Voronoi Diagram of Polygonal Chains Under the Discrete Fréchet Distance

Sergey Bereg [*] Kevin Buchin [†] Maike Buchin [†] Marina Gavrilova [‡]
Binhai Zhu [§]

Abstract

Polygonal chains are fundamental objects in many applications like pattern recognition and protein structure alignment. A well-known measure to characterize the similarity of two polygonal chains is the famous (continuous/discrete) Fréchet distance. In this paper, for the first time, we consider the Voronoi diagram of polygonal chains in d-dimension under the discrete Fréchet distance. Given a set $\mathcal{C}$ of n polygonal chains in d-dimension, each with at most k vertices, we prove fundamental properties of such a Voronoi diagram $VD_F(\mathcal{C})$. Our main results are summarized as follows.

- The combinatorial complexity of $VD_F(\mathcal{C})$ is at most $O(n^{dk+\epsilon})$.
- The combinatorial complexity of $VD_F(\mathcal{C})$ is at least $\Omega(n^{dk})$ for dimension $d = 1, 2$; and $\Omega(n^{d(k-1)+2})$ for dimension $d > 2$.

Keywords: Voronoi diagram, Fréchet distance, discrete Fréchet distance, combinatorial complexity, protein structure alignment

[*]Department of Computer Science, University of Texas at Dallas, Richardson, TX 75083, USA. Email: besp@utdallas.edu.

[†]Institute of Computer Science, Freie Universität Berlin, Germany. Email: buchin,mbuchin@inf.fu-berlin.de.

[‡]Department of Computer Science, University of Calgary, Calgary, Alberta T2N 1N4, Canada. Email: marina@cpsc.ucalgary.ca.

[§]Department of Computer Science, Montana State University, Bozeman, MT 59717-3880, USA. Email: bhz@cs.montana.edu.

November 10, 2007

1 Introduction

The Fréchet distance was first defined by Maurice Fréchet in 1906 [8]. While known as a famous distance measure in the field of mathematics (more specifically, abstract spaces), it was first applied in measuring the similarity of polygonal curves by Alt and Godau in 1992 [1]. In general, the Fréchet distance between 2D polygonal chains (polylines) can be computed in polynomial time [1, 2], even under translation or rotation (though the running time is much higher) [3, 16]. While computing (approximating) Fréchet distance for surfaces is NP-hard [9], it is polynomially solvable for restricted surfaces [5].

In 1994, Eiter and Mannila defined the *discrete Fréchet distance* between two polygonal chains A and B in d-dimension. This simplified distance is always realized by two vertices in A and B [7]. They showed that with dynamic programming the discrete Fréchet distance between polygonal chains A and B can be computed in $O(|A||B|)$ time. In [10], Indyk defined a similar discrete Fréchet distance in some metric space and showed how to compute approximate nearest neighbors using that distance.

Recently, Jiang, Xu and Zhu applied the discrete Fréchet distance in aligning the backbones of proteins (which are called the *protein structure-structure alignment* problem [11] and *protein local structure alignment* respectively [17]). In fact, in these applications the discrete Fréchet distance makes more sense as the backbone of a protein is simply a polygonal chain in 3D, with each vertex being the alpha-carbon atom of a residue. So if the (continuous) Fréchet distance is realized by an alpha-carbon atom and some other point which does not represent an atom, it is not meaningful biologically. Jiang, *et al.* showed that given two planar polygonal chains the minimum discrete Fréchet distance between them, under both translation and rotation, can be computed in polynomial time. They also applied some ideas therein to design an efficient heuristic for the original protein structure-structure alignment problem in 3D and the empirical results showed that their alignment is more accurate compared with previously known solutions.

On the other hand, a lot is still unknown regarding the discrete/continuous Fréchet distance. For instance, even though the Voronoi diagram has been studied for many objects and distance measures, it has not yet been studied for polygonal chains under the discrete/continuous Fréchet distance. This problem is fundamental, it has potential applications, e.g., in protein structure alignment, especially with the ever increasing computational power. Imagine that we have some polylines A_1, A_2, A_3... in space. If we can construct the Voronoi diagram for $A_1, A_2, A_3, ...$ in space, then given a new polyline B we can easily compute all the approximate alignments of B with the A_i's. The movement of B defines a subspace (each point in the subspace represents a copy of B) and if we sample this subspace evenly then all we need to do is to locate all these sample points in the Voronoi diagram for the A_i's.

Unfortunately, nothing is known about the Voronoi diagram under the discrete/continuous

Fréchet distance, even for the simplest case of line segments. In this paper, we will present the first set of such results by proving some fundamental properties for both the general case and some special case. We believe that these results will be essential for us to design efficient algorithms for computing/approximating the Voronoi diagram under the Fréchet distance. In this paper, all of our results are under the discrete Fréchet distance unless otherwise specified.

2 Preliminaries

Given two polygonal chains A, B with $|A| = k$ and $|B| = l$ vertices respectively, we aim at measuring the similarity of A and B (possibly under translation and rotation) such that their distance is minimized under certain measure. Among the various distance measures, the Hausdorff distance is known to be better suited for matching two point sets than for matching two polygonal chains; the (continuous) Fréchet distance is a superior measure for matching two polygonal chains, but it is not quite easy to compute [1].

