Fréchet Distance for Uncertain Curves

Kevin Buchin
Department of Mathematics and Computer Science, TU Eindhoven, Netherlands
k.a.buchin@tue.nl

Chenglin Fan
Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA
cxf160130@utdallas.edu

Maarten Löffler
Department of Information and Computing Sciences, Utrecht University, Netherlands
m.loffler@uu.nl

Aleksandr Popov
Department of Mathematics and Computer Science, TU Eindhoven, Netherlands
a.popov@tue.nl

Benjamin Raichel
Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA
benjamin.raichel@utdallas.edu

Marcel Roeloffzen
Department of Mathematics and Computer Science, TU Eindhoven, Netherlands
m.j.m.roeloffzen@tue.nl

Abstract

In this paper we study a wide range of variants for computing the (discrete and continuous) Fréchet distance between uncertain curves. We define an uncertain curve as a sequence of *uncertainty regions*, where each region is a disk, a line segment, or a set of points. A *realisation* of a curve is a polyline connecting one point from each region. Given an uncertain curve and a second (certain or uncertain) curve, we seek to compute the lower and upper bound Fréchet distance, which are the minimum and maximum Fréchet distance for any realisations of the curves.

We prove that both problems are NP-hard for the continuous Fréchet distance, and the upper bound problem remains hard for the discrete Fréchet distance. In contrast, the lower bound discrete Fréchet distance can be computed in polynomial time using dynamic programming. Furthermore, we show that computing the expected discrete or continuous Fréchet distance is #P-hard when the uncertainty regions are modelled as point sets or line segments.

On the positive side, we argue that in any constant dimension there is a FPTAS for the lower bound problem when Δ/δ is polynomially bounded, where δ is the Fréchet distance and Δ bounds the diameter of the regions. We then argue there is a near-linear-time 3-approximation for the decision problem when the regions are convex and roughly δ-separated. Finally, we study the setting with Sakoe–Chiba bands, restricting the alignment of the two curves, and give polynomial-time algorithms for upper bound and expected (discrete) Fréchet distance for point-set-modelled uncertainty regions.

2012 ACM Subject Classification Theory of computation → Computational geometry

Keywords and phrases Curves, Uncertainty, Fréchet Distance, Hardness

Digital Object Identifier 10.4230/LIPIcs.ICALP.2020.20

Category Track A: Algorithms, Complexity and Games

Related Version The full version is available at `https://arxiv.org/abs/2004.11862` [13].

Funding *Chenglin Fan*: NSF CRII Award 1566137 and CAREER Award 1750780
Aleksandr Popov: Netherlands Organisation for Scientific Research (NWO), project no. 612.001.801
Benjamin Raichel: NSF CRII Award 1566137 and CAREER Award 1750780
Marcel Roeloffzen: Netherlands Organisation for Scientific Research (NWO), project no. 628.011.005

47th International Colloquium on Automata, Languages, and Programming (ICALP 2020).
Editors: Artur Czumaj, Anuj Dawar, and Emanuela Merelli; Article No. 20; pp. 20:1–20:20
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

In this paper we investigate the well-studied topic of curve similarity in the context of the burgeoning area of geometric computing under uncertainty. While classical algorithms in computational geometry typically assume the input point locations are known exactly, in recent years there has been a concentrated effort to adapt these algorithms to uncertain inputs, which can more faithfully model real-world inputs. The need to model such uncertain inputs is perhaps no more clear than for the location data of a moving object obtained from physical devices, which is inherently imprecise due to issues such as measurement error, sampling error, and network latency [42, 43]. Moreover, to ensure location privacy, one may purposely add uncertainty to the data by adding noise or reporting positions as geometric regions rather than points. (See the survey by Krumm [35] and the references therein.)

Here we consider both the continuous and discrete Fréchet distance for uncertain curves. Given the applications above, our uncertain input is given as a sequence of compact regions, from which a polygonal curve is *realised* by selecting one point from each region. Our goal is to find, for a given pair of uncertain curves, the upper bound, lower bound, and expected Fréchet distance, where the upper (resp. lower) bound Fréchet distance is the maximum (resp. minimum) distance over any realisation. For the expected Fréchet distance we assume a probability distribution is provided that describes how each vertex on a curve is chosen from the compact region. Previously, Ahn et al. [5] considered the lower bound problem for the discrete Fréchet distance, giving a polynomial-time algorithm for points in constant dimension. The authors also gave efficient approximation algorithms for the discrete upper bound Fréchet distance for uncertain inputs, where the approximation factor depends on the spread of the region diameters or how well-separated they are. Subsequently, Fan and Zhu showed that the discrete upper bound Fréchet distance is NP-hard for uncertain inputs modelled as thin rectangles [25]. To our knowledge, we are the first to consider either variant for the continuous Fréchet case, and the first to consider the expected Fréchet distance.

1.1 Previous Work

Geometric computing under uncertainty. The two most common models of geometric uncertainty are the locational model [36] and the existential model [46, 48]. In the existential model the location of an uncertain point is known, but the point may not be present; in the locational model we know that each uncertain point exists, but not its exact location.

In this paper we consider the locational model. Each uncertain point is a set of potential locations. We call an uncertain point *indecisive* if the set of potential locations is finite, or *imprecise* if the set is not finite but is a convex region. A *realisation* of a set of uncertain points is a selection of one point from each uncertain point. The goal is typically to compute the realisation of a set of uncertain points that minimises or maximises some quantity (e.g. area, distance, perimeter) of some underlying geometric structure (e.g. convex hull, MST) [1, 7, 14, 17, 21, 23, 28, 34, 37, 38, 39, 45, 47]. By assigning a probability distribution to uncertain points, one can also consider the expectation or distribution of various measures [2, 4, 32, 41].

