)

Figure 9:

intersection point between F_δ^- and the boundary of *coarse*(s). Refer to Figure 9(a). We claim that $F_\delta^-(p,q)$ lies below ℓ_1. If ℓ_1 does not intersect $F_\delta^- \cap$ *coarse*(s), then this is clearly true. Otherwise, by Lemma 4.5(iii), the magnitude of the slope of the tangent at p is at most $2\sin^{-1}\eta_1$. Since the slope of ℓ_1 is more than $3\sin^{-1}\eta_1 + \sin^{-1}\eta_2$, F_δ^- crosses ℓ_1 at p from above to below. So $F_\delta^-(p,q)$ lies below ℓ_1.

We show that $\|p-q\| \leq \psi_h\sqrt{f(\tilde{s})}/2$ with probability at least $1 - n^{-\Omega(\ln^\omega n/f_{\max})}$. Notice that pq is parallel to the tangent to F_δ^- at some point on $F_\delta^-(p,q)$. By Lemma 4.5(iii), the tangent to $F_\delta^-(p,q)$ turns by an angle at most $4\sin^{-1}(0.06) < \pi/2$ from p to q. This implies that $F_\delta^-(p,q)$ is monotone with respect to the direction perpendicular to pq. We divide pq into three equal segments. Let u and v be the intersection points between $F_\delta^-(p,q)$ and the perpendiculars of pq at the dividing points. Assume that v follows u along $F_\delta^-(p,q)$. Refer to Figure 9(b). Suppose that $\|p-q\| > \psi_h\sqrt{f(\tilde{s})}/2$. Then

$$|F_\delta^-(u,v)| \geq \frac{\|p-q\|}{3} \geq \frac{\psi_h\sqrt{f(\tilde{s})}}{6}. \tag{4}$$

Since $f(\tilde{u}) < 1.1f(\tilde{s})$ by Lemma 4.5(v), $|F_\delta^-(u,v)| > \psi_h\sqrt{f(\tilde{u})}/7$. Consider a $(\lambda_k/\sqrt{f_{\max}})$-grid where $k = h/231$ and $\tilde{u}$ is a cut-point. (Note that $\lambda_k = \psi_h/77$.) Let C be the $(\lambda_k/\sqrt{f_{\max}})$-cell that touches $F_\delta^-(u,v)$ and the normal segment through u. By Lemma 3.5, the diameter of C is at most $11\lambda_k\sqrt{f(\tilde{u})} = \psi_h\sqrt{f(\tilde{u})}/7 < |F_\delta^-(u,v)|$. So the bottom side of C lies inside $F_\delta^-(u,v)$. Let $\mathcal{R}$ be the region inside *coarse*(s) that lies below ℓ_1 and above $F_\delta^-(p,q)$. From any point $x \in F_\delta^-(u,v)$, if we shoot a ray along the normal at x into $\mathcal{R}$, either the ray will leave C first or the ray will hit ℓ_1 or the boundary of *coarse*(s) in $\mathcal{R}$. We are to prove that the distances from x to ℓ_1 and the boundary of *coarse*(s) in $\mathcal{R}$ are more than $2\lambda_k\delta$, which is at least $2\lambda_k\delta/\sqrt{f_{\max}}$. This shows that the ray always leaves C first, so C lies completely inside *coarse*(s) and below ℓ_1. Then the upper bound on $\|p-q\|$ follows as C is empty with probability at most $n^{-\Omega(\ln^\omega n/f_{\max})}$ by Lemma 3.8(i).

Consider the distance from any point $x \in F_\delta^-(u,v)$ to ℓ_1. By Lemma 4.5(iii), the angle between ℓ_1 and the tangent at p (measured by rotating ℓ_1 in the clockwise direction) is at least $3\sin^{-1}\eta_1 + \sin^{-1}\eta_2 - 2\sin^{-1}\eta_1 = \sin^{-1}\eta_1 + \sin^{-1}\eta_2$ and at most $\pi/2 + 2\sin^{-1}\eta_1$. By

Lemma 4.5(vii), the acute angle between px and the tangent at p is at most $\sin^{-1}\eta_2$. So the angle between px and ℓ_1 is at least $\sin^{-1}\eta_1$ and at most $\pi/2 + 2\sin^{-1}\eta_1 + \sin^{-1}\eta_2$. This implies that the distance from x to ℓ_1 is at least $\|p-x\| \cdot \min\{\eta_1\,,\, \cos(2\sin^{-1}\eta_1 + \sin^{-1}\eta_2)\}$. By Lemma 4.5(i), $\eta_1 \le 0.06 < \cos(3\sin^{-1}(0.06)) \le \cos(2\sin^{-1}\eta_1 + \sin^{-1}\eta_2)$. Therefore, the distance from x to ℓ_1 is at least $\|p-x\| \cdot \eta_1 > 5\rho\delta \cdot \|p-x\| \ge 25\delta \cdot (\|p-q\|/3) \overset{(4)}{>} 4\delta\psi_h\sqrt{f(\tilde{s})}$. Since $\lambda_k = \psi_h/77$, this distance is greater than $2\lambda_k\delta$.

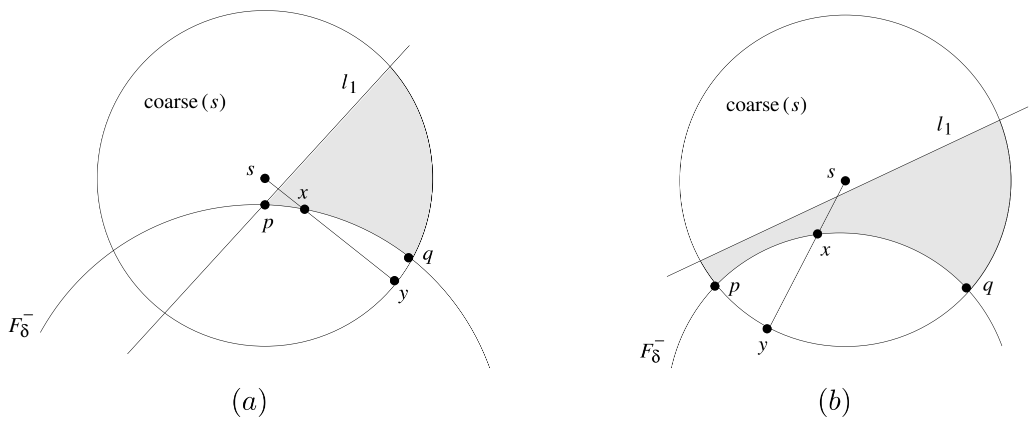


Figure 10:

Next, we consider the distance d from any point $x \in F_\delta^-(u,v)$ to the boundary of *coarse*(s) in $\mathcal{R}$. Take a radius sy of *coarse*(s) that passes through x. Suppose that y lies outside $\mathcal{R}$. Refer to Figure 10. If ℓ_1 intersects $F_\delta^- \cap$ *coarse*(s) at p (Figure 10(a)), then $d = \|q-x\|$. If ℓ_1 does not intersect $F_\delta^- \cap$ *coarse*(s) (Figure 10(b)), then $d = \min\{\|p-x\|\,,\,\|q-x\|\}$. Thus, by (4), $d \ge \|p-q\|/3 \ge \psi_h\sqrt{f(\tilde{s})}/6 > 2\lambda_k\delta$. The remaining possibility is that y lies on the boundary of $\mathcal{R}$. Then either sy is tangent to F_δ^- at x or sy intersects $F_\delta^- \cap$ *coarse*(s) at least twice. So xy is parallel to the tangent at some point on $F_\delta^- \cap$ *coarse*(s). By Lemma 4.5(iii), the acute angle between xy and the tangent at x is at most $4\sin^{-1}\eta_1$. By Lemma 4.5(vii), the acute angle between qx and the tangent at x is at most $\sin^{-1}\eta_2$. So the angle between qx and xy is at most $4\sin^{-1}\eta_1 + \sin^{-1}\eta_2$. It follows that $d = \|x-y\| \ge \|q-x\| \cdot \cos(4\sin^{-1}\eta_1 + \sin^{-1}\eta_2) \ge \|q-x\| \cdot \cos(5\sin^{-1}(0.06)) > 0.9 \cdot \|q-x\| \ge 0.9 \cdot (\|p-q\|/3) \ge 0.15\psi_h\sqrt{f(\tilde{s})} > 2\lambda_k\delta$.

In all, C lies below ℓ_1 and inside *coarse*(s). So C must be empty which occurs with probability at most $n^{-\Omega(\ln^\omega n/f_{\max})}$ by Lemma 3.8(i). It follows that $\|p-q\| \le \psi_h\sqrt{f(\tilde{s})}/2$ with probability at least $1 - n^{-\Omega(\ln^\omega n/f_{\max})}$. By Lemma 4.5(vi), the horizontal distance between q and the left intersection point between F_δ^- and the boundary of *coarse*(s) is at least $1.6 \cdot \mathrm{radius}(\textit{coarse}(s)) \ge 1.6\psi_h\sqrt{f(\tilde{s})} > \|p-q\|$. We conclude that ℓ_1 intersects $F_\delta^- \cap$ *coarse*(s) exactly once at p.

Refer to Figure 9(a) and Figure 11. Let y be the leftmost intersection point between F_δ^+ and the boundary of *coarse*(s). Symmetrically, we can also show that ℓ_2 intersects $F_\delta^+ \cap$ *coarse*(s) exactly once at some point z, $F_\delta^+(y,z)$ lies above ℓ_2, and $\|y-z\| \le \psi_h\sqrt{f(\tilde{s})}/2$ with probability at least $1 - n^{-\Omega(\ln^\omega n/f_{\max})}$.