Let X be the Euclidean plane $\mathbb{R}^d$; let $d(a, b)$ denote the Euclidean distance between two points $a, b \in X$. The (continuous) Fréchet distance between two parametric curves $f : [0, 1] \to X$ and $g : [0, 1] \to X$ is

$$\delta_{\mathcal{F}}(f, g) = \inf_{\alpha,\beta} \max_{s\in[0,1]} d(f(\alpha(s)), g(\beta(s))),$$

where α and β range over all continuous non-decreasing real functions with $\alpha(0) = \beta(0) = 0$ and $\alpha(1) = \beta(1) = 1$.

Imagine that a person and a dog walk along two different paths while connected by a leash; they always move forward, though (possibly) at different paces. The minimum possible length of the leash is the Fréchet distance between the two paths. To compute the Fréchet distance between two polygonal curves A and B (in the Euclidean plane) of $|A|$ and $|B|$ vertices, respectively, Alt and Godau [1] presented an $O(|A||B| \log^2(|A||B|))$ time algorithm. Later this bound was reduced to $O(|A||B| \log(|A||B|))$ time [2].

We now define the discrete Fréchet distance following [7].

Definition 2.1 *Given a polygonal chain (polyline) in d-dimension $P = \langle p_1, \ldots, p_k\rangle$ of k vertices, a m-**walk** along P partitions the path into m disjoint non-empty subchains $\{\mathcal{P}_i\}_{i=1..m}$ such that $\mathcal{P}_i = \langle p_{k_{i-1}+1}, \ldots, p_{k_i}\rangle$ and $0 = k_0 < k_1 < \cdots < k_m = k$.*

*Given two polylines in d-dimension $A = \langle a_1, \ldots, a_k\rangle$ and $B = \langle b_1, \ldots, b_l\rangle$, a **paired walk** along A and B is a m-walk $\{\mathcal{A}_i\}_{i=1..m}$ along A and a m-walk $\{\mathcal{B}_i\}_{i=1..m}$ along B for some m, such that, for $1 \le i \le m$, $|\mathcal{A}_i| = 1$ or $|\mathcal{B}_i| = 1$ (that is, $\mathcal{A}_i$ or $\mathcal{B}_i$ contains exactly one vertex). The **cost** of a paired walk $W = \{(\mathcal{A}_i, \mathcal{B}_i)\}$ along two paths A and B is*

$$d_F^W(A, B) = \max_i \max_{(a,b)\in\mathcal{A}_i\times\mathcal{B}_i} d(a, b).$$

The ***discrete Fréchet distance*** *between two polylines A and B in d-dimension is*

$$d_F(A,B) = \min_W d_F^W(A,B).$$

The paired walk that achieves the discrete Fréchet distance between two paths A and B is also called the ***Fréchet alignment*** *of A and B.*

Consider the scenario in which the person walks along A and the dog along B. Intuitively, the definition of the paired walk is based on three cases:

1. $|\mathcal{B}_i| > |\mathcal{A}_i| = 1$: the person stays and the dog moves forward;
2. $|\mathcal{A}_i| > |\mathcal{B}_i| = 1$: the person moves forward and the dog stays;
3. $|\mathcal{A}_i| = |\mathcal{B}_i| = 1$: both the person and the dog move forward.

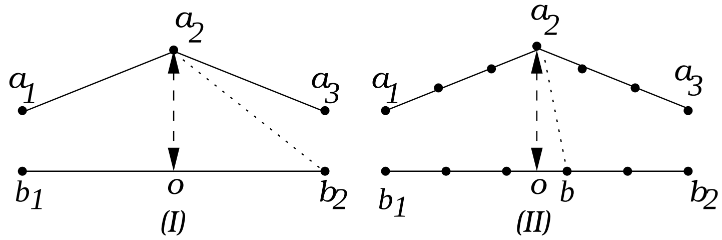

Figure 1: The relationship between discrete and continuous Fréchet distances.

Eiter and Mannila presented a simple dynamic programming algorithm to compute $d_F(A,B)$ in $O(|A||B|) = O(kl)$ time [7]. Recently, Jiang *et al.* showed that in 2D the minimum discrete Fréchet distance between A and B under translation can be computed in $O(k^3l^3\log(k+l))$ time, and under both translation and rotation it can be computed in $O(k^4l^4\log(k+l))$ time [11]. They are significantly faster than the corresponding bounds for the continuous Fréchet distance. In 2D, for the continuous Fréchet distance, under translation, the current fastest algorithm for computing the minimum (continuous) Fréchet distance between A, B takes $O((kl)^3(k+l)^2\log(k+l))$ time [3]; under both translation and rotation, the bound is $O((k+l)^{11}\log(k+l))$ time [16].

We comment that while the discrete Fréchet distance could be arbitrarily larger than the corresponding continuous Fréchet distance (e.g., in Figure 1 (I), they are $d(a_2, b_2)$ and $d(a_2, o)$ respectively), by adding sample points on the polylines, one can easily obtain a close approximation of the continuous Fréchet distance using the discrete Fréchet distance (e.g., one can use $d(a_2, b)$ in Figure 1 (II) to approximate $d(a_2, o)$). This fact has been pointed out in [7, 10]. Moreover, the discrete Fréchet distance is a more natural measure for matching the geometric shapes of biological sequences such as proteins.