Fréchet distance. Computing the Fréchet distance between two precise curves can be done in near-quadratic time [3, 6, 12], and assuming the Strong Exponential Time Hypothesis (SETH) it cannot be computed or even approximated well in strongly subquadratic time [9, 15]. However, for several restricted versions the Fréchet distance can be calculated more quickly, for example for c-packed curves [20], when the edges are long [29], or when the alignment of

Table 1 Hardness results for the decision problems in this paper. Ahn et al. [5] solve the lower bound problem for disks, but their algorithm extends to the indecisive curves as well as line segment imprecision.

		indecisive	imprecise: disks	imprecise: line segments
discrete Fréchet distance	LB	Polynomial [5]	Polynomial [5]	Polynomial [5]
	UB	NP-complete	NP-complete	NP-complete
	Exp	#P-hard	—	#P-hard
Fréchet distance	LB	Polynomial	—	NP-complete
	UB	NP-complete	NP-complete	NP-complete
	Exp	#P-hard	—	—

curves is restricted [11, 40]. Many variants of the problem have been considered: Fréchet distance with shortcuts [16, 19], weak Fréchet distance [6], discrete Fréchet distance [3, 22], Fréchet gap distance [24], Fréchet distance under translations [10, 26], and more.

1.2 Our Contributions

In this paper we present an extensive study of the Fréchet distance for uncertain curves. We provide a wide range of hardness results and present several approximations and polynomial-time solutions to restricted versions. We are the first to consider the continuous Fréchet distance in the uncertain setting, as well as the first to consider the expected Fréchet distance.

On the negative side, we present a plethora of hardness results (Table 1; details follow in Section 2). The hardness of the lower bound case is curious: while the variants discrete Fréchet distance on imprecise inputs [5] and, as we prove, continuous Fréchet distance on indecisive inputs both permit a simple dynamic programming solution, the variant continuous Fréchet distance on imprecise input has just enough (literal) wiggle room to show NP-hardness by reduction from SubsetSum.

We complement the lower bound hardness result by two approximation algorithms (Section 3). The first is a FPTAS for general uncertain curves in constant dimension when the ratio between the diameter of the uncertain points and the lower bound Fréchet distance is polynomially bounded. The second is a 3-approximation for separated imprecise curves, but uses a simpler greedy approach that runs in near-linear time.

The NP-hardness of the upper bound by a reduction from CNF-SAT is less surprising, but requires a careful set-up and analysis of the geometry to then extend it to a reduction from #CNF-SAT to the expected (discrete or continuous) Fréchet distance. However, by adding the common constraint that the alignment between the curves needs to stay within a Sakoe–Chiba [44] band of constant width (see Section 4 for definition and results), we can solve these problems in polynomial time for indecisive curves. Sakoe–Chiba bands are frequently used for time-series data [8, 33, 44] and trajectories [11, 18], when the alignment should (or is expected to) not vary too much from a certain "natural" alignment.

1.3 Preliminaries

Curves. Denote $[n] \equiv \{1, 2, \dots, n\}$. Consider a *sequence* of d-dimensional points $\pi = \langle p_1, p_2, \dots, p_n \rangle$. A *polygonal curve* π is defined by these points by linearly interpolating between the successive points and can be seen as a continuous function: $\pi(i + \alpha) = (1 - \alpha)p_i + \alpha p_{i+1}$ for $i \in [n-1]$ and $\alpha \in [0, 1]$. The *length* of such a curve is the number of

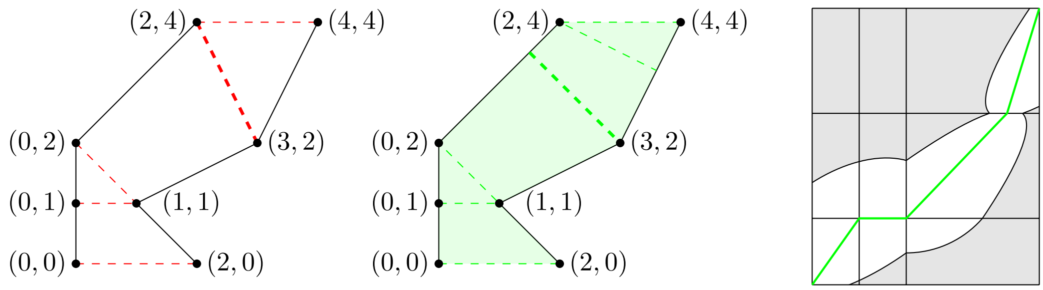


Figure 1 Left: Discrete Fréchet distance, where an optimal coupling is shown in dashed red lines. Middle: Fréchet distance, dashed green lines indicate specific values for δ for optimal functions ϕ_1, ϕ_2. Right: Free-space diagram for threshold $\delta = 2.15$. One can draw a monotonous path from the lower left corner to the upper right corner of the diagram, so the Fréchet distance between trajectories is below the threshold.

its vertices, $|\pi| = n$. Where we deem important to distinguish between points that are a part of the curve and other points, we denote the polygonal curve by $\pi = \langle \pi_1, \pi_2, \ldots, \pi_n \rangle$. We denote the *concatenation* of two polygonal curves π and σ of lengths n and m by $\pi \parallel \sigma$; the new curve follows π, then has a segment between $\pi(n)$ and $\sigma(1)$, and then follows σ. Similarly, $p \parallel q$ (or simply pq) denotes the line segment between points p and q. We can generalise this notation:

$$\mathop{\Big\|}_{i\in[n]} p_i \equiv p_1 \parallel p_2 \parallel \cdots \parallel p_n \equiv \pi \,.$$

We denote a *subcurve* from vertex i to j of curve π as $\pi[i : j] \equiv p_i \parallel p_{i+1} \parallel \cdots \parallel p_j$.

Metrics definitions. Given two points $x, y \in \mathbb{R}^d$, denote their Euclidean distance by $\|x - y\|$. For two compact sets $X, Y \subset \mathbb{R}^d$, denote their distance by $\|X - Y\| = \min_{x\in X, y\in Y} \|x - y\|$. Throughout we treat the dimension d as a small constant.