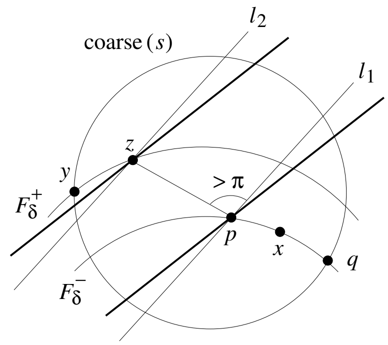


Figure 11:

Consider the projections of $F_\delta^+(y,z)$ and $F_\delta^-(p,q)$ onto the horizontal diameter of $coarse(s)$ through s. By Lemma 4.5(vi), the projections of y and q are at distance at least $0.8 \cdot \text{radius}(coarse(s))$ from s. Thus, the distance between the projections of $F_\delta^+(y,z)$ and $F_\delta^-(p,q)$ is at least $1.6 \cdot \text{radius}(coarse(s)) - \|p-q\| - \|y-z\| \geq 1.6 \cdot \text{radius}(coarse(s)) - \psi_h\sqrt{f(\tilde{s})} \geq 1.6 \cdot \text{radius}(coarse(s)) - \text{radius}(coarse(s)) > \text{radius}(coarse(s))/\rho$. That is, this distance is greater than the width of $strip(s)$. But then we can rotate ℓ_1 and ℓ_2 around p and z, respectively, in the clockwise direction to reduce the width of $strip(s)$ while not losing any sample inside $coarse(s)$. See Figure 11. This is impossible. It follows that, under the condition that $\max\{2\sqrt{\rho}\delta, \psi_h\sqrt{f(\tilde{s})}\} \leq \text{radius}(coarse(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$, the acute angle between the direction of $strip(s)$ and the tangent at $\tilde{s}$ is at most $3\sin^{-1}\eta_1 + \sin^{-1}\eta_2$ with probability at least $1 - O(n^{\Omega(\ln^\omega n/f_{\max})})$. By Lemmas 4.2, 4.3, and 4.4, the inequalities $\max\{2\sqrt{\rho}\delta, \psi_h\sqrt{f(\tilde{s})}\} \leq \text{radius}(coarse(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$ hold with probability at least $1 - O(n^{\Omega(\ln^\omega n)/f_{\max}})$. So the lemma follows. ◻

5 Guarantees

The results in Section 4 show that after the step Coarse Neighborhood, the algorithm already has a normal estimate at each noisy sample with an error roughly equal to $\Theta(\delta)$. However, this error bound does not tend to zero as the sampling density increases. This explains the need for the step Refined Neighborhood in the algorithm. This step will improve the normal estimate so that the error tends to zero as the sampling density increases. This will allow us to prove the pointwise convergence. Afterwards, we prove that the reconstruction is homeomorphic to the true curve. Finally, we combine these results to prove our main result that the reconstruction is faithful.

5.1 Pointwise convergence

Recall that our algorithm computes a center point for each sample. Eventually, a subset of these center points become the vertices of the output curve. Our goal is to show that all center points converge to F as n tends to ∞. To this end, we show that our algorithm aligns *refined*(s) approximately well with the normal at $\tilde{s}$. Then we prove the pointwise convergence. (See Lemmas 5.3 and 5.4.) We first prove two technical lemmas, Lemmas 5.1 and 5.2.

5.1.1 Technical lemmas

In the step Refined Neighborhood, we align *candidate*(s,θ) with the normal at $\tilde{s}$ by varying θ within $[-\pi/10, \pi/10]$. Recall that θ is the signed acute angle between the upward direction of *candidate*(s,θ) and N_s, where N_s is the upward direction perpendicular to *strip*(s). Let *angle*(*strip*(s)) denote the signed acute angle between N_s and the upward normal at $\tilde{s}$. If N_s points to the right of the upward normal at $\tilde{s}$, *angle*(*strip*(s)) is positive. Otherwise, *angle*(*strip*(s)) is negative. We define $\theta_s = \theta + \mathit{angle}(\mathit{strip}(s))$. That is, θ_s is the signed acute angle between the upward direction of *candidate*(s,θ) and the upward normal at $\tilde{s}$. The sign of θ_s is determined in the same way as *angle*(*strip*(s)). For any F_α and for any point $p \in F_\alpha \cap \mathit{candidate}(s,\theta)$, let γ_p be the signed acute angle between the upward direction of *candidate*(s,θ) and the upward normal at $\tilde{p}$. The sign of γ_p is determined in the same way as *angle*(*strip*(s)).

Lemma 5.1 *Assume that $\delta \leq 1/(25\rho^2)$ and $\rho \geq 5$. Let s be a sample. Let W_s be the width of candidate(s,θ). For sufficiently large n, the following hold with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$ throughout the variation of θ within $[-\pi/10, \pi/10]$.*

(i) $W_s \leq 0.1 f(\tilde{s})$.

(ii) $\theta_s \in [-\pi/5, \pi/5]$ and $\theta_s = 0$ for some $\theta \in [-\pi/10, \pi/10]$.

(iii) Any line, which is parallel to candidate(s,θ) and inside candidate(s,θ), intersects $F_\alpha \cap$ coarse(s) for any α exactly once.

(iv) For any F_α and for any point $p \in F_\alpha \cap$ candidate(s,θ), $\theta_s - 0.2|\theta_s| - 3W_s/f(\tilde{s}) \leq \gamma_p \leq \theta_s + 0.2|\theta_s| + 3W_s/f(\tilde{s})$.

Proof. We first assume that $\max\{2\sqrt{\rho}\delta, \psi_h\sqrt{f(\tilde{s})}\} \leq \mathrm{radius}(\mathit{coarse}(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$ and $\mathrm{radius}(\mathit{initial}(s)) \leq \psi_m\sqrt{f(\tilde{s})}$. We will take the consideration of the probabilities of their occurrences later.

Since $W_s \leq \sqrt{\mathrm{radius}(\mathit{initial}(s))} \leq \sqrt{\psi_m} f(\tilde{s})^{1/4}$ and $\psi_m \leq 0.01$ for sufficiently large n, $W_s \leq 0.1 f(\tilde{s})$. This proves (i).

By Lemma 4.6, for sufficiently large n, $|\mathit{angle}(\mathit{strip}(s))| \leq 4\sin^{-1}(0.06) < \pi/10$. Since $\theta \in [-\pi/10, \pi/10]$, $\theta_s = \theta + \mathit{angle}(\mathit{strip}(s)) \in [-\pi/5, \pi/5]$ and $\theta_s = 0$ for some θ. This proves (ii).

Consider (iii). Let ℓ be a line that is parallel to *candidate*(s,θ) and inside *candidate*(s,θ). We first prove that ℓ intersects F_α. Refer to Figure 12. Without loss of generality, assume that the normal at $\tilde{s}$ is vertical, the slope of *candidate*(s,θ) is positive, and ℓ is below s. Let s_1 and s_2 be the points on F^+_δ and F^-_δ, respectively, such that $\tilde{s_1} = \tilde{s_2} = \tilde{s}$. Shoot two rays upward from s_1 with slopes $\pm\sin^{-1}(0.03)$. Also, shoot two rays downward from s_2 with slopes $\pm\sin^{-1}(0.03)$. Let $\mathcal{R}$ be the region inside *coarse*(s) bounded by these four rays. By

Lemma 4.5(iv), $F_\alpha \cap coarse(s)$ lies inside $\mathcal{R}$. Let x be the upper right vertex of $\mathcal{R}$. Let y be the right endpoint of a horizontal chord through s_1. Let L be the line that passes through x and is parallel to ℓ. Let L' be the line that passes through s and is parallel to ℓ. Let z be the point on L such that $s_1 z$ is perpendicular to L.

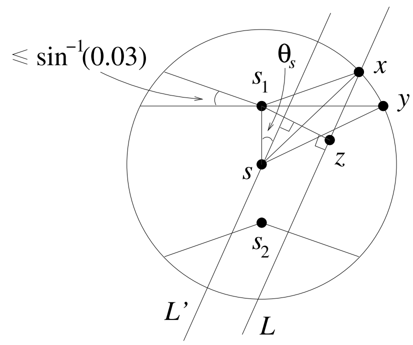


Figure 12:

We claim that L' is above L and L and L' intersect both the upper and lower boundaries of $\mathcal{R}$. By Lemma 4.5(iv), $\angle x s_1 y \le \sin^{-1}(0.03)$, so $\angle xsy \le 2\sin^{-1}(0.03)$. Observe that $\cos\angle s_1 s y = \frac{\|s-s_1\|}{\|s-y\|} \le \frac{2\delta}{\text{radius}(coarse(s))}$. Since radius$(coarse(s)) \ge 2\sqrt{\rho}\delta$, $\cos\angle s_1 s y \le 1/\sqrt{\rho} \le 1/\sqrt{5}$ which implies that $\angle s_1 s y > \pi/3$. Since $\angle s_1 s x = \angle s_1 s y - \angle xsy$, we get

$$\angle s_1 s x \ge \pi/3 - 2\sin^{-1}(0.03) > \pi/5 \ge |\theta_s|. \tag{5}$$

So L' cuts through the angle between ss_1 and sx. It follows that L' is above L. Observe that L' intersects $s_1 x$. By symmetry, L' intersects the left downward ray from s_2 too. We conclude that L and L' intersect both the upper and lower boundaries of $\mathcal{R}$.

Since $|\theta_s| \le \pi/5$ and $\angle sxz = \angle s_1 s x - |\theta_s|$, by (5), $\angle sxz \ge \pi/3 - 2\sin^{-1}(0.03) - \pi/5 > 0.3$. The distance from s to L is equal to $\|s-x\|\cdot\sin\angle sxz > \|s-x\|\cdot\sin(0.3) > 0.2\cdot\text{radius}(coarse(s))$. Recall that ℓ lies below s by our assumption. The distance between ℓ and s is at most $W_s/2$ and our algorithm enforces that $W_s/2 \le \text{radius}(coarse(s))/6$. So ℓ lies between L' and L. Since L and L' intersect both the upper and lower boundaries of $\mathcal{R}$, so does ℓ. It follows that ℓ must intersect $F_\alpha \cap coarse(s)$.