In the remaining part of this paper, for the first time, we investigate the Voronoi diagram of a set of polygonal chains (polylines) in d-dimension. While the Voronoi diagram is a central structure in geometric computing and has been widely studied [13], it still attracts a lot of attention recently [4, 12] (including a new annual International Symposium on Voronoi Diagrams in Science and Engineering). We hope that our work will facilitate the understanding of the continuous and discrete Fréchet distance and further enable their applications in various areas, like pattern recognition and computational biology.

The Voronoi diagram of polygonal curves can be represented using the correspondence

$$\begin{aligned} \text{polygonal curve with } k \text{ vertices in } \mathbb{R}^d &\leftrightarrow \text{point in } \mathbb{R}^{dk} \\ \langle (x_{11},\ldots,x_{1d}),\ldots,(x_{k1},\ldots,x_{kd})\rangle &\leftrightarrow (x_{11},\ldots,x_{1d},\ldots,x_{k1},\ldots,x_{kd}). \end{aligned}$$

In the following, we always use the parameters n, d, and k to denote n input curves in $\mathbb{R}^d$, each with at most k vertices. We give upper and lower bounds on the combinatorial complexity of the Voronoi diagram of the input curves, which is a partition of the space in $\mathbb{R}^{dk}$ into Voronoi regions associated with each input curve. By a Voronoi region we mean a set of curves with a common set of nearest neighbors under the discrete Fréchet distance in the given set of input curves.

3 The Combinatorial Upper Bound of $\boldsymbol{VD}_F(\mathcal{C})$

In this section, we prove the combinatorial upper bound of $VD_F(\mathcal{C})$. We first show the case for $d = 2$, and then we sketch how to generalize the result to any fixed d-dimension.

3.1 The Combinatorial Upper Bound of $\boldsymbol{VD}_F(\mathcal{C})$ for $d = 2$

Let $A_k = \langle a_1, a_2, \ldots, a_k\rangle$ and $B_l = \langle b_1, b_2, \ldots, b_l\rangle$ be two polygonal chains in the plane where $a_i = (x(a_i), y(a_i)), b_j = (x(b_j), y(b_j))$ and $k, l \geq 1$. We first have the following lemma, which is easy to prove.

Lemma 3.1 *Let $A_2 = \langle a_1, a_2\rangle$ and $B_2 = \langle b_1, b_2\rangle$ be two line segments in the plane, then*

$$d_F(A_2, B_2) = \max(d(a_1, b_1), d(a_2, b_2)).$$

For general polylines, we can generalize the above lemma as follows. Notice that as $d_F()$ is a min-max-max measure and as we will be using the fact that the Voronoi diagram is a minimization of distance functions, the following lemma is essential.

Lemma 3.2 *Let $A_k = \langle a_1, a_2, \ldots, a_k\rangle$ and $B_l = \langle b_1, b_2, \ldots, b_l\rangle$ be two polygonal chains in the plane where $k, l \geq 1$. The discrete Fréchet distance between A_k and B_l can be computed as*

$$d_F(A_k, B_l) = \begin{cases} \max\{d(a_i, b_1), i = 1, 2, \ldots, k\} & \text{if } l = 1, \\ \max\{d(a_1, b_j), j = 1, 2, \ldots, l\} & \text{if } k = 1, \\ \max\{d(a_k, b_l), \min(d_F(A_{k-1}, B_{l-1}), d_F(A_k, B_{l-1}), d_F(A_{k-1}, B_l))\} & \text{if } k, l > 1. \end{cases} \quad (1)$$

Proof. Omitted due to space constraint. □

Based on the above lemma, we investigate the combinatorial complexity of $VD_F(\mathcal{C})$, the Voronoi diagram of a set $\mathcal{C}$ of n planar polylines each with at most k vertices. Following [6, 15], a Voronoi diagram is a minimization of distance functions to the sites (in this case the polylines in $\mathcal{C}$). We first briefly review a result on the upper bound of lower envelopes in high dimensions by Sharir [15].

Let $\Sigma = \{\sigma_1, \ldots, \sigma_n\}$ be a collection of n $(d-1)$-dimensional algebraic surface patches in d-space. Let $\mathcal{A}(\Sigma)$ be the arrangement of Σ. The result in [15] holds upon the following three conditions.

(i) Each σ_i is monotone in the $x_1x_2 \ldots x_{d-1}$-direction (i.e. any line parallel to x_d-axis intersects σ_i in at most one point). Moreover, each σ_i is a portion of a $(d-1)$-dimensional algebraic surface of constant maximum degree b.

(ii) The projection of σ_i in x_d-direction onto the hyperplane $x_d = 0$ is a semi-algebraic set defined in terms of a constant number of $(d-1)$-variate polynomials of constant maximum degree.

(iii) The surface patches σ_i are in *general position* meaning that the coefficients of the polynomials defining the surfaces and their boundaries are algebraically independent over the rationals.

Theorem 3.1 *[15] Let Σ be a collection of n $(d-1)$-dimensional algebraic surface patches in d-space, which satisfy the above conditions (i),(ii), and (iii). Then the number of vertices of $\mathcal{A}(\Sigma)$ that lie at the lower envelope (i.e., level one) is $O(n^{d-1+\epsilon})$, for any $\epsilon > 0$.*

We now show a general upper bound on the combinatorial complexity of $VD_F(\mathcal{C})$.