Let Φ_n denote the set of all *reparametrisations* of length n, defined as continuous non-decreasing functions $\phi : [0, 1] \to [1, n]$ where $\phi(0) = 1$ and $\phi(1) = n$. Given a pair of curves π and σ of lengths n and m, respectively, and corresponding reparametrisations $\phi_1 \in \Phi_n$ and $\phi_2 \in \Phi_m$, define $\mathrm{width}_{\phi_1,\phi_2}(\pi, \sigma) = \max_{t\in[0,1]} \|\pi(\phi_1(t)) - \sigma(\phi_2(t))\|$.

The width represents the maximum distance between two points traversing the curves from start to end according to ϕ_1 and ϕ_2 (which allow varying speed, but no backtracking). The *Fréchet distance* $d_{\mathrm{F}}(\pi, \sigma)$ is defined as the minimum possible width over all such traversals:

$$d_{\mathrm{F}}(\pi, \sigma) = \inf_{\phi_1\in\Phi_n, \phi_2\in\Phi_m} \mathrm{width}_{\phi_1,\phi_2}(\pi, \sigma) = \inf_{\phi_1\in\Phi_n, \phi_2\in\Phi_m} \max_{t\in[0,1]} \|\pi(\phi_1(t)) - \sigma(\phi_2(t))\| \,.$$

The *discrete Fréchet distance* $d_{\mathrm{dF}}(\pi, \sigma)$ is defined similarly, except that we do not traverse edges of the curves, but must jump from one vertex to the next on either or both curves. We define a valid *coupling* as a sequence $c = \langle (p_1, q_1), \ldots, (p_r, q_r) \rangle$ of pairs from $[n] \times [m]$ where $(p_1, q_1) = (1, 1)$, $(p_r, q_r) = (n, m)$, and, for any $i \in [r - 1]$ we have $(p_{i+1}, q_{i+1}) \in \{(p_i + 1, q_i), (p_i, q_i + 1), (p_i + 1, q_i + 1)\}$. Let $\mathcal{C}$ be the set of all valid couplings on curves of lengths n and m, then

$$d_{\mathrm{dF}}(\pi, \sigma) = \inf_{c\in\mathcal{C}} \max_{s\in[|c|]} \|\pi(p_s) - \sigma(q_s)\| \,,$$

where $c_s = (p_s, q_s)$ for all $s \in [|c|]$. Both distances, as well as a common approach to computing Fréchet distance, are illustrated in Figure 1.

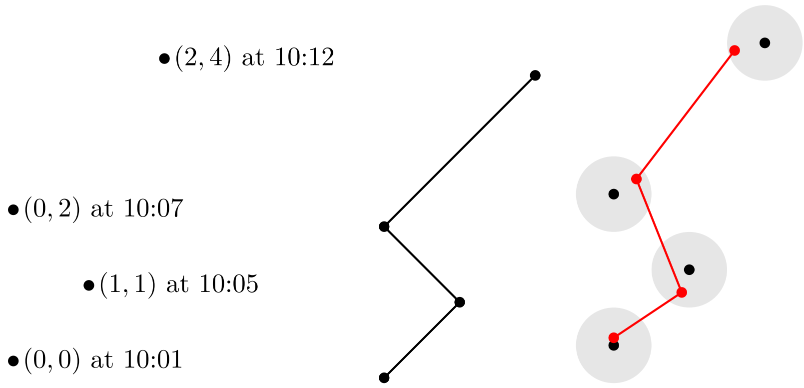


Figure 2 Left: Trajectory data. Middle: Polygonal curve on the data. Right: Imprecise curve with disks as imprecision regions and real curve.

Uncertainty model. An *uncertain* point is commonly represented as a compact region $U \subset \mathbb{R}^d$. Usually, it is a finite set of points, a disk, a rectangle, or a line segment. The intuition is that only one point from this region represents the true location of the point; however, we do not know which one. A *realisation* p of such a point is one of the points from the region U. When needed we assume the realisations are drawn from U according to a known probability distribution $\mathbb{P}$. We denote the diameter of any compact set (e.g. an uncertain point) $U \subset \mathbb{R}^d$ by $\operatorname{diam}(U) = \max_{p,q \in U} \|p - q\|$. An *indecisive* point is a special case of an uncertain point: it is a set of points $U = \{p_1, \dots, p_k\}$, where each point $p_i \in \mathbb{R}^d$ for $i \in [k]$. Similarly, an *imprecise* point is a compact convex region $U \subset \mathbb{R}^d$. We will often use disks or line segments as such regions. Note that a precise point is a special case of an indecisive point (set of size one) and an imprecise point (disk of radius zero).

Uncertain curves and distances. Define an *uncertain curve* as a sequence of uncertain points $\mathcal{U} = \langle U_1, \dots, U_n \rangle$. A *realisation* $\pi \Subset \mathcal{U}$ of an uncertain curve is a polygonal curve $\pi = \langle p_1, \dots, p_n \rangle$, where each p_i is a realisation of the corresponding uncertain point U_i. We denote the set of all realisations of an uncertain curve $\mathcal{U}$ by $\operatorname{Real}(\mathcal{U})$ (see Figure 2). In a probabilistic setting, we write $\pi \Subset_{\mathbb{P}} \mathcal{U}$ to denote that each point of π gets drawn from the corresponding uncertainty region independently according to distribution $\mathbb{P}$.