Next, we show that ℓ intersects $F_\alpha \cap coarse(s)$ exactly once. If not, ℓ is parallel to the tangent at some point on $F_\alpha \cap coarse(s)$. By Lemma 4.5(iii), the angle between ℓ and the vertical is at least $\pi/2 - 2\sin^{-1}(0.06) > \pi/5$, contradicting the fact that $|\theta_s| \le \pi/5$.

Consider (iv). Let ℓ be a line that is parallel to $candidate(s,\theta)$ and passes through s. By (iii), ℓ intersects F_α at some point b. We first prove that $\theta_s - 0.2|\theta_s| \le \gamma_b \le \theta_s + 0.2|\theta_s|$. Let s_1 be the point on F_α such that $\tilde{s} = \tilde{s_1}$. Assume that the tangent at s is horizontal, s is above s_1, and b is to the left of s. Let C be the circle tangent to F_α at s_1 that lies below s_1, is centered at x, and has radius $f(\tilde{s}) - \delta$. By Lemma 3.1, F_α does not intersect the interior of C. Refer to Figure 13(a). Let sa be a tangent to C that lies on the left of x. We claim that $\angle asx > |\theta_s|$. Otherwise, $\|s-x\| \ge \|a-x\|/\sin(\pi/5) = (f(\tilde{s})-\delta)/\sin(\pi/5) > f(\tilde{s})+\delta \ge \|s-x\|$, contradiction. So sb lies between sa and sx. Let sr be the extension of sb such that r lies on

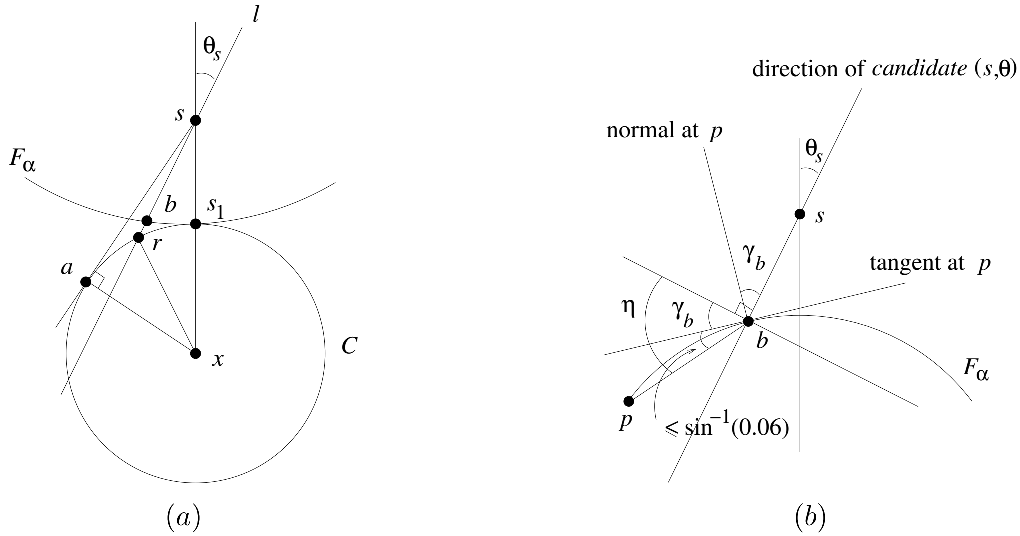


Figure 13:

C. We have $\|a-s\| = \sqrt{\|s-x\|^2 - \|a-x\|^2} \leq \sqrt{(f(\tilde{s})+\delta)^2 - (f(\tilde{s})-\delta)^2} = 2\sqrt{\delta f(\tilde{s})}$. Thus, $\|r-s\| \leq \|a-s\| \leq 2\sqrt{\delta f(\tilde{s})}$. Observe that

$$\angle rxs = \sin^{-1}\frac{\|r-s\|\cdot\sin|\theta_s|}{\|r-x\|} \leq \sin^{-1}\frac{2\sqrt{\delta f(\tilde{s})}\cdot|\theta_s|}{\|r-x\|}.$$

Since $\delta \leq 1/(25\rho^2)$ and $|\theta_s| \leq \pi/5$, we have

$$\frac{2\sqrt{\delta f(\tilde{s})}\cdot|\theta_s|}{\|r-x\|} = \frac{2\sqrt{\delta f(\tilde{s})}\cdot|\theta_s|}{f(\tilde{s})-\delta} \leq \frac{2\sqrt{\delta}\cdot|\theta_s|}{\sqrt{f(\tilde{s})}-\delta/\sqrt{f(\tilde{s})}} \leq \frac{2\sqrt{\delta}\cdot|\theta_s|}{1-\delta} < 0.06. \quad (6)$$

Combing (6) with the following fact

$$x \leq 0.6 \Rightarrow \sin^{-1}x < 1.1x, \quad (7)$$

we get $\angle rxs < \frac{2.2\sqrt{\delta f(\tilde{s})}\cdot|\theta_s|}{\|r-x\|}$. Since $\|b-s_1\| \leq \|r-s_1\| = \|r-x\|\cdot 2\sin\frac{\angle rxs}{2}$, we get

$$\|b-s_1\| \leq \|r-x\|\cdot\angle rxs \leq 2.2\sqrt{\delta f(\tilde{s})}\cdot|\theta_s|.$$

Let γ' be the acute angle between the normals at b and s_1. By Lemma 3.3, $\gamma' \leq 2\sin^{-1}\frac{\|b-s_1\|}{(1-\alpha)f(\tilde{s})} \leq 2\sin^{-1}\frac{2.2\sqrt{\delta}\cdot|\theta_s|}{1-\alpha} \leq 2\sin^{-1}\frac{2.2\sqrt{\delta}\cdot|\theta_s|}{1-\delta}$. By (6) and (7), we conclude that $\gamma' < \frac{4.84\sqrt{\delta}\cdot|\theta_s|}{1-\delta} < 0.2|\theta_s|$. It follows that

$$\theta_s - 0.2|\theta_s| \leq \theta_s - \gamma' \leq \gamma_b \leq \theta_s + \gamma' \leq \theta_s + 0.2|\theta_s|.$$

Next, we prove the upper and lower bounds for γ_p for any point $p \in F_\alpha \cap \mathit{candidate}(s,\theta)$. Let η be the acute angle between bp and the line that passes through b and is perpendicular to $\mathit{candidate}(s,\theta)$. See Figure 13(b). By Lemma 4.5(vii), the acute angle between bp and the tangent at b is at most $\sin^{-1}(0.06)$. It follows that $\eta \leq \gamma_b + \sin^{-1}(0.06) \leq \theta_s + 0.2|\theta_s| + \sin^{-1}(0.06) \leq 1.2(\pi/5) + \sin^{-1}(0.06) < 0.9$. Thus,

$$\|b-p\| \leq \frac{W_s}{2\cos\eta} < 0.9W_s.$$

Note that $W_s \leq \text{radius}(\textit{coarse}(s))/3 \leq (5\rho\delta + \psi_m)f(\tilde{s})/3$, which is less than $0.02f(\tilde{s})$ by Lemma 4.5(i). Also, by Lemma 4.5(v), $f(\tilde{p}) \geq 0.9f(\tilde{s})$. It follows that

$$\|b - p\| < 0.9W_s \leq 0.02f(\tilde{p}). \tag{8}$$

So we can invoke Lemma 3.3 to bound the angle γ'' between the normals at b and p:

$$\gamma'' \leq 2\sin^{-1}\frac{\|b-p\|}{(1-\alpha)f(\tilde{p})} \leq 2\sin^{-1}\frac{0.9W_s}{(1-\alpha)f(\tilde{p})} \leq 2\sin^{-1}\frac{W_s}{f(\tilde{p})}.$$

By (8), $W_s/f(\tilde{p}) < 0.03$. So by (7), we get $\gamma'' \leq 2.2W_s/f(\tilde{p})$. Since $f(\tilde{p}) \geq 0.9f(\tilde{s})$, we conclude that $\gamma'' < 3W_s/f(\tilde{s})$. This implies that

$$\theta_s - 0.2|\theta_s| - 3W_s/f(\tilde{s}) \leq \gamma_b - \gamma'' \leq \gamma_p \leq \gamma_b + \gamma'' \leq \theta_s + 0.2|\theta_s| + 3W_s/f(\tilde{s}).$$

Finally, we have proved the lemma under the conditions that $\max\{2\sqrt{\rho}\delta, \psi_h\sqrt{f(\tilde{s})}\} \leq \text{radius}(\textit{coarse}(s)) \leq 5\rho\delta + \psi_m\sqrt{f(\tilde{s})}$ and $\text{radius}(\textit{initial}(s)) \leq \psi_m\sqrt{f(\tilde{s})}$. These conditions hold with probabilities at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$ by Lemmas 4.2, 4.3, and 4.4. So the lemma follows. ◻

Lemma 5.2 *Assume that $\delta \leq 1/(25\rho^2)$ and $\rho \geq 5$. Let s be a sample. Let H be a strip that is parallel to candidate(s,θ) and lies inside candidate(s,θ). For any F_α and for any two points u and v on on $F_\alpha \cap H$, whenever n is sufficiently large, the following hold with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.*

(i) $\|u - v\| < 3\,\text{width}(H)$.

(ii) The angle between the normals at u and v is at most $9\,\text{width}(H)$.