We use the correspondence of polygonal chains with k vertices in $\mathbb{R}^d$ and points in $\mathbb{R}^{kd}$ given in Section 2. Let $c_1, c_2, \ldots, c_k$ be the sequence of vertices of a polygonal chain C and let $(x(c_i), y(c_i))$ be the coordinates of vertex $c_i, i = 1, \ldots, k$.

Lemma 3.3 *Let $B \in \mathbb{R}^{2l}$ be a polygonal chain of l vertices $b_1, \ldots, b_l$ in the plane where $b_i = (x(b_i), y(b_i))$. Let $f : \mathbb{R}^{2k} \to \mathbb{R}$ be the distance function defined as*

$$f(C) = d_F(C, B),$$

where $C = \langle c_1, \ldots, c_k \rangle \in \mathbb{R}^{2k}$ is a polygonal chain with k vertices and $c_i = (x(c_i), y(c_i)), i = 1, \ldots, k$. The space $\mathbb{R}^{2k}$ can be partitioned into at most $(kl)!$ semi-algebraic sets $R_1, R_2, R_3, \ldots$ such that the function $f(C)$ with domain restricted to any R_i is algebraic. Thus, the function $f(C)$ satisfies the above condition (i) and (ii).

Proof. We consider $\binom{k}{2} \cdot \binom{l}{2}$ manifolds in $\mathbb{R}^{2k}$ defined as

$$(x(c_i) - x(b_j))^2 + (y(c_i) - y(b_j))^2 = (x(c_{i'}) - x(b_{j'}))^2 + (y(c_{i'}) - y(b_{j'}))^2$$

for every four integer i, i', j, j' such that $1 \le i < i' \le k$ and $1 \le j < j' \le l$. They partition $\mathbb{R}^{2k}$ into at most $(kl)!$ semi-algebraic sets $R_1, R_2, R_3, \ldots$ corresponding to the order of distances between c_i and b_j for all $1 \le i \le k$ and $1 \le j \le l$.

Equation (1) in Lemma 3.2 implies that the function $f(C)$ restricted to a domain R_m satisfies

$$f(C) = \sqrt{(x(c_i) - x(b_j))^2 + (y(c_i) - y(b_j))^2}$$

for some pair i, j. The lemma follows. □

We now prove the following theorem regarding the combinatorial upper bound for $VD_F(\mathcal{C})$.

Theorem 3.2 *Let $\mathcal{C}$ be a collection of n polygonal chains $C_1, \ldots, C_n$ each with at most k vertices in the plane. The combinatorial complexity of the Voronoi diagram* $\mathrm{VD}_F(\mathcal{C})$ *is $O(n^{2k+\epsilon})$, for any $\epsilon > 0$.*

Proof. Due to space constraint, we only give a sketch of the proof. Every polygonal chain C with k vertices in the plane as a point in $\mathbb{R}^{2k}$. The Voronoi diagram $VD_F(\mathcal{C})$ can be viewed as a diagram in $\mathbb{R}^{2k}$. It is well-known that a Voronoi diagram can be interpreted as a minimization of distance functions to the sites [6, 15]. Let $f_i(C) = d_F(C, C_i)$, where $C \in \mathbb{R}^{2k}$ is a polygonal chain with k vertices. By Lemma 3.3 the function $f_i(C)$ satisfies the conditions (i) and (ii). And condition (iii) can be satisfied by perturbing the functions f_i as in [15]. Therefore, Voronoi diagram $VD_F(\mathcal{C})$ corresponds to the lower envelope in the arrangement of the surfaces $z = f_i(C)$ in $\mathbb{R}^{2k+1}$. By Theorem 3.1 the combinatorial complexity of $VD_F(\mathcal{C})$ is $O(n^{2k+\epsilon})$. □

3.2 The Combinatorial Upper Bound of $VD_F(\mathcal{C})$ in d-dimension

For protein-related applications, the input are polygonal chains in 3D. So it makes sense to consider the cases when $d > 2$. It turns out that Lemmas 3.1 and 3.2 hold for polygonal chains in $\mathbb{R}^d$. Similar to Lemma 3.3 we can prove

Lemma 3.4 *Let $B \in \mathbb{R}^{dl}$ be a polygonal chain of l vertices $b_1, \ldots, b_l$ in $\mathbb{R}^d$, where $b_i = (x_1(b_i), x_2(b_i), ..., x_d(b_i))$. Let $f : \mathbb{R}^{dk} \to \mathbb{R}$ be the distance function defined as*

$$f(C) = d_F(C, B),$$

where $C = \langle c_1, \ldots, c_k \rangle \in \mathbb{R}^{dk}$ is a polygonal chain with k vertices and $c_i = (x_1(c_i), x_2(c_i), ..., x_d(c_i))$,$i = 1, \ldots, k$. The space $\mathbb{R}^{dk}$ can be partitioned into at most $(kl)!$ semi-algebraic sets $R_1, R_2, R_3, \ldots$ such that the function $f(C)$ with domain restricted to any R_i is algebraic in d-dimension. Thus, the function $f(C)$ satisfies the above condition (i) and (ii).

Using Lemma 3.4 we can prove similar to Theorem 3.2 the following bound.