For uncertain curves $\mathcal{U}$ and $\mathcal{V}$, define the upper bound, lower bound, and expected discrete Fréchet distance (and extend to continuous Fréchet distance $d_{\mathrm{F}}^{\max}$, $d_{\mathrm{F}}^{\min}$, $d_{\mathrm{F}}^{\mathbb{E}(\mathbb{P})}$ using d_{F}) as:

$$d_{\mathrm{dF}}^{\max}(\mathcal{U}, \mathcal{V}) = \max_{\pi \Subset \mathcal{U}, \sigma \Subset \mathcal{V}} d_{\mathrm{dF}}(\pi, \sigma)\,, \qquad d_{\mathrm{dF}}^{\min}(\mathcal{U}, \mathcal{V}) = \min_{\pi \Subset \mathcal{U}, \sigma \Subset \mathcal{V}} d_{\mathrm{dF}}(\pi, \sigma)\,,$$

$$d_{\mathrm{dF}}^{\mathbb{E}(\mathbb{P})}(\mathcal{U}, \mathcal{V}) = \mathbb{E}_{\pi \Subset_{\mathbb{P}} \mathcal{U}, \sigma \Subset_{\mathbb{P}} \mathcal{V}}[d_{\mathrm{dF}}(\pi, \sigma)]\,.$$

If the distribution is clear from the context, we write $d_{\mathrm{F}}^{\mathbb{E}}$ and $d_{\mathrm{dF}}^{\mathbb{E}}$. The definitions above also apply if one of the curves is precise, as a precise curve is a special case of an uncertain curve.

2 Hardness Results

In this section, we first discuss the hardness results for the upper bound and expected value of the continuous and discrete Fréchet distance for indecisive and imprecise curves. We then show hardness of finding the lower bound continuous Fréchet distance on imprecise curves.

2.1 Upper Bound and Expected Fréchet Distance

We present proofs of NP-hardness and #P-hardness for the upper bound and expected Fréchet distance for both indecisive and imprecise curves by showing polynomial-time reductions from CNF-SAT and #CNF-SAT (counting version). We consider the upper bound problem for indecisive curves and then illustrate how the construction can be used to show #P-hardness for the expected Fréchet distance (both discrete and continuous). We then illustrate how the construction can be adapted to show hardness for imprecise curves. All our constructions are in two dimensions. The missing proofs can be found in the full version [13].

2.1.1 Upper Bound Fréchet Distance on Indecisive Curves

Define the following problem and its continuous counterpart, using $d_{\mathrm{F}}^{\max}$ instead:

▶ **Problem 1.** *Upper Bound Discrete Fréchet: Given two uncertain curves $\mathcal{U}$ and $\mathcal{V}$ and a threshold $\delta \in \mathbb{R}^+$, decide if $d_{\mathrm{dF}}^{\max}(\mathcal{U}, \mathcal{V}) > \delta$.*

Suppose we are given a CNF-SAT formula C with n clauses, C_1 to C_n, on m boolean variables, x_1 to x_m. We pick some value $0.12 \le \varepsilon < 0.25$.[1] Construct a *variable* curve, where each variable corresponds to an indecisive point with locations $(0, 0.5 + \varepsilon)$ and $(0, -0.5 - \varepsilon)$; the locations are interpreted as assigning the variable True and False. Any realisation of the curve corresponds to a variable assignment. Each indecisive point is followed by a precise point that is far away, to force synchronisation with the other curve:

$$\mathrm{VG}_j = \{(0, 0.5 + \varepsilon), (0, -0.5 - \varepsilon)\} \parallel (2, 0)\,.$$

Consider a specific clause C_i of the formula. We define an *assignment gadget* $\mathrm{AG}_{i,j}$ for each variable x_j and clause C_i depending on how the variable occurs in the clause.

$$\mathrm{AG}_{i,j} = \begin{cases} (0, -0.5) \parallel (1, 0) & \text{if } x_j \text{ is a literal of } C_i, \\ (0, 0.5) \parallel (1, 0) & \text{if } \neg x_j \text{ is a literal of } C_i, \\ (0, 0) \parallel (1, 0) & \text{otherwise.} \end{cases}$$

Note that if assignment $x_j = \textsc{True}$ makes a clause C_i true, then the first precise point of the corresponding assignment gadget appears at distance $1 + \varepsilon$ from the realisation corresponding to setting $x_j = \textsc{True}$ of the indecisive point in VG_j. We can repeat the construction, yielding a *variable clause gadget* and an *assignment clause gadget:*

$$\mathrm{VCG} = (-2, 0) \parallel \Big\Vert_{j \in [m]} \mathrm{VG}_j\,, \qquad \mathrm{ACG}_i = (-1, 0) \parallel \Big\Vert_{j \in [m]} \mathrm{AG}_{i,j}\,.$$

Consider the Fréchet distance between the two gadgets. Observe that matching a synchronisation point from one gadget with a non-synchronisation point in the other yields a distance more than $1 + \varepsilon$, whereas matching synchronisation points pairwise and non-synchronisation points pairwise will yield the distance at most $1 + \varepsilon$. So we only consider one-to-one couplings, i.e. we match point i on one curve to point i on the other curve, for all i.

Now, if a realisation corresponds to a satisfying assignment, then for some x_j we have picked the realisation that is opposite from the coupled point on the clause curve, yielding the bottleneck distance of $1 + \varepsilon$. If the realisation corresponds to a non-satisfying assignment, then the synchronisation points establish the bottleneck, yielding the distance 1. So, we can clearly distinguish between a satisfying and a non-satisfying assignment for a clause.

[1] This range is determined by the relative distances in the construction.

Next, we define the *variable curve* and the *clause curve* as follows:

$$\mathrm{VC} = (0,0) \parallel \mathrm{VCG} \parallel (0,0)\,, \qquad \mathrm{CC} = \mathop{\Big\|}_{i\in[n]} \mathrm{ACG}_i\,.$$

Observe that the synchronisation points at $(-2, 0)$ and $(-1, 0)$ ensure that for any optimal coupling we match up VCG with some ACG_i as described before. Also note that all the points on CC are within distance 1 from $(0, 0)$. Therefore, we can always pick any one of n clauses to align with VCG, and couple the remaining points to $(0, 0)$; the bottleneck distance will then be determined by the distance between VCG and the chosen ACG_i.

Now consider a specific realisation of VCG. If the corresponding assignment does not satisfy C, then we can synchronise VCG with a clause that is false to obtain a distance of 1. If the assignment corresponding to the realisation satisfies all clauses, we must synchronise VCG with a satisfied clause, which yields a distance of $1 + \varepsilon$. The construction is shown in Figures 3 and 4.

We can use similar reasoning to arrive at the same conclusion if we compute the Fréchet distance instead. The necessary adaptations are presented in the full version [13].