(iii) The acute angle between uv and the tangent to F_α at u is at most $5\,\text{width}(H)$.

Proof. Let ϕ be the acute angle between uv and the tangent to F_α at u. Let η be the acute angle between uv and the direction of *candidate*(s,θ). By Lemma 4.5(vii), $\phi \leq \sin^{-1}(0.06)$. So $\eta \geq \pi/2 - \gamma_u - \phi \geq \pi/2 - \gamma_u - \sin^{-1}(0.06)$. By Lemma 5.1(i), (ii), and (iv), $\eta \geq \pi/2 - 1.2(\pi/5) - 3(0.1) - \sin^{-1}(0.06) > 0.4$. Thus, $\|u - v\| \leq \frac{\text{width}(H)}{\sin\eta} \leq \frac{\text{width}(H)}{\sin(0.4)} < 3\,\text{width}(H)$. This proves (i).

Consider (ii). Note that $W_s \leq \text{radius}(\textit{coarse}(s))/3 \leq (5\rho\delta + \psi_m)f(\tilde{s})/3$. So by (i), $\|u - v\| \leq 3W_s \leq (5\rho\delta + \psi_m)f(\tilde{s})$. By Lemma 4.5(i) and (v), $5\rho\delta + \psi_m \leq 0.05$ and $f(\tilde{u}) \geq 0.9f(\tilde{s})$. It follows that

$$\|u - v\| < 0.06f(\tilde{u}). \tag{9}$$

Thus, we can invoke Lemma 3.3 to bound the angle ξ between the normals at u and v:

$$\xi \leq 2\sin^{-1}\frac{\|u-v\|}{(1-\alpha)f(\tilde{u})} \leq 2\sin^{-1}\frac{3\,\text{width}(H)}{0.9(1-\alpha)f(\tilde{s})} < 2\sin^{-1}\frac{4\,\text{width}(H)}{f(\tilde{s})}.$$

Since $4\,\text{width}(H)/f(\tilde{s}) \leq 4W_s/f(\tilde{s})$ which is at most 0.4 by Lemma 5.1(i), we can apply (7) to conclude that $\xi < 9\,\text{width}(H)/f(\tilde{s}) \leq 9\,\text{width}(H)$. This proves (ii).

Finally, by (9), we can invoke Lemma 3.2(ii) to bound the acute angle between uv and the tangent at u. This angle is at most $\sin^{-1}\frac{\|u-v\|}{2(1-\alpha)f(\tilde{u})}$ which is less than $\xi/2$. ◻

5.1.2 Convergence lemmas

Our algorithm varies θ so as to minimize the height of *rectangle*(s, θ). Let θ^* denote the minimizing angle. Recall that *refined*(s) = *rectangle*(s, θ^*). Let θ^*_s denote θ^* + *angle*(*strip*(s)). We apply the technical lemmas in the previous subsection to show that θ^*_s is very small, i.e., *refined*(s) aligns quite well with the normal direction at $\tilde{s}$.

Lemma 5.3 *Assume that $\delta \leq 1/(25\rho^2)$ and $\rho \geq 5$. Let s be a sample. Let W_s be the width of refined(s). For sufficiently large n, $|\theta^*_s| \leq 23W_s$ with probability at least $1 - O(n^{\Omega(\ln^\omega n / f_{\max})})$.*

Proof. We rotate the plane such that *candidate*(s, θ^*) is vertical. Suppose that $|\theta^*_s| > 23W_s$. We first assume that Lemmas 4.2, 4.3, 4.4, 5.1, and 5.2 hold deterministically and show that a contradiction arises with probability at least $1 - O(n^{\Omega(\ln^\omega n / f_{\max})})$. Since these lemmas hold with probability at least $1 - O(n^{\Omega(\ln^\omega n / f_{\max})})$, we can then conclude that $|\theta^*_s| > 23W_s$ occurs with probability at most $O(n^{\Omega(\ln^\omega n / f_{\max})})$.

Without loss of generality, we assume that $\theta^*_s > 0$. That is, the upward normal at s points to the left. Let L be the left boundary line of *candidate*(s, θ^*). By Lemma 5.1(iii), L intersects $F^-_\delta \cap$ *coarse*(s) exactly once. We use p to denote the point $L \cap F^-_\delta \cap$ *coarse*(s). We first prove a general claim which will be useful later.

Claim 1 *Orient space such that candidate(s, θ) is vertical. If $\theta_s > 23W_s$, then for any α, F_α rises strictly from left to right.*

Proof. Take any point $z \in F_\alpha \cap$ *candidate*(s, θ). By Lemma 5.1(iv), $\gamma_z \geq 0.8\theta_s - 3W_s$, which is positive as $\theta_s \geq 23W_s$ by assumption. Therefore, the upward normal at z points to the left, so the slope of the tangent to F_α at z is positive. ◻

Let h and m be the constants in Lemma 4.2. Let $k = h/3024$. Let H_1 be the strip inside *candidate*(s, θ^*) such that H_1 is bounded by L on the left and width$(H_1) = W_s/3$. Let H be the strip inside *candidate*(s, θ^*) that is bounded by L on the left and has width $28\lambda_k\sqrt{f(\tilde{s})}$. Refer to Figure 14. Since radius(*initial*(s)) $\leq \psi_m\sqrt{f(\tilde{s})}$, radius(*initial*$(s)$) < 1 for sufficiently large n. So

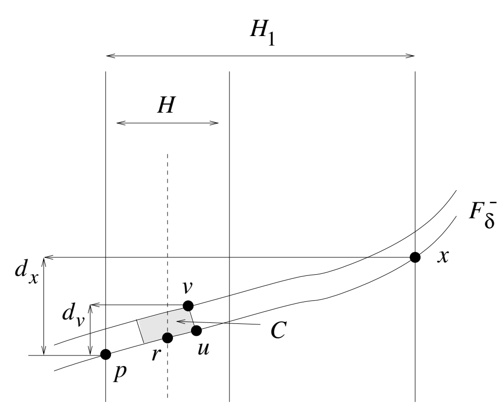


Figure 14:

$\sqrt{\text{radius}(\mathit{initial}(s))} > \text{radius}(\mathit{initial}(s))$. Since $W_s = \min\{\sqrt{\text{radius}(\mathit{initial}(s))}, \frac{\text{radius}(\mathit{coarse}(s))}{3}\}$, $W_s \geq \text{radius}(\mathit{initial}(s))/3 \geq \lambda_h\sqrt{f(\tilde{s})}/9$. We get

$$\text{width}(H) = 28\lambda_k\sqrt{f(\tilde{s})} = \frac{\lambda_h\sqrt{f(\tilde{s})}}{108} \leq \frac{W_s}{12}. \tag{10}$$

Thus, H lies inside H_1. Take any $(\lambda_k/\sqrt{f_{\max}})$-grid. By Lemma 5.1(iii), F_δ^- crosses H completely. Let r be the intersection point between F_δ^- and the center line of H. Let C be the $(\lambda_k/\sqrt{f_{\max}})$-cell that contains r. The distance from r to the boundary of H is $14\lambda_k\sqrt{f(\tilde{s})}$. By Lemma 3.5, the diameter of C is at most $11\lambda_k f(\tilde{r})/\sqrt{f_{\max}} \leq 11\lambda_k\sqrt{f(\tilde{r})}$. Since $f(\tilde{r}) \leq 1.1f(\tilde{s})$ by Lemma 4.5(v), the diameter of C is less than $14\lambda_k\sqrt{f(\tilde{s})}$. It follows that C lies inside H.

Let u be the rightmost vertex of C on F_δ^-. Let v be the vertex of C different from u on the normal segment at u. Let x be the intersection point between F_δ^- and the right boundary line of H_1. We are to prove that x lies above C. Since C is non-empty with very high probability, the lower side of *rectangle*(s,θ^*) should intersect F_δ^- below x then. This will allow us to rotate *rectangle*(s,θ^*) to reduce its height, yielding the desired contradiction.

By Claim 1, v is the highest point in C and x is the highest point on $F_\delta^-(p,x)$. Let d_v and d_x be the height of v and x from p, respectively. Let ϕ be the acute angle between pu and the horizontal line through p. Since ϕ is at most the sum of γ_p and the angle between pu and the tangent at p, by Lemma 5.2(iii), we have $\phi \leq \gamma_p + 5\,\text{width}(H)$. By Lemma 5.2(i), $\|p-u\| \leq 3\,\text{width}(H)$. Observe that $d_v \leq \|p-u\|\cdot\sin\phi + \|u-v\|$. So $d_v < 3\phi\,\text{width}(H) + 2\lambda_k\delta < 3\gamma_p\,\text{width}(H) + 15\text{width}(H)^2 + 2\lambda_k\delta$. By (10), we get $d_v < W_s\gamma_p/4 + 5W_s^2/48 + 2\lambda_k\delta$. We bound $2\lambda_k\delta$ as follows. Recall that $W_s = \min\{\sqrt{\text{radius}(\mathit{initial}(s))}, \text{radius}(\mathit{coarse}(s))/3\}$. If $W_s = \sqrt{\text{radius}(\mathit{initial}(s))}$, by Lemma 4.2, $W_s \geq \sqrt{\lambda_h/3}f(\tilde{s})^{1/4} \geq \sqrt{\lambda_h/3}$. So $2\lambda_k\delta < 2\lambda_k = \lambda_h/1512 < 0.002W_s^2$. If $W_s = \text{radius}(\mathit{coarse}(s))/3$, by Lemma 4.4, $W_s \geq 2\sqrt{\rho}\delta/3$ and $W_s \geq \lambda_h\sqrt{f(\tilde{s})}/3 \geq \lambda_h/3$. We get $\lambda_k = \lambda_h/3024 \leq W_s/1008$ and $2\delta \leq 3W_s/\sqrt{\rho} \leq 3W_s/\sqrt{5}$, so $2\lambda_k\delta < 0.002W_s^2$. We conclude that

$$d_v < \frac{W_s\gamma_p}{4} + 0.2W_s^2.$$

Observe that px is parallel to the tangent at some point z on $F_\delta^-(p,x)$. By Lemma 5.2(ii), $\gamma_z \geq \gamma_p - 9\,\text{width}(H_1) = \gamma_p - 3W_s$. Since $d_x = \text{width}(H_1)\cdot\tan\gamma_z = (W_s/3)\cdot\tan\gamma_z$, we get

$$d_x \geq \frac{W_s\gamma_z}{3} \geq \frac{W_s\gamma_p}{3} - W_s^2.$$

Since $\theta_s^* > 23W_s$ by our assumption, Lemma 5.1(iv) implies that $\gamma_p \geq 0.8\theta_s^* - 3W_s > 15W_s$. Therefore, $d_x - d_v > W_s\gamma_p/12 - 1.2W_s^2 > 0$. It follows that x lies above C.