Theorem 3.3 *Let $\mathcal{C}$ be a collection of n polygonal chains $C_1, \ldots, C_n$ each with at most k vertices in $\mathbb{R}^d$. The combinatorial complexity of the Voronoi diagram* $\mathrm{VD}_F(\mathcal{C})$ *is $O(n^{dk+\epsilon})$, for any $\epsilon > 0$.*

4 The Combinatorial Lower Bounds of $VD_F(\mathcal{C})$

We now present a general lower bound for $VD_F(\mathcal{C})$. In fact we first show a result that even a slice of $VD_F(\mathcal{C})$ could contain a L_∞ Voronoi diagram in k dimensions, whose combinatorial complexity is $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$. This result is somehow a 'folklore' as the relationship between discrete Fréchet distance and L_∞-distance is known to many researchers, e.g., in [10]. We treat this more like a warm-up of our lower bound constructions.

Schaudt and Drysdale proved that a L_∞ Voronoi diagram in k dimensions has combinatorial complexity of $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$ [14]. Let $S = \{p_1, p_2, \ldots, p_n\}$ be a set of n points in $\mathbb{R}^k$ such that the L_∞ Voronoi diagram of S has complexity of $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$. Let $M > 0$ be a real number such that the hypercube $[-M, M]^k$ contains S and all the Voronoi vertices of the L_∞ Voronoi diagram of S. We consider a k-dimensional flat F of $\mathbb{R}^{2k}$ defined as $F = \{(a_1, M, a_2, 2M, \ldots, a_k, kM) \mid a_1, \ldots, a_k \in \mathbb{R}\}$ and a projection $\pi : F \to \mathbb{R}^k$ defined as $\pi(b) = (b_1, b_3, \ldots, b_{2k-1})$, for $b = (b_1, b_2, b_3, ..., b_{2k-1}, b_{2k})$.

Let $\mathcal{C} = \{C_1, C_2, ..., C_n\}$, each C_i being a planar polygonal chain with k vertices. Let $C_i = \langle c_{i1}, c_{i2}, ..., c_{ik} \rangle$ and $c_{im} = (x(c_{im}), y(c_{im}))$, for $m = 1, 2, ..., k$. We set $c_{im} = (p_{im}, mM)$, for $1 \le i \le n, 1 \le m \le k$. Clearly, every $C_i \in F$. With C_i we associate a point $C'_i = \pi(C_i)$ in $\mathbb{R}^k$. We show that the intersection of F and a $VD_F(\mathcal{C})$ has complexity of $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$.

Consider a point $T \in F$ such that $T' = \pi(T)$ is a L_∞ Voronoi vertex of S in $\mathbb{R}^k$. Then $T' \in [-M, M]^k$. At this point, the question is: what is the discrete Fréchet distance between T and a chain C_i? Note that $d_F^W(T, C_i) < M$ if and only if $W = W_0$ where $W_0 = \{(t_m, c_{im}) \mid m = 1, \ldots, k\}$. Therefore $d_F(T, C_i) = d_F^{W_0}(T, C_i) = \max\{|x(t_1) - x(c_{i1})|, |x(t_2) - x(c_{i2})|, ..., |x(t_j) - x(c_{ij})|, ..., |x(t_k) - x(c_{ik})|\}$. This is exactly the L_∞-distance between T' and C'_i, or $d_F(T, C_i) = d_\infty(T', C'_i)$. Then the slice of $VD_F(\mathcal{C})$ contains the L_∞ Voronoi diagram of S in k dimensions. We thus have the following theorem.

Theorem 4.1 *The combinatorial complexity of* $\mathrm{VD}_F(\mathcal{C})$ *for a set* $\mathcal{C}$ *of* n *planar polygonal chains with* k *vertices is* $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$*; in fact even a* k*-dimensional slice of* $\mathrm{VD}_F(\mathcal{C})$ *can have a combinatorial complexity of* $\Omega(n^{\lfloor \frac{k+1}{2} \rfloor})$.

We comment that this lower bound is apparently not tight and we next show a lower bound construction with does not make use of the L_∞ Voronoi diagram. We summarize our result as follows.

Theorem 4.2 *For any* d, k, n*, there is a set of* n *polygonal curves in* $\mathbb{R}^d$ *with* k *vertices each whose Voronoi diagram under the discrete Fréchet distance,* $\mathrm{VD}_F(\mathcal{C})$*, has combinatorial complexity* $\Omega(n^{dk})$ *for* $d = 1, 2$ *and* $k \in \mathbb{N}$ *and complexity* $\Omega(n^{d(k-1)+2})$ *for* $d > 2$ *and* $k \in \mathbb{N}$.

The above new lower bounds nearly close the gap to the $O(n^{dk+\epsilon})$ upper bound. However, it remains an open problem to close the gap.

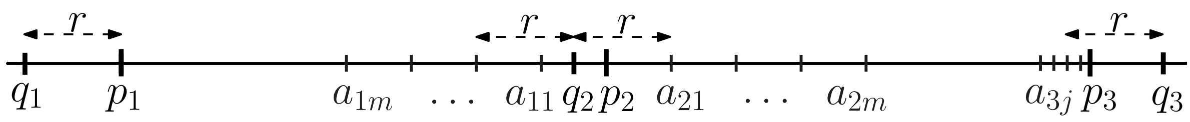

Figure 2: Construction for $d = 1$ and $k = 3$.

We show the lower bounds in Theorem 4.2 first for dimension $d = 1$ (Lemma 4.1) and then for dimensions $d \geq 2$ (Lemma 4.2). For both lower bounds we construct a set S of n curves. Then we construct $g(n)$ query curves which all lie in different Voronoi regions of the Voronoi diagram of S. This implies that the Voronoi diagram has complexity $\Omega(g(n))$.