▶ **Theorem 2.** *The problems* Upper Bound Discrete Fréchet *and* Upper Bound Continuous Fréchet *for indecisive curves are NP-hard.*

2.1.2 Expected Fréchet Distance on Indecisive Curves

We show that finding expected discrete Fréchet distance is #P-hard by providing a polynomial-time reduction from #CNF-SAT, i.e. the problem of finding the number of satisfying assignments to a CNF-SAT formula. Missing details can be found in the full version [13]. Define the following problem and its continuous counterpart:

▶ **Problem 3.** Expected Discrete Fréchet: *Find $d_{\mathrm{dF}}^{\mathbb{E}(\mathbb{U})}(\mathcal{U}, \mathcal{V})$ for uncertain curves $\mathcal{U}, \mathcal{V}$.*

The main idea is to derive an expression for the number of satisfying assignments in terms of $d_{\mathrm{dF}}^{\mathbb{E}(\mathbb{U})}(\mathrm{VC}, \mathrm{CC})$. This works, since there is a one-to-one correspondence between boolean variable assignment and a choice of realisation of VC, so counting the number of satisfying assignments corresponds to finding the proportion of realisations yielding large Fréchet distance. We can establish the result for Expected Continuous Fréchet similarly.

▶ **Theorem 4.** *The problems* Expected Discrete Fréchet *and* Expected Continuous Fréchet *for indecisive curves are #P-hard.*

2.1.3 Imprecise Curves

We have so far considered indecisive points; instead, we can look at imprecise points, namely, line segments or disks. We can show similar hardness results in that setting. We alter the construction – instead of the point $\{(0, 0.5 + \varepsilon), (0, -0.5 - \varepsilon)\}$, we either have the disk centred at $(0, 0)$ with radius $0.5 + \varepsilon$ or the line segment connecting $(0, -0.5 - \varepsilon)$ and $(0, 0.5 + \varepsilon)$. Observe that the locations of the indecisive point are still on the disk or the line segment. We can show that the upper bound decision problem is NP-hard by showing that we can always consider only the extreme locations on the imprecise points that coincide with the locations of the indecisive points.

▶ **Theorem 5.** *The problems* Upper Bound Discrete Fréchet *and* Upper Bound Continuous Fréchet *for imprecise curves modelled as line segments or disks are NP-hard.*

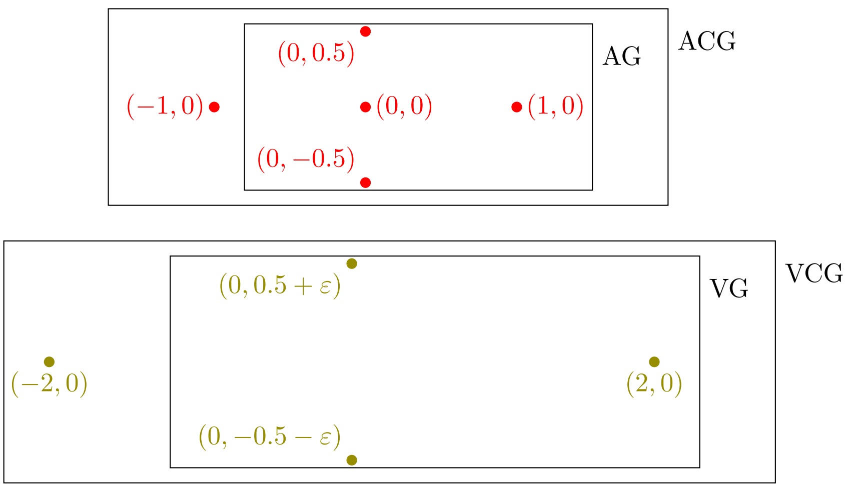


Figure 3 Illustration of gadgets used in the basic construction.

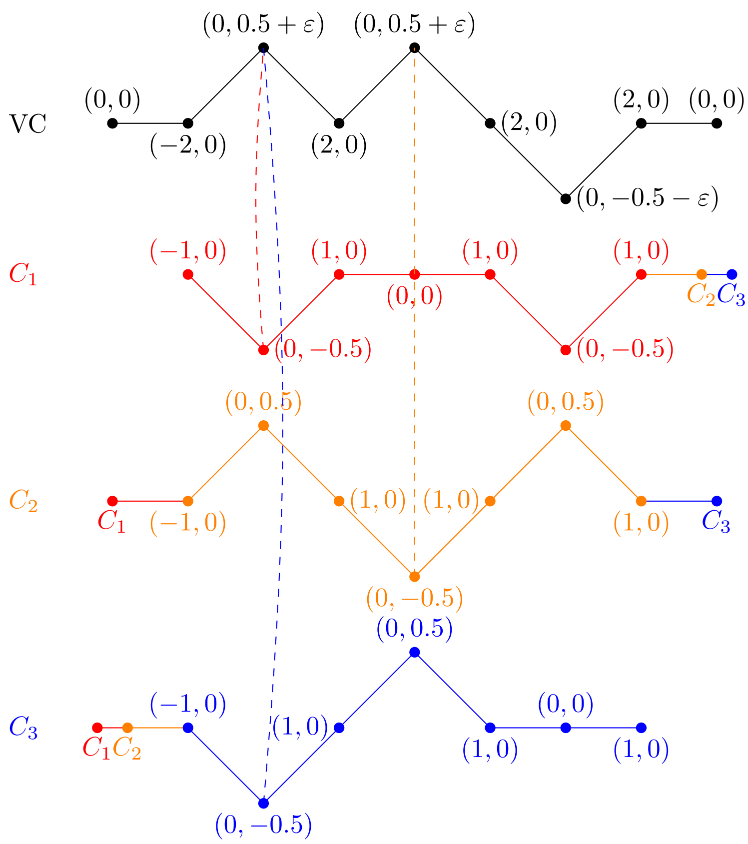


Figure 4 Realisation of VC for assignment $x_1 = \textsc{True}$, $x_2 = \textsc{True}$, $x_3 = \textsc{False}$ and the CC for formula $C = (x_1 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee \neg x_3) \wedge (x_1 \vee \neg x_2)$. Note that $C = \textsc{True}$ with the given variable assignment. Also note that we can choose any of C_1, C_2, C_3 to align with VC; we always get the bottleneck distance of $1 + \varepsilon$, as all three are satisfied, so here $d_{\mathrm{dF}}(\mathrm{VC}, \mathrm{CC}) = 1 + \varepsilon$.