Since C is a $(\lambda_k/\sqrt{f_{\max}})$-cell, by Lemma 3.8(i), C contains some sample with probability at least $1 - n^{\Omega(\ln^\omega n/f_{\max})}$. Thus, the lower side of *rectangle*(s,θ^*) lies below x with probability at least $1 - n^{\Omega(\ln^\omega n/f_{\max})}$. On the other hand, the lower side of *rectangle*(s,θ^*) cannot lie below $F_\delta^- \cap H_1$, otherwise it could be raised to reduce the height of *rectangle*(s,θ^*), contradiction. So the lower side of *rectangle*(s,θ^*) intersects $F_\delta^- \cap H_1$ at some point a. See Figure 15.

Let H_2 be the strip inside *candidate*(s,θ^*) such that H_2 is bounded by the right boundary line of *candidate*(s,θ^*) on the right and $\text{width}(H_2) = W_s/3$. By a symmetric argument, we can prove that the upper side of *rectangle*(s,θ^*) intersects $F_\delta^+ \cap H_2$ at a point b.

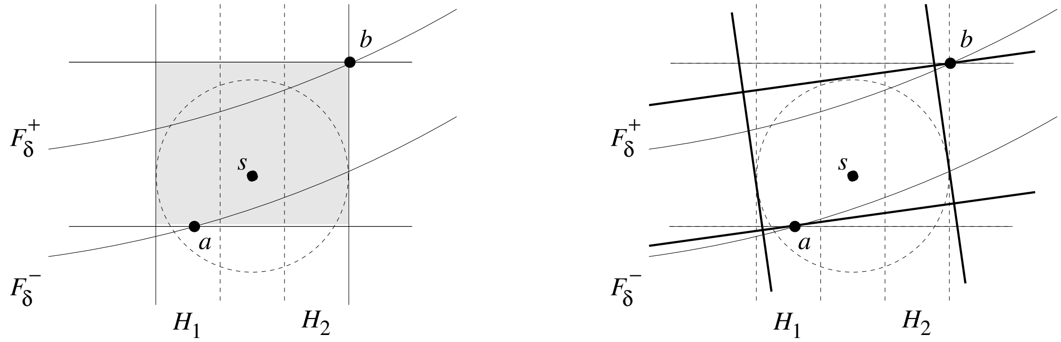


Figure 15:

As shown in Figure 15, we slightly decrease θ from θ^*, i.e., rotate the candidate neighborhood in the anti-clockwise direction. By Lemma 5.1(ii), θ_s can reach zero during the variation of θ. Thus, as $\theta_s^* > 0$, decreasing θ from θ^* is legal. Moreover, as $\theta_s^* > 23W_s$, the small rotation keeps θ_s greater than $23W_s$. Correspondingly, we rotate the lower and upper sides of $\mathit{rectangle}(s, \theta^*)$ around a and b, respectively, to obtain a rectangle R. Orient space such that the new candidate neighborhood becomes vertical. By Claim 1, F_δ^- rises strictly from left to right, so F_δ^- crosses the lower side of R at most once at a from below to above. Similarly, F_δ^+ crosses the upper side of R at most once at b from below to above. This implies that R contains all the samples inside the new candidate neighborhood . Since a is on the left of b and below b, the anti-clockwise rotation makes the height of R strictly less than the height of $\mathit{rectangle}(s, \theta^*)$. This contradicts the assumption that the height of $\mathit{rectangle}(s, \theta^*)$ is already the minimum possible. ◻

Once *refined*(s) is aligned well with the normal at $\tilde{s}$, it is intuitively true that the center point of *refined*(s) should lie very close to $\tilde{s}$. The following lemma proves this formally.

Lemma 5.4 *Assume that* $\delta \leq 1/(25\rho^2)$ *and* $\rho \geq 5$. *Let* s *be a sample. Let* W_s *be the width of refined*(s). *For sufficiently large* n, *the distance between the center point of refined*(s) *and* $\tilde{s}$ *is at most* $(138\delta + 3)W_s$ *with probability at least* $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.

Proof. We first assume that Lemmas 4.2, 4.3, 4.4, 5.1, 5.2, and 5.3 hold deterministically and show that the lemma is true with probability at least $1 - O(n^{\Omega(\ln^\omega n/f_{\max})})$. Since these lemmas hold with probability at least $1 - O(n^{\Omega(\ln^\omega n/f_{\max})})$, the lemma follows.

Assume that s lies on F_α^+ and the normal at $\tilde{s}$ is vertical. Let r_d (resp. r_u) be the ray that shoots downward (resp. upward) from s and makes an angle θ_s^* 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. Let s_1 be the point on F_δ^- such that $\tilde{s_1} = \tilde{s}$. Without loss of generality, we assume that $\theta_s^* \geq 0$. Refer to Figure 16.

First, we bound the distance between the midpoint of xz and y. By Lemma 4.5(iv), the acute angle between s_1z and the tangent at s_1 (the horizontal) is at most $\sin^{-1}(0.03)$. It follows that $\angle ss_1z \leq \pi/2 + \sin^{-1}(0.03)$. So $\angle szs_1 = \pi - \theta_s^* - \angle ss_1z \geq \pi/2 - \theta_s^* - \sin^{-1}(0.03)$, which is greater than 0.9 as $\theta_s^* \leq \pi/5$ by Lemma 5.1(ii). By applying sine law to the shaded triangle in

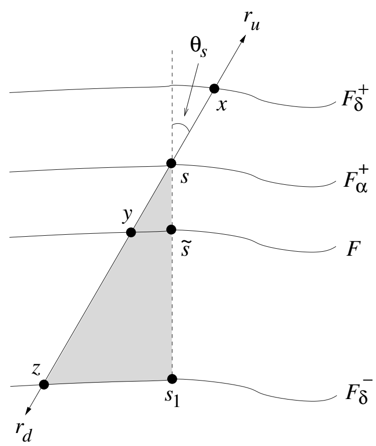


Figure 16: For the proof of Lemma 5.4.

Figure 16, we get

$$\|s_1 - z\| = \frac{\|s - s_1\| \cdot \sin \theta_s^*}{\sin \angle szs_1} \leq \frac{(\delta + \alpha)\theta_s^*}{\sin(0.9)} < 2(\delta + \alpha)\theta_s^*. \tag{11}$$

Similarly, we get

$$\|\tilde{s} - y\| = \frac{\|s - \tilde{s}\| \cdot \sin \theta_s^*}{\sin \angle sys_1} \leq \frac{\alpha\theta_s^*}{\sin(0.9)} < 2\alpha\theta_s^*. \tag{12}$$

By triangle inequality, $\|s - s_1\| - \|s_1 - z\| \leq \|s - z\| \leq \|s - s_1\| + \|s_1 - z\|$. Then (11) yields

$$(\delta + \alpha) - 2(\delta + \alpha)\theta_s^* \leq \|s - z\| \leq (\delta + \alpha) + 2(\delta + \alpha)\theta_s^*. \tag{13}$$

We can use a similar argument to show that

$$(\delta - \alpha) - 2(\delta - \alpha)\theta_s^* \leq \|s - x\| \leq (\delta - \alpha) + 2(\delta - \alpha)\theta_s^*, \tag{14}$$

$$\alpha - 2\alpha\theta_s^* \leq \|s - y\| \leq \alpha + 2\alpha\theta_s^*. \tag{15}$$

Let d_x and d_y be the distances from the midpoint of xz to x and y, respectively. Since $\|x - z\| = \|s - x\| + \|s - z\|$, by (13) and (14), we get $2\delta - 4\delta\theta_s^* \leq \|x - z\| \leq 2\delta + 4\delta\theta_s^*$. Therefore, $\delta - 2\delta\theta_s^* \leq d_x \leq \delta + 2\delta\theta_s^*$. Since $\|x - y\| = \|s - x\| + \|s - y\|$, by (14) and (15), we get $\delta - 2\delta\theta_s^* \leq \|x - y\| \leq \delta + 2\delta\theta_s^*$. We conclude that

$$d_y = |d_x - \|x - y\|| \leq 4\delta\theta_s^*. \tag{16}$$

Second, we bound the distance between the center point s^* of *refined*(s) and y. Although s^* lies on the support line of xz, it may not coincide with the midpoint of xz. There are two cases.

Case 1: the upper side of *refined*(s) lies above x. The upper side of *refined*(s) must intersect $F_\delta^+ \cap$ *candidate*(s, θ^*) at some point v, otherwise we could lower it to reduce the height of *refined*(s), contradiction. Since $\|x - v\| \leq 3W_s$ by Lemma 5.2(i), the distance between x and the upper side of *refined*(s) is at most $3W_s$.