Lemma 4.1 *For all n and k, there is a set of n polygonal curves in $\mathbb{R}^1$ with k vertices each whose Voronoi diagram under the discrete Fréchet distance has at least $\lfloor \frac{n}{k} \rfloor^k$ Voronoi regions.*

Proof. We construct a set S of n curves with k vertices each for $n = m \cdot k$ with $m \in \mathbb{N}$. S will be a union of k sets $S_1, \ldots, S_k$ of m curves each. We show that the Voronoi diagram of S contains m^k Voronoi regions.

The construction for $k = 3$ is shown in Figure 2. We place k points $p_1, \ldots, p_k$ with distance $2m$ between consecutive points on the real line. A curve in S has the form $\langle p_1, \ldots, p_{i-1}, p'_i, p_{i+1}, p_k \rangle$ for some $i \in \{1, \ldots, k\}$ and point p'_i close to p_i. Our construction uses the following points, curves, and sets of curves. See Figure 2 for an illustration for $k = 3$.

$$
\begin{aligned}
p_i &= (i-1)2m && \text{for } i = 1, \ldots, k \\
a_{1j} &= p_2 - j, \quad a_{2j} = p_2 + j && \text{for } j = 1, \ldots, m \\
a_{ij} &= p_i - j/(m+1) && \text{for } i = 3, \ldots, k,\ j = 1, \ldots, m \\
S_{ij} &= \langle p_1, a_{ij}, p_2, \ldots, p_k \rangle && \text{for } i = 1, 2,\ j = 1, \ldots, m \\
S_{ij} &= \langle p_1, \ldots, p_{i-1}, a_{ij}, p_{i+1}, \ldots, p_k \rangle && \text{for } i = 3, \ldots, k,\ j = 1, \ldots, m \\
S_i &= \{S_{i1}, \ldots, S_{im}\} && \text{for } i = 1, \ldots, k
\end{aligned}
$$

We claim that for all $1 \leq j_1, \ldots, j_k \leq m$ a query curve Q exists whose set of nearest neighbors in S under the discrete Fréchet distance, denoted by $N_S(Q)$, is

$$N_S(Q) = \{S_{11}, \ldots, S_{1j_1}, \ldots, S_{k1}, \ldots, S_{kj_k}\}. \tag{2}$$

Since these are m^k different sets, this implies that there are at least m^k Voronoi regions.

The query curve Q will have k vertices $q_1, \ldots, q_k$ with q_i close to p_i for $i = 1, \ldots, k$. The discrete Fréchet distance of Q to any curve in S will be realized by a bijection mapping each p_i or p'_i to q_i. Because the p_i are placed at large pairwise distances, this is the best possible matching of the vertices for the discrete Fréchet distance.

Let $r = (a_{2j_2} - a_{1j_1})/2$ denote half the distance between a_{1j_1} and a_{2j_2}. We choose the first vertex of Q as $q_1 = -r$. The second vertex q_2 we choose as midpoint between a_{1j_1} and a_{2j_2}, i.e.,

$q_2 = (a_{1j_1} + a_{2j_2})/2$. Since $p_1 = 0$, the distance between p_1 and q_1 is r. Because all curves in S start at p_1, this is the smallest possible discrete Fréchet distance between Q and any curve in S. We now construct the remaining points of Q, such that the curves in $N_S(Q)$ are exactly those given in equation 2 and these have discrete Fréchet distance r to Q.

We have already constructed q_2 such that it has distance at most r to the points $a_{11}, \ldots, a_{1j_1}$ and $a_{21}, \ldots, a_{2j_2}$ (cf. Figure 2). Now we choose the remaining points q_i as $q_i = a_{ij_i} + r$ for $i = 3, \ldots, k$. Then the point q_i has distance at most r to the points $p_i, a_{i1}, \ldots, a_{ij_i}$ for $i = 3, \ldots, k$. $\square$

Lemma 4.2 *For all n, k and for all $d \geq 2$, there is a set of n polygonal curves in $\mathbb{R}^d$ with k vertices each whose Voronoi diagram under the discrete Fréchet distance has at least $\lfloor \frac{n}{d(k-1)+2} \rfloor^{d(k-1)+2}$ Voronoi regions.*

Proof. We first give the construction for dimension $d = 2$ and then show how to generalize it for $d > 2$.

Construction for $\mathbf{d = 2}$. We construct the set S as union of $2k = d(k-1) + 2$ sets $S_1, \ldots, S_{2k}$ of m curves each for $m \in \mathbb{N}$.

First, we place k points $p_1, \ldots, p_k$ at sufficient pairwise distance in $\mathbb{R}^2$, that is, at distance $4r$ for some distance $r > 0$. Let $a_{11}, \ldots, a_{1m}$ be points evenly distributed on the circle of radius $2r$ around p_1. Let a_{21}, a_{31}, and a_{41} be points evenly distributed on the circle with radius r around p_2. Let the points $a_{22}, \ldots, a_{2m}$ lie on the line through a_{21} and p_2 moved away from a_{21} by at most $\varepsilon > 0$ as in Figure 3. The distance ε is sufficiently small for our construction, namely $\varepsilon < r \left(\frac{1}{\cos(\frac{\pi}{m})} - 1 \right)$ (assuming $m > 2$). Place the points $a_{32}, \ldots, a_{3m}$ and $a_{42}, \ldots, a_{4m}$ analogously.