We can also consider the value of expected Fréchet distance on imprecise points. We show the result only for points modelled as line segments; in principle, we believe that for disks a similar result holds, but the specifics of our reduction do not allow for clean computations.

We cannot immediately use our construction: we treat subsegments at the ends of the imprecision segments as TRUE and FALSE, but we have no interpretation for points in the centre part of a segment. So, we want to separate the realisations that pick any such invalid points. To that aim, we introduce extra gadgets to the clause curve that act as clauses, but catch these invalid realisations, so each of them yields the distance of 1. Now we have three distinct cases: realisation is satisfying, non-satisfying, or invalid. We can derive the expression connecting $d_{\mathrm{dF}}^{\mathbb{E}(\mathbb{U})}$ and the number of satisfying assignments.

▶ **Theorem 6.** *The problem* EXPECTED DISCRETE FRÉCHET *for imprecise curves modelled as line segments is #P-hard.*

2.2 Lower Bound Fréchet Distance

In this section, we prove that computing the lower bound Fréchet distance is NP-hard. The missing proofs can be found in the full version [13]. Unlike the upper bound proofs, this reduction uses the NP-hard problem SUBSET-SUM. Consider the following problems.

▶ **Problem 7.** LOWER BOUND CONTINUOUS FRÉCHET: *Given a polygonal curve π with n vertices, an uncertain curve $\mathcal{U}$ with m vertices, and a threshold $\delta > 0$, decide if $d_{\mathrm{F}}^{\min}(\pi, \mathcal{U}) \le \delta$.*

▶ **Problem 8.** SUBSET-SUM: *Given a set $S = \{s_1, \dots, s_n\}$ of n positive integers and a target integer τ, decide if there exists an index set I such that $\sum_{i \in I} s_i = \tau$.*

2.2.1 An Intermediate Problem

We start by reducing SUBSET-SUM to a more geometric intermediate curve-based problem.

▶ **Definition 9.** *Let $\alpha > 0$ be some value, and let $\sigma = \langle \sigma_1, \dots, \sigma_{2n+1} \rangle$ be a polygonal curve. Call σ an* α-regular curve *if for all $1 \le i \le 2n+1$, the x-coordinate of σ_i is $i \cdot \alpha$. Let $Y = \{y_1, \dots, y_n\}$ be a set of n positive integers. Call σ a* Y-respecting curve *if:*

1. *For all $1 \le i \le n$, σ passes through the point $((2i + 1/2)\alpha, 0)$.*
2. *For all $1 \le i \le n$, σ either passes through the point $((2i - 1/2)\alpha, 0)$ or $((2i - 1/2)\alpha, -y_i)$.*

Intuitively, the above definition requires σ to pass through $((2i + 1/2)\alpha, 0)$ as it reflects the y-coordinate about the line $y = 0$ (see Figure 5). Thus, if the curve also passes through $((2i - 1/2)\alpha, 0)$, the two reflections cancel each other. If it passes through $((2i - 1/2)\alpha, -y_i)$, the lemma below argues that y_i shows up in the final vertex height.

▶ **Lemma 10.** *Let σ be a Y-respecting α-regular curve, and let I be the subset of indices i such that σ passes through $((2i - 1/2)\alpha, -y_i)$. If $\sigma_1 = (\alpha, 0)$, then $\sigma_{2n+1} = ((2n+1)\alpha, 2\sum_{i \in I} y_i)$.*

The following is needed in the next section, and follows from the proof of the above.

▶ **Corollary 11.** *For a set $Y = \{y_1, \dots, y_n\}$, let $M = \sum_{i=1}^{n} y_i$. For any vertex σ_i of a Y-respecting α-regular curve, its y-coordinate is at most $2M$ and at least $-2M$.*

▶ **Problem 12.** RR-CURVE: *Given a set $Y = \{y_1, \dots, y_n\}$ of n positive integers, a value $\alpha = \alpha(Y) > 0$, and an integer τ, decide if there is a Y-respecting α-regular curve $\sigma = \langle \sigma_1, \dots, \sigma_{2n+1} \rangle$ such that $\sigma_1 = (\alpha, 0)$ and $\sigma_{2n+1} = ((2n+1)\alpha, 2\tau)$.*

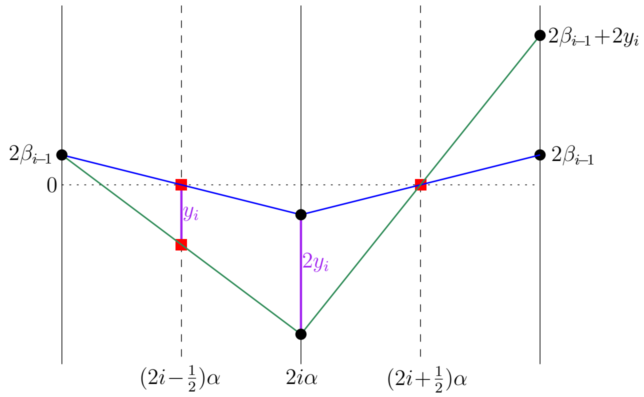


Figure 5 Passing through $((2i-1/2)\alpha, 0)$ does not change the height, and passing through $((2i-1/2)\alpha, -y_i)$ adds $2y_i$.

By Lemma 10, Subset-Sum immediately reduces to the above problem by setting $Y = S$. Note that for this reduction it suffices to use any positive constant for α; however, we allow α to depend on Y, as this will ultimately be needed in our reduction to Problem 7.