Case 2: the upper side of *refined*(s) lies below x. Let h be the constant in Lemma 4.2. Let $k = h/252$. Take any $(\lambda_k/\sqrt{f_{\max}})$-grid. Let C be the cell that contains x.

We claim that C lies inside *candidate*(s,θ^*). Since radius(*initial*(s)) $\leq \psi_m\sqrt{f(\tilde{s})}$, radius(*initial*$(s)$) < 1 for sufficiently large n. So $\sqrt{\text{radius}(\mathit{initial}(s))} > \text{radius}(\mathit{initial}(s))$. Thus, $W_s = \min\{\sqrt{\text{radius}(\mathit{initial}(s))}, \text{radius}(\mathit{coarse}(s))/3\} \geq \text{radius}(\mathit{initial}(s))/3$, which is at least $\lambda_h\sqrt{f(\tilde{s})}/9$. By Lemma 3.5, the diameter of C is at most $11\lambda_k f(\tilde{x})/\sqrt{f_{\max}} \leq 11\lambda_k\sqrt{f(\tilde{x})}$. Since $f(\tilde{x}) \geq 0.9f(\tilde{s})$ by Lemma 4.5(v), the diameter of C is less than $13\lambda_k\sqrt{f(\tilde{s})}$. Since $W_s \geq \lambda_h\sqrt{f(\tilde{s})}/9 = 28\lambda_k\sqrt{f(\tilde{s})}$, C must lie inside *candidate*(s,θ^*).

Since C is a $(\lambda_k/\sqrt{f_{\max}})$-cell, by Lemma 3.8(i), C contains some sample with probability at least $1 - n^{-\Omega(\ln^\omega n/f_{\max})}$. Thus, the upper side of *refined*(s) cannot lie below C. It follows that the distance between x and the upper side of *refined*(s) is at most the diameter of C, which has been shown to be less than $W_s/2$.

Hence, the position of the upper side of *refined*(s) may cause s^* to be displaced from the midpoint of xz by a distance of at most $3W_s/2$. The position of the lower side of *refined*(s) has the same effect. So the distance between s^* and the midpoint of xz is at most $3W_s$. It follows that $\|s^* - y\| \leq d_y + 3W_s$. By (16), we get $\|s^* - y\| \leq 4\delta\theta_s^* + 3W_s$. Starting with triangle inequality, we obtain

$$\begin{aligned}\|\tilde{s} - s^*\| &\leq \|s^* - y\| + \|\tilde{s} - y\| \\ &\leq 4\delta\theta_s^* + 3W_s + \|\tilde{s} - y\| \\ &\stackrel{(12)}{\leq} 6\delta\theta_s^* + 3W_s.\end{aligned}$$

Since $\theta_s^* \leq 23W_s$ by Lemma 5.3, we conclude that $\|\tilde{s} - s^*\| \leq (138\delta + 3)W_s$. ◻

5.2 More convergence properties and homeomorphism

In this section, we prove more convergence properties which leads to the proof that the output curve of the NN-crust algorithm is homeomorphic to F. For each sample s, we use s^* to denote the center point of *refined*(s). We briefly review the processing of the center points. We first sort the center points in decreasing order of the widths of their corresponding refined neighborhoods. Then we scan the sorted list to select a subset of center points. When the current center point s^* is selected, we delete all center points p^* from the sorted list such that $\|p^* - s^*\| \leq \text{width}(\mathit{refined}(s))^{1/3}$.

In the end, we call two selected center points s^* and t^* *adjacent* if $F(\tilde{s},\tilde{t})$ or $F(\tilde{t},\tilde{s})$ does not contain $\tilde{u}$ for any other selected center point u^*. We use G to denote the polygonal curve that connects adjacent selected center points. Note that the degree of every vertex in G is exactly two. Clearly, if we connect $\tilde{s}$ and $\tilde{t}$ for every pair of adjacent selected center points s^* and t^*, we obtain a polygonal curve G' that is homeomorphic to F. Our goal is to show that the output curve of the NN-crust algorithm is exactly G. Since there is a bijection between G and G', the homeomorphism result follows.

We need to establish several technical lemmas (Lemma 5.5–5.10) before proving the homeomorphism results (Lemma 5.11 and Corollary 5.1). Throughout this section, we assume that width(*refined*(*s*)) < 1 for any sample s, which is true for sufficiently large n. There are a few consequences. First, for any constants $a > b > 0$, width(*refined*(*s*))$^a <$ width(*refined*(*s*))^b. Second, radius(*initial*(*s*)) < 1 which implies that $\sqrt{\text{radius}(\textit{initial}(s))} \geq \text{radius}(\textit{initial}(s))$. Third, $\text{width}(\textit{refined}(s)) = \min\{\sqrt{\text{radius}(\textit{initial}(s))}, \text{radius}(\textit{coarse}(s))/3\} \geq \text{radius}(\textit{initial}(s))/3$.

We first relate the widths of refined neighborhoods for two nearby center points (not necessarily selected).

Lemma 5.5 *Let p^* and q^* be two center points. If $\|\tilde{p} - \tilde{q}\| \leq f(\tilde{p})/2$, there exists a constant $\mu_1 > 0$ such that $W_q \leq \mu_1 f(\tilde{p})\sqrt{W_p}$ with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.*

Proof. We prove the lemma by assuming that Lemma 4.2, 4.3, and 4.4 hold deterministically. The probability bound then follows from the probability bounds in these lemmas. For $i = p$ or q, let $R_i = \text{radius}(\textit{coarse}(i))$ and let $r_i = \text{radius}(\textit{initial}(i))$. The Lipschitz condition implies that $f(\tilde{p})/2 \leq f(\tilde{q}) \leq 3f(\tilde{p})/2$. Let h and m be the constants in Lemma 4.2.

Suppose that $W_p = \sqrt{r_p}$. By Lemma 4.2, we have

$$W_p = \sqrt{r_p} \geq \sqrt{\frac{\lambda_h \sqrt{f(\tilde{p})}}{3}} = \sqrt{\frac{h\lambda_m \sqrt{f(\tilde{p})}}{3m}}.$$

Note that $W_q \leq \sqrt{r_q}$ and $r_q \leq \sqrt{11\lambda_m f(\tilde{q})}$ by Lemma 4.2. So we get

$$W_p \geq \sqrt{\frac{h\sqrt{f(\tilde{p})}}{33mf(\tilde{q})}} \cdot r_q \geq \sqrt{\frac{2h}{99m\sqrt{f(\tilde{p})}}} \cdot W_q^2 \geq \sqrt{\frac{2h}{99m}} \cdot \frac{W_q^2}{f(\tilde{p})}.$$

Suppose that $W_p = R_p/3$. First, since $R_p \geq 2\sqrt{\rho}\delta$ by Lemma 4.4, we get $\rho\delta \leq 3\sqrt{\rho}W_p/2$. Second, $W_p \geq R_p/3 \geq r_p/3$ which is at least $\lambda_h\sqrt{f(\tilde{p})}/9$ by Lemma 4.2. So we get $\sqrt{\lambda_m f(\tilde{p})} = \sqrt{m\lambda_h f(\tilde{p})/h} \leq 3\sqrt{mW_p/h} \cdot f(\tilde{p})^{1/4} \leq 3\sqrt{mW_p/h} \cdot f(\tilde{p})$. Finally, since $W_q \leq R_q/3$, by Lemma 4.3, we get

$$\begin{aligned} W_q &\leq \frac{5\rho\delta}{3} + \frac{\sqrt{11\lambda_m f(\tilde{q})}}{3} \\ &\leq \frac{5\rho\delta}{3} + \sqrt{\frac{11\lambda_m f(\tilde{p})}{6}} \\ &\leq \frac{5\sqrt{\rho}W_p}{2} + \sqrt{\frac{33mW_p}{2h}} \cdot f(\tilde{p}). \end{aligned}$$

□

The next result shows that the selected center points cannot be too close to each other.

Lemma 5.6 *Let p^* and q^* be two selected center points. Then $\|p^* - q^*\| > \max\{W_p^{1/3}, W_q^{1/3}\}$.*

Proof. Assume without loss of generality that p^* was selected before q^*. Since q^* was selected subsequently, q^* was not eliminated by the selection of p^*. Thus, $\|p^* - q^*\| > W_p^{1/3} \geq W_q^{1/3}$. □

Next, we bound the angle between x^*y^* and $\tilde{x}\tilde{y}$ and the angle $\angle x^*y^*z^*$ for three center points x^*, y^*, and z^*.

Lemma 5.7 *Let x^* and y^* be two center points such that $\|\tilde{x} - \tilde{y}\| \leq f(\tilde{y})/2$ and $\|x^* - y^*\| \geq W_y^{1/3}$. Then the acute angle between x^*y^* and $\tilde{x}\tilde{y}$ tends to zero as n tends to ∞ with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.*

Proof. We prove the lemma by assuming that Lemmas 5.4 and 5.5 hold deterministically. The probability bound then follows from the probability bounds in these lemmas.