For $i \geq 5$ the points a_{ij} are placed as follows. The points $a_{(i-1)j}$ and a_{ij} for $i = 2l$ are placed close to the point p_l. Then we place $a_{(i-1)1}$ and a_{i1} on the intersection of the coordinate axes originating in p_l with the circle of radius r around p_l. The points $a_{(i-1)j}, a_{ij}$ for $j \geq 2$ are placed on these axes, moved away from $a_{(i-1)1}, a_{i1}$ by at most $\delta > 0$. The distance δ is also sufficiently small for our construction, namely it is $\delta \leq (\sqrt{2} - 1)r$.

We can now define the curves in S. As in the construction for $d = 1$, the curves in S visit all but one of the points $p_1, \ldots, p_k$, and in the one point deviate slightly. We define

$$
\begin{aligned}
S_{1j} &= \langle a_{1j}, p_2, \ldots, p_k \rangle & &\text{for } j = 1, \ldots, m \\
S_{ij} &= \langle p_1, a_{ij}, p_2, \ldots, p_k \rangle & &\text{for } i = 2, 3, 4,\ j = 1, \ldots, m \\
S_{ij} &= \langle p_1, \ldots, p_{\lceil \frac{i}{2} \rceil - 1}, a_{ij}, p_{\lceil \frac{i}{2} \rceil}, \ldots, p_k \rangle & &\text{for } i = 5, \ldots, k,\ j = 1, \ldots, m \\
S_i &= \{S_{i1}, \ldots, S_{im}\} & &\text{for } i = 1, \ldots, 2k
\end{aligned}
$$

Again we claim that for all $1 \leq j_1, \ldots, j_{2k} \leq m$ a query curve Q exists whose set of nearest

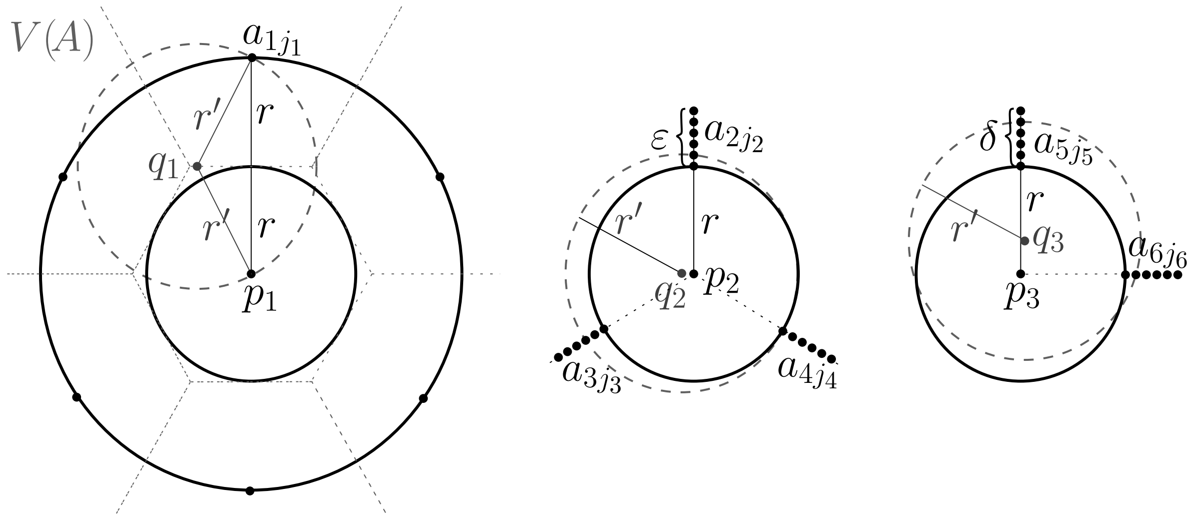

Figure 3: Construction for $d = 2$ and $k = 3$.

neighbors in S is

$$N_S(Q) = \{S_{1j_1}, S_{21}, \ldots, S_{2j_2}, \ldots, S_{(2m)1}, \ldots, S_{(2m)j_{(2m)}}\}.$$

This will imply that there are at least m^{2k} different Voronoi regions in the Voronoi diagram of S.

As second point q_2 of Q we choose the midpoint of the circle defined by the three points a_{2j_2}, a_{3j_3}, and a_{4j_4}. Let $r' > r$ be the radius of this circle. Note that $r' \le r + \varepsilon$ and that this circle contains the point p_2. Thus, the points $p_2, a_{21}, \ldots, a_{2j_2}, a_{31}, \ldots, a_{3j_3}$, and $a_{41}, \ldots, a_{4j_4}$ have distance at most r' to the point q_2.

As first point q_1 of Q we choose a point that has distance r' to both p_1 and a_{1j_1}, and a larger distance to all other a_{1j}. Consider the Voronoi diagram of the points $p, a_{11}, \ldots, a_{1m}$. Consider the edge between the cells of p and of a_{1j_1}. Because we chose ε sufficiently small, namely $\varepsilon < r\left(\frac{1}{\cos\left(\frac{\pi}{m}\right)} - 1\right)$, and because $r' \le r + \varepsilon$, there are two points in the interior of this edge with distance r' to p_1 and a_{1j_1}. We choose q_1 as one of these two points. Then the distance of q_1 to all a_{1j} for $j \neq j_1$ is larger than r'.