▶ **Theorem 13.** *For any $\alpha(Y) > 0$, RR-Curve is NP-hard.*

2.2.2 Reduction to Lower Bound Fréchet Distance

Let α, τ, $Y = \{y_1, \ldots, y_n\}$ be an instance of RR-Curve. In this section, we show how to reduce it to an instance δ, π, $\mathcal{U}$ of Problem 7, where the uncertain regions in $\mathcal{U}$ are vertical line segments. The main idea is to use $\mathcal{U}$ to define an α-regular curve, and use π to enforce that it is Y-respecting. Specifically, let $M = \sum_{i=1}^{n} y_i$. Then $\mathcal{U} = \langle v_1, \ldots, v_{2n+1}\rangle$, where each v_i is a vertical segment whose horizontal coordinate is $i\alpha$ and whose vertical extent is given by the interval $[-2M, 2M]$. By Corollary 11, we have the following simple observation.

▶ **Observation 14.** *The set of all Y-respecting α-regular curves is a subset of* $\mathrm{Real}(\mathcal{U})$.

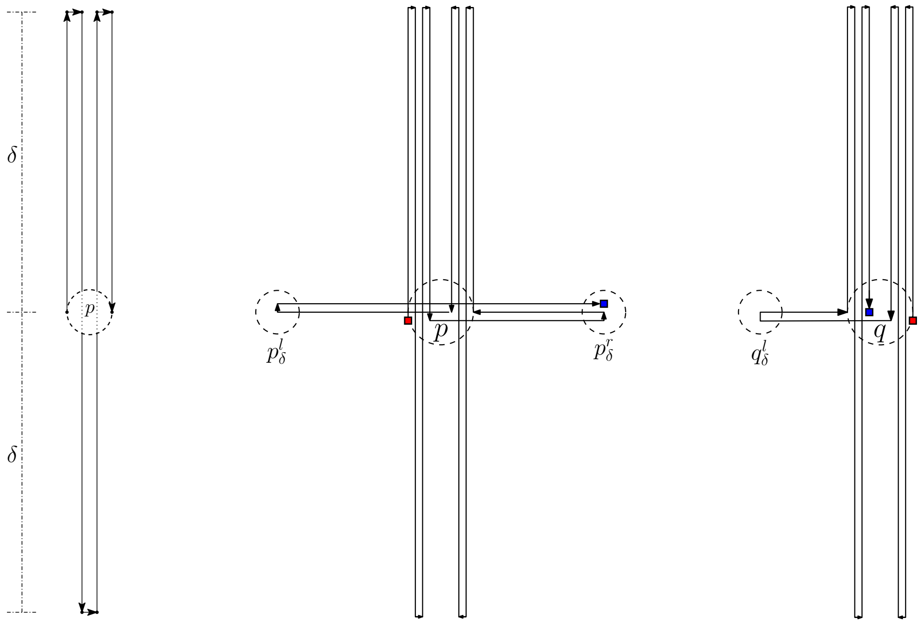
Thus, the main challenge is to define π to enforce that the realisation is Y-respecting. To that end, we first describe a gadget forcing the realisation to pass through a specified point.

▶ **Definition 15.** *For any point $p = (x, y) \in \mathbb{R}^2$ and value $\delta > 0$, let the δ* gadget *at p, denoted $\mathrm{g}_\delta(p)$, be the curve: $(x, y) \parallel (x, y+\delta) \parallel (x, y-\delta) \parallel (x, y+\delta) \parallel (x, y)$. See Figure 6a.*

▶ **Lemma 16.** *Let $p = (x, y) \in \mathbb{R}^2$ be a point, and let ℓ be any line segment. Then if $d_{\mathrm{F}}(\ell, \mathrm{g}_\delta(p)) \le \delta$, then ℓ must pass through p.*

For our uncertain curve to be Y-respecting, it must pass through all points of the form $((2i+1/2)\alpha, 0)$. This condition is satisfied by the lemma above by placing a δ gadget at each such point. The second condition of a Y-respecting curve is that it passes through $((2i-1/2)\alpha, 0)$ or $((2i-1/2)\alpha, -y_i)$. This condition is much harder to encode, and requires putting several δ gadgets together to create a composite gadget, which we now describe.

▶ **Definition 17.** *For any point $p = (x, y) \in \mathbb{R}^2$ and value $\delta > 0$, let $p^l_\delta = (x - \delta/2, y)$ and $p^r_\delta = (x + \delta/2, y)$. Define the δ* lower composite gadget *at p, denoted $\mathrm{lcg}_\delta(p)$, to be the curve $\mathrm{g}_\delta(p) \parallel p^r_\delta \parallel \mathrm{g}_\delta(p) \parallel p^l_\delta \parallel p^r_\delta$. See Figure 6b. Define the δ* upper composite gadget *at q, denoted $\mathrm{ucg}_\delta(q)$, to be the curve $\mathrm{g}_\delta(q) \parallel q^l_\delta \parallel \mathrm{g}_\delta(q)$. See Figure 6c. Define the δ* composite gadget *of p and q, denoted $\mathrm{cg}_\delta(p, q)$, to be the curve $\mathrm{lcg}_\delta(p) \parallel \mathrm{ucg}_\delta(q)$.*



(a) $\mathrm{g}_\delta(p)$. **(b)** $\mathrm{lcg}_\delta(p)$. **(c)** $\mathrm{ucg}_\delta(q)$.

Figure 6 Depiction of gadgets $\mathrm{g}_\delta(p)$, $\mathrm{lcg}_\delta(p)$, and $\mathrm{ucg}_\delta(p)$. Circles represent zero-area points. For the right two figures, the red / blue square represents the starting / ending point.