We translate x^*y^* to align y^* with $\tilde{y}$. Let z denote the point $x^* + \tilde{y} - y^*$. Let $k = 138\delta + 3$. By triangle inequality and Lemma 5.4, $\|\tilde{x} - z\| \leq \|x^* - \tilde{x}\| + \|y^* - \tilde{y}\| \leq kW_x + kW_y$. Since $\|\tilde{x} - \tilde{y}\| \leq f(\tilde{y})/2$, by Lemma 5.5, $W_x \leq \mu_1 f(\tilde{y})\sqrt{W_y}$. So $\|\tilde{x} - z\| \leq k\mu_1 f(\tilde{y})\sqrt{W_y} + kW_y$, which is smaller than $W_y^{1/3} \leq \|x^* - y^*\|$ for sufficiently large n. Thus, $\tilde{x}z$ is not the longest side of the triangle $\tilde{x}\tilde{y}z$. It follows that $\angle\tilde{x}\tilde{y}z$ is acute. Since $\|\tilde{x} - z\|$ is an upper bound on the height of z from $\tilde{x}\tilde{y}$, we have $\angle\tilde{x}\tilde{y}z \leq \sin^{-1}\frac{\|\tilde{x}-z\|}{\|\tilde{y}-z\|} = \sin^{-1}\frac{\|\tilde{x}-z\|}{\|x^*-y^*\|} \leq \sin^{-1}(k\mu_1 f(\tilde{y})W_y^{1/6} + kW_y^{2/3})$. We conclude that $\angle\tilde{x}\tilde{y}z$ tends to zero as n tends to ∞. $\square$

Lemma 5.8 *Let x^*, y^*, and z^* be three center points such that $\tilde{y} \in F(\tilde{x}, \tilde{z})$, $\|\tilde{x} - \tilde{z}\| \leq \max\{f(\tilde{x})/4, f(\tilde{z})/4\}$, $\|x^* - y^*\| \geq W_y^{1/3}$, and $\|y^* - z^*\| \geq W_y^{1/3}$. For sufficiently large n, the angle $\angle x^*y^*z^*$ is obtuse with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.*

Proof. We first show that $\|\tilde{x} - \tilde{z}\| \leq \min\{f(\tilde{x})/3, f(\tilde{z})/3\}$. Assume that $\|\tilde{x} - \tilde{z}\| \leq f(\tilde{x})/4$. By the Lipschitz condition, we have $f(\tilde{z}) \geq 3f(\tilde{x})/4$. Therefore, $\|\tilde{x} - \tilde{z}\| \leq f(\tilde{x})/4 \leq f(\tilde{z})/3$.

Let D be the disk centered at $\tilde{x}$ with radius $f(\tilde{x})/3$. Observe that $F(\tilde{x}, \tilde{z})$ lies completely inside D. Otherwise, the medial axis of F intersects the interior of D which implies that $f(\tilde{x}) \leq f(\tilde{x})/3$, contradiction. So $\|\tilde{x} - \tilde{y}\| \leq f(\tilde{x})/3$. The Lipschitz condition implies that $f(\tilde{y}) \geq 2f(\tilde{x})/3$.

We claim that the angle $\angle\tilde{x}\tilde{y}\tilde{z}$ is obtuse. The line segments $\tilde{x}\tilde{y}$ and $\tilde{y}\tilde{z}$ are parallel to the tangents at some points on $F(\tilde{x}, \tilde{y})$ and $F(\tilde{y}, \tilde{z})$, respectively. Lemma 3.3 implies that $\angle\tilde{x}\tilde{y}\tilde{z} \geq \pi - 4\sin^{-1}\frac{\text{radius}(D)}{f(\tilde{x})} = \pi - 4\sin^{-1}(1/3) > \pi/2$.

Since $\|\tilde{x} - \tilde{y}\| \leq f(\tilde{x})/3 \leq f(\tilde{y})/2$, by Lemma 5.7, the angle between x^*y^* and $\tilde{x}\tilde{y}$ tends to zero as n tends to ∞ with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$. A symmetric argument shows that the angle between y^*z^* and $\tilde{y}\tilde{z}$ tends to zero with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$ as n tends to ∞. Thus, $\angle x^*y^*z^*$ converges to $\angle\tilde{x}\tilde{y}\tilde{z}$ which is obtuse. $\square$

The next lemma provides an upper bound on the the edge lengths in G.

Lemma 5.9 *Let e be an edge in G connecting two center points p^* and q^*. For sufficiently large n, there exists a constant $\mu_2 > 0$ such that* $\text{length}(e) \leq \mu_2 f(\tilde{p})W_p^{1/3} + \mu_2 f(\tilde{q})W_q^{1/3}$ *with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.*

Proof. Note that p^* and q^* are adjacent and they are selected by the algorithm. Let $k = 138\delta+3$. Let D_p be the disk centered at p^* with radius $(1 + k\mu_1 f(\tilde{p}))W_p^{1/3}$. Let D_q be the disk centered at q^* with radius $(1 + k\mu_1 f(\tilde{q}))W_q^{1/3}$. By Lemma 5.4, $\|\tilde{p} - p^*\| \leq kW_p$ which is less than $W_p^{1/3}$ for sufficiently large n. So $\tilde{p}$ lies inside D_p. Similarly, $\tilde{q}$ lies inside D_q.

If D_p intersects D_q, then $\|p^* - q^*\| \leq (1 + \mu_1 f(\tilde{p}))W_p^{1/3} + (1 + \mu_1 f(\tilde{q}))W_q^{1/3}$ and we are done. Suppose that D_p does not intersect D_q. We claim that $F(\tilde{p}, \tilde{q}) \cap D_p$ is connected. Otherwise, the medial axis of F intersects the interior of D_p which implies that $f(\tilde{p}) \leq \text{radius}(D_p)$ which is less than $f(\tilde{p})$ for sufficiently large n, contradiction. Similarly, $F(\tilde{p}, \tilde{q}) \cap D_q$ is connected. It follows that $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ is also connected. There are two cases.

Case 1: $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ does not contain $\tilde{u}$ for any sample u. Let y be the endpoint of $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ that lies on D_p. Let h be the constant in Lemma 4.2. Take a λ_h-partition such that y is the first cut-point. Since $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ does not contain $\tilde{u}$ for any sample u, by Lemma 3.8(i), $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ does not contain $F(y, c_1)$, where c_1 is the second cut-point, with probability at least $1 - O(n^{-\Omega(\ln^\omega n)})$. It follows that

$$|F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)| < \lambda_h^2 f(y). \tag{17}$$

Since $\|\tilde{p} - y\| \leq 2\,\text{radius}(D_p) = 2(1 + k\mu_1 f(\tilde{p}))W_p^{1/3}$, $\|\tilde{p} - y\| \leq f(\tilde{p})/2$ for sufficiently large n. Thus, $f(y) \leq 3f(\tilde{p})/2$, so $\lambda_h^2 f(y) < 3\lambda_h^2 f(\tilde{p})/2$. Since $W_p \geq \text{radius}(\mathit{initial}(p))/3$ which is at least $\lambda_h\sqrt{f(\tilde{p})}/9$ by Lemma 4.2, we have $\lambda_h^2 f(\tilde{y}) \leq 243W_p^2/2$. Substituting into (17), we get

$$|F(\tilde{p}, \tilde{q})| \leq 243W_p^2/2 + 2\,\text{radius}(D_p) + 2\,\text{radius}(D_q).$$

By Lemma 5.4, $\|\tilde{p} - p^*\| \leq kW_p$ and $\|\tilde{q} - q^*\| \leq kW_q$. We conclude that $\|p^* - q^*\| \leq \|\tilde{p} - p^*\| + |F(\tilde{p}, \tilde{q})| + \|\tilde{q} - q^*\| \leq \mu_2 f(\tilde{p})W_p^{1/3} + \mu_2 f(\tilde{q})W_q^{1/3}$ for some constant $\mu_2 > 0$.

Case 2: $F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$ contains $\tilde{u}$ for some sample u. We show that this case is impossible if Lemmas 5.5 and 5.8 hold deterministically. It follows that case 2 occurs with probability at most $O(n^{-\Omega(\ln^\omega n/f_{\max})})$. We first claim that $\|p^* - u^*\| > W_p^{1/3}$. If not, Lemma 5.5 implies that $W_u \leq \mu_1 f(\tilde{p})\sqrt{W_p}$ for sufficiently large n. But then $\|p^* - \tilde{u}\| \leq \|p^* - u^*\| + \|\tilde{u} - u^*\| \leq W_p^{1/3} + kW_u \leq W_p^{1/3} + k\mu_1 f(\tilde{p})\sqrt{W_p}$. This is a contradiction as $\tilde{u}$ lies outside D_p. Similarly, $\|q^* - u^*\| > W_q^{1/3}$. So u^* is not eliminated by the selection of p^* and q^*.

Next, take any selected center point z^* different from p^* and q^* such that $\tilde{q} \in F(\tilde{u}, \tilde{z})$. We show that u^* is not eliminated by the selection of z^*. Assume to the contrary that this is false. So $\|u^* - z^*\| \leq W_z^{1/3}$. By Lemma 5.5, $W_u \leq \mu_1 f(\tilde{z})\sqrt{W_z}$ for sufficiently large n. Let $k' = 1 + k + k\mu_1$. Then $\|\tilde{u} - \tilde{z}\| \leq \|u^* - z^*\| + \|z^* - \tilde{z}\| + \|u^* - \tilde{u}\| \leq W_z^{1/3} + kW_z + kW_u \leq W_z^{1/3} + kW_z + k\mu_1 f(\tilde{z})\sqrt{W_z} \leq k' f(\tilde{z})W_z^{1/3}$. For sufficiently large n, $k' f(\tilde{z})W_z^{1/3} \leq f(\tilde{z})/4$. By Lemma 5.8, the angle $\angle u^* q^* z^*$ is obtuse. It follows that $\|q^* - z^*\| < \|u^* - z^*\| \leq W_z^{1/3}$, contradicting Lemma 5.6.

Symmetrically, we can show that u^* is not eliminated by any selected center point z^* different from p^* and q^* such that $\tilde{p} \in F(\tilde{z}, \tilde{u})$. In all, our algorithm should select another center point u^* such that $\tilde{u} \in F(\tilde{p}, \tilde{q}) - (D_p \cup D_q)$. This contradicts the assumption that p^* and q^* are adjacent in G.

$\sqcup\!\!\!\sqcap$

We are ready to show that the output curve of the NN-crust algorithm is exactly G. This will allow us to show that the output curve is homeomorphic to the underlying smooth closed curve.