Now we choose the remaining points q_i of Q for $i = 3, \ldots, k$. Let $l = 2i, l' = 2i - 1$. There are two circles with radius r' that touch the points a_{lj_l} and $a_{l'j'_l}$. As q_i we choose the midpoint of the one circle that contains the point p_i. Then a point a_{lj} or $a_{l'j}$ has distance at most r' to q_i exactly if $l \le j_l$ or $l' \le j'_l$, respectively.

Construction for $\mathbf{d > 2}$. The construction can be generalized to $d > 2$ giving a lower bound of $m^{d(k-1)+2}$ for $m \cdot d(k-1) + 2$ curves.

The construction at p_1 remains the same. At p_2 we place $d + 1$ sets of points. Then one point from each set, i.e., $d + 1$ points, define a d-ball. At p_i for $i \ge 3$ we place d sets of points. Then one point from each set, i.e., d points, define a ball of the radius given by the choice of q_2. In total, this gives us $m^{d(k-1)+2}$ choices: m choices at p_1, m^{d+1} choices at p_2 and m^d choices each at $p_3, \ldots, p_k$. □

5 Concluding Remarks

In this paper, for the first time, we study the Voronoi diagram of polygonal chains under the discrete Fréchet distance. We show combinatorial upper and lower bounds for such a Voronoi diagram. We conjecture that the upper bound is tight (up to ε) as we have shown here for dimension $d = 2$. For closing the gap, consider how the construction for $d = 2$ is generalized to $d > 2$ in the proof of Lemma 4.2. While the constructions at the vertices $p_2, \ldots, p_k$ are replaced by higher-dimensional analogs, the construction at p_1 stays the same as in the two-dimensional case. Improving the construction at p_1 might close the gap.

At this point, to make the diagram useful, we need to develop efficient (approximation) algorithms to construct such a Voronoi diagram for decent k (say $k = 10 \sim 20$) so that one can first use a $(k-1)$-link chain to approximate a general input polyline. Although the running time might still look too high (due to the $O(k)$ exponent in the running time), the good news is that in many applications like protein structural alignment, n is not very large. Designing such an (exact or approximate) algorithm is another major open problem alone this line.

Acknowledgment

We thank Minghui Jiang for pointing out a gap in an earlier version of the arguments for Theorem 4.1.

References

[1] H. Alt and M. Godau. Measuring the resemblance of polygonal curves. In *Proceedings of the 8th Annual Symposium on Computational Geometry (SoCG'92)*, pages 102–109, 1992.

[2] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *Intl. J. Computational Geometry and Applications*, **5**:75-91, 1995.

[3] H. Alt, C. Knauer and C. Wenk. Matching polygonal curves with respect to the Fréchet distance. In *Proceedings of the 18th Annual Symposium on Theoretical Aspects of Computer Science (STACS'01)*, pages 63–74, 2001.

[4] T. Asano, J. Matousek and T. Tokuyama. The distance trisector curve. In *Proceedings of the 38th Annual Symposium on Theory of Comput.(STOC'06)*, pages 336–343, 2006.

[5] K. Buchin, M. Buchin and C. Wenk. Computing the Fréchet distance between simple polygons in polynomial time. In *Proceedings of the 22nd Annual Symposium on Computational Geometry (SoCG'06)*, pages 80–87, 2006.

[6] H. Edelsbrunner and R. Seidel. Voronoi diagrams and arrangements. *Discrete Comput. Geom.*, 1:25–44, 1986.

[7] T. Eiter and H. Mannila. Computing discrete Fréchet distance. *Tech. Report CD-TR 94/64, Information Systems Department, Technical University of Vienna*, 1994.

[8] M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Mathematico di Palermo*, **22**:1-74, 1906.

[9] M. Godau. On the complexity of measuring the similarity between geometric objects in higher dimensions. *PhD thesis, Freie Universitaet Berlin*, 1998.

[10] P. Indyk. Approximate nearest neighbor algorithms for Fréchet distance via product metrics. In *Proceedings of the 18th Annual Symposium on Computational Geometry (SoCG'02)*, pages 102–106, 2002.

[11] M. Jiang, Y. Xu and B. Zhu. Protein structure-structure alignment with discrete Fréchet distance. In *Proceedings of the 5th Asia-Pacific Bioinformatics Conf. (APBC'07)*, pages 131–141, 2007.

[12] F. Nielsen, J.-D. Boissonnat, and R. Nock. On Bregman Voronoi diagrams. In *Proc. 18th ACM-SIAM Sympos. Discrete Algorithms*, pages 746–755, 2007.

[13] F.P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, 1985.

[14] B. Schaudt and S. Drysdale. Higher dimensional Delaunay diagrams for convex distance functions. In *Proceedings of the 4th Canadian Conf. Comp. Geom. (CCCG'92)*, pages 274–279, 1992.

[15] M. Sharir. Almost tight upper bounds for lower envelopes in higher dimensions. *Discrete Comp. Geom.*, **12**:327-345, 1994.

[16] C. Wenk. Shape Matching in Higher Dimensions. *PhD thesis, Freie Universitaet Berlin*, 2002.

[17] B. Zhu. Protein local structure alignment under the discrete Fréchet distance. *J. Computational Biology*, 14(10):1343-1351, 2007.