Lemma 5.10 *Let p^* and q^* be two selected center points that are not adjacent in G. For sufficiently large n, if $\|p^* - q^*\| \leq f(\tilde{p})/4$, there is an edge e in G incident to p^* such that the angle between e and p^*q^* is acute and* $\text{length}(e) < \|p^* - q^*\|$ *with probability at least* $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$.

Proof. Since p^* and q^* are not adjacent in G, there is some selected center point u^* adjacent to p^* such that $\tilde{u}$ lies on $F(\tilde{p}, \tilde{q})$ or $F(\tilde{q}, \tilde{p})$, say $F(\tilde{p}, \tilde{q})$. By Lemma 5.6, $\|p^* - u^*\| > W_u^{1/3}$ and $\|q^* - u^*\| > W_u^{1/3}$. By Lemma 5.8, the angle $\angle p^*u^*q^*$ is obtuse with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$. It follows that $\angle u^*p^*q^*$ is acute and $\|p^* - u^*\| < \|p^* - q^*\|$. $\sqcup\!\!\!\sqcap$

Lemma 5.11 *For sufficiently large n, the output curve obtained by running the NN-crust algorithm on the selected center points is exactly G with probability at least* $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$.

Proof. We first prove the lemma assuming that Lemmas 5.4, 5.8, 5.9, and 5.10 hold deterministically. We will discuss the probability bound later.

Let p^* be a selected center point. Let p^*u^* and p^*v^* be the edges of G incident to p^*. Without loss of generality, we assume that $\tilde{p}$ lies on $F(\tilde{u}, \tilde{v})$. By Lemma 5.6, $\|p^* - u^*\| > W_p^{1/3}$ and $\|p^* - v^*\| > W_p^{1/3}$.

By Lemmas 5.4 and 5.9, $\|\tilde{p} - \tilde{u}\| \leq \|\tilde{p} - p^*\| + \|\tilde{u} - u^*\| + \|p^* - u^*\| \leq kW_p + kW_u + \mu_2 f(\tilde{p})W_p^{1/3} + \mu_2 f(\tilde{u})W_u^{1/3}$, which is less than $(f(\tilde{p}) + f(\tilde{u}))/30$ for sufficiently large n. The Lipschitz condition implies that

$$0.9f(\tilde{p}) < f(\tilde{u}) < 1.1f(\tilde{p}).$$

So we get

$$\|\tilde{p} - \tilde{u}\| \leq \frac{f(\tilde{p}) + f(\tilde{u})}{30} < 0.1f(\tilde{p}), \qquad \|p^* - u^*\| \leq \frac{f(\tilde{p}) + f(\tilde{u})}{30} < 0.1f(\tilde{p}).$$

Similarly, we can show that

$$\|\tilde{p} - \tilde{v}\| < 0.1f(\tilde{p}), \qquad \|p^* - v^*\| < 0.1f(\tilde{p}).$$

Let p^*q^* be an edge computed by the NN-crust algorithm when it processes the vertex p^*. Assume to the contrary that p^*q^* is not an edge in G. If p^*q^* is computed in step 1 of the NN-crust algorithm, then q^* is the nearest neighbor of p^*. So $\|p^* - q^*\| \leq \|p^* - u^*\| < 0.1f(\tilde{p})$. By Lemma 5.10, there is another edge e in G such that $\text{length}(e) < \|p^* - q^*\|$, contradiction. Suppose that p^*q^* is computed in step 2 of the NN-crust algorithm. As we have just shown, the step 1 of the NN-crust algorithm already outputs an edge, say p^*u^*, of G where u^* is the nearest neighbor of p^*. Observe that $\|\tilde{u} - \tilde{v}\| \leq \|\tilde{p} - \tilde{u}\| + \|\tilde{p} - \tilde{v}\| < 0.2f(\tilde{p}) < 0.25f(\tilde{u})$. By Lemma 5.8, $\angle u^*p^*v^*$ is obtuse. By the NN-crust algorithm, $\angle u^*p^*q^*$ is also obtuse. Since the NN-crust algorithm prefers p^*q^* to p^*v^*, $\|p^* - q^*\| \leq \|p^* - v^*\| < 0.1f(\tilde{p})$. By Lemma 5.10, G has an edge e incident to p^* that is shorter than p^*q^* (p^*v^* too) and makes an acute angle with p^*q^*. The edge e is not p^*v^* as e is shorter than p^*v^*. The edge e is not p^*u^* as $\angle u^*p^*q^*$ is obtuse. But then the degree of p in G is at least three, contradiction.

We have shown that each output edge belongs to G. Since the NN-crust algorithm guarantees that each vertex in the output curve has degree at least two, the output curve and G have the same number of edges. So the output curve is exactly G.

Since Lemmas 5.4, 5.8, 5.9, and 5.10 hold with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$, the output edges incident to p^* are edges of G with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$. Since there are $O(n)$ output vertices, the probability that this holds for all vertices is at least $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$. ∎

Corollary 5.1 *For sufficiently large n, the output curve obtained by running the NN-crust algorithm on the selected center points is homeomorphic to the underlying smooth closed curve with probability at least $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$.*

Proof. We have shown that the output curve is G. Let G' be the curve obtained by connecting $\tilde{p}$ and $\tilde{q}$ for each edge p^*q^* of G. G' is homeomorphic to the underlying smooth closed curve as p^* and q^* are adjacent in G. Clearly, G is homeomorphic to G' as there is a bijection between the edges of G and G'. ∎

5.3 Main theorem

We make use of the results in the previous subsections to prove the main theorem in this paper.

Theorem 5.1 *Assume that $\delta \leq 1/(25\rho^2)$ and $\rho \geq 5$. Given n noisy samples from a smooth closed curve, when n is sufficiently large, our algorithm computes a polygonal closed curve that satisfies the following properties with probability at least $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$:*

- *Each output vertex s^* converges to $\tilde{s}$.*
- *For each output edge r^*s^*, its slope converges to the slope of the tangent at $\tilde{s}$.*
- *The output curve is homeomorphic to the smooth closed curve.*

Proof. By Lemma 5.4, an output vertex s^* converges to $\tilde{s}$ with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$. Since there are $O(n)$ output vertices, the pointwise convergence occurs with probability at least $1 - O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}} - 1)})$. Next, we analyze the angular differences between the tangents of the smooth closed curve and the output curve.

Let r^*s^* be an output edge. By Lemma 5.9, with probability at least $1 - O(n^{-\Omega(\ln^\omega n/f_{\max})})$, we have

$$\|r^* - s^*\| \leq \mu_2 f(\tilde{r}) W_r^{1/3} + \mu_2 f(\tilde{s}) W_s^{1/3}. \quad (18)$$

Using the above, the triangle inequality, and Lemma 5.4, we get

$$\begin{aligned} \|\tilde{r}-\tilde{s}\| &\leq \|\tilde{r}-r^*\| + \|\tilde{s}-s^*\| + \|r^*-s^*\| && (19) \\ &\leq kW_r + kW_s + \mu_2 f(\tilde{r})W_r^{1/3} + \mu_2 f(\tilde{s})W_s^{1/3}. && (20) \end{aligned}$$

By (18), $\|r^*-s^*\| < f(\tilde{r})/5+f(\tilde{s})/5$ for sufficiently large n. The Lipschitz condition implies that $f(\tilde{r}) < 1.5f(\tilde{s})$. So $\|r^*-s^*\| < f(\tilde{s})/2$. Thus, Lemma 5.5 applies and yields $W_r \leq \mu_1 f(\tilde{s})\sqrt{W_s}$ with probability at least $1-O(n^{-\Omega(\ln^\omega n/f_{\max})})$. Substituting into (20), we conclude that

$$\|\tilde{r}-\tilde{s}\| \leq \mu_3 f(\tilde{s})^{4/3} W_s^{1/6}, \qquad (21)$$

for some constant $\mu_3 > 0$.

Let θ be the angle between $\tilde{r}\tilde{s}$ and the tangent at $\tilde{s}$. By Lemma 3.2(ii), we have $\theta \leq \sin^{-1}\frac{\mu_3 f(\tilde{s})^{1/3}W_s^{1/6}}{2}$. Let θ' be the acute angle between r^*s^* and $\tilde{r}\tilde{s}$. By (21), $\|\tilde{r}-\tilde{s}\| \leq f(\tilde{s})/2$ for sufficiently large n. Thus, by Lemma 5.7, θ' tends to zero as n tends to ∞ with probability at least $1-O(n^{-\Omega(\ln^\omega n/f_{\max})})$. We conclude that $\theta+\theta'$ tends to zero as n tends to ∞, so the slope of r^*s^* converges to the slope of the tangent at $\tilde{s}$. Since there are $O(n)$ output edges, the convergence of their slopes occur with probability at least $1-O(n^{-\Omega(\frac{\ln^\omega n}{f_{\max}}-1)})$.

The output curve is homeomorphic to the smooth closed curve by Corollary 5.1. □

6 Conclusion

We have presented an algorithm to reconstruct polygonal closed curves from noisy samples drawn from a set of smooth closed curves. Although we have assumed that there is only one smooth closed curve in our analysis for notational simplicity, the analysis can be carried over to the general case. A straightforward implementation of our algorithm takes $O(n^3)$ time. We view the analysis as our major contribution as it is the first result that deals with faithful curve reconstruction from noisy samples. Since the analysis is already quite involved, we did not spend much effort in looking for a faster algorithm.

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 in the normal directions. It would be interesting to investigate if other noise models are amenable to analysis.

We expect that our approach will also help in reconstructing curves with features such as corners, branchings and terminals (with or without noise). Another research direction is to study the reconstruction of surfaces from noisy samples.

References

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

[2] 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.

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

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

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

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

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

[8] 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.

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

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