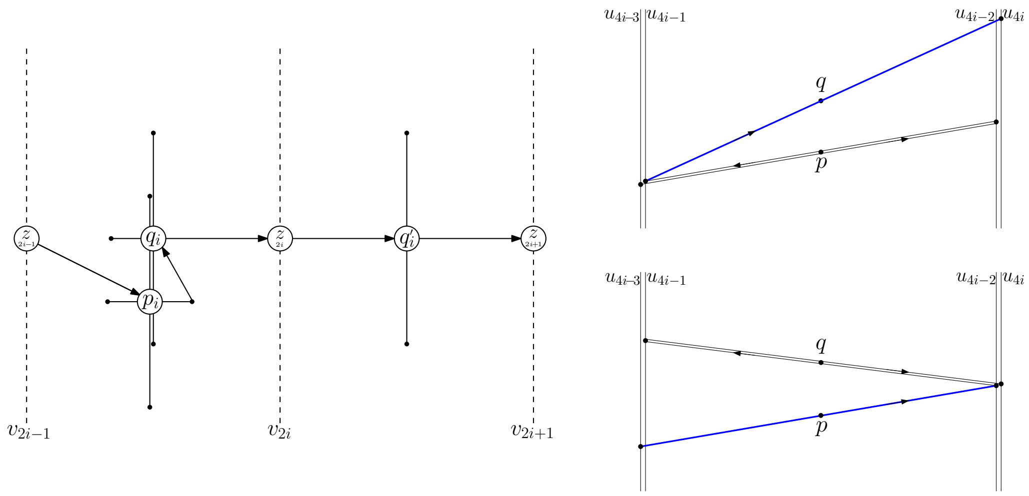
To use this composite gadget, we centre the lower gadget at height $-y_i$ and the upper gadget directly above it at height zero. As the two gadgets are on top of each other, ultimately we require our uncertain curve to go back and forth once between consecutive vertical line segments, for which we have the following key property.

▶ **Lemma 18.** *Let $p = (x_p, -y_p)$ and $q = (x_p, 0)$ be points in $\mathbb{R}^2$. Let $\sigma = \langle a, b, c, d\rangle$ be a three-segment curve such that $b_x > x_p + \delta$ and $c_x < x_p - \delta$. If $d_{\mathrm{F}}(\sigma, \mathrm{cg}_\delta(p, q)) \le \delta$, then:*
(i) *the segment ab must pass through p,*
(ii) *the segment cd must pass through q, and*
(iii) *the segment bc must either pass through p or through q.*
In particular, either ab and bc are on the same line, or cd and bc are on the same line.

Let v_l, v_r be vertical segments lying to the left and right of $\mathrm{cg}_\delta(p, q)$ further than δ away, and let z_l, z_r be the points on v_l and v_r at the same height as q. Consider the uncertain curve $\mathcal{U} = \langle U_1, U_2, U_3, U_4\rangle$, where $U_1 = U_3 = v_l$ and $U_2 = U_4 = v_r$. By Lemma 18, if there is a curve $\sigma \Subset \mathcal{U}$ such that $d_{\mathrm{F}}(\sigma, z_l \parallel \mathrm{cg}_\delta(p, q) \parallel z_r) \le \delta$, then implicitly it defines a single edge from v_l to v_r either passing through p or passing through q (see Figure 7b, whose notation is defined below). The following lemma acts as a rough converse of Lemma 18.

▶ **Lemma 19.** *Let $p = (x_p, -y_p)$ and $q = (x_p, 0)$ be points in $\mathbb{R}^2$, with $y_p \le \delta/4$. Let $\sigma = \langle p, b, c, q\rangle$ be a curve such that $x_p + \delta < b_x \le x_p + 1.1\delta$, $x_p - 1.1\delta \le c_x < x_p - \delta$, and $-\delta/2 \le b_y, c_y \le \delta/2$. If bc passes through either p or q, then $d_{\mathrm{F}}(\sigma, \mathrm{cg}_\delta(p, q)) \le \delta$.*

We now give the reduction from RR-Curve to Problem 7, whose correctness follows from the lemmas and discussion above. (See the full version [13] for details.) Let $\alpha(Y)$, τ, $Y = \{y_1, \ldots, y_n\}$ be an instance of RR-Curve. For the reduction to Problem 7, we set $\delta = 4M$, where $M = \sum_{i=1}^{n} y_i$. Theorem 13 allows us to choose how to set $\alpha(Y)$, and we



(a) Pictorial representation of λ_i.

(b) The two solutions.

Figure 7 On the left, λ_i. On the right, the two possible solutions with Fréchet distance at most δ. The top (resp. bottom) corresponds to an α-regular curve passing through q (resp. p).

set $\alpha = 2.1\delta = 8.4M$. Let $V = \{v_1, \ldots, v_{2n+1}\}$ be a set of vertical line segments where all upper (resp. lower) endpoints of the segments have height $2M$ (resp. $-2M$), and for all i, the x-coordinate of v_i is $i\alpha$. Let $\mathcal{U} = \langle U_1, \ldots, U_{4n+1}\rangle$ be the uncertain curve such that $U_{4n+1} = v_{2n+1}$, and for all $1 \le i \le n$, $U_{4i-3} = v_{2i-1}$, $U_{4i-2} = v_{2i}$, $U_{4i-1} = v_{2i-1}$, and $U_{4i} = v_{2i}$. For $1 \le i \le 2n+1$, define the points $z_i = (i\alpha, 0)$, and for $1 \le i \le n$, define $q_i = ((2i - 1/2)\alpha, 0)$, $q_i' = ((2i + 1/2)\alpha, 0)$, and $p_i = ((2i - 1/2)\alpha, -y_i)$. For a given value $1 \le i \le n$, consider the curve $\lambda_i = z_{2i-1} \parallel \mathrm{cg}_\delta(p_i, q_i) \parallel z_{2i} \parallel \mathrm{g}_\delta(q_i')$ (see Figure 7a). Let $s = (\alpha, 0)$ and $t = ((2n+1)\alpha, 2\tau)$. Then $\pi = \mathrm{g}_\delta(s) \parallel \lambda_1 \parallel \lambda_2 \parallel \cdots \parallel \lambda_{n-1} \parallel \lambda_n \parallel \mathrm{g}_\delta(t)$.

▶ **Theorem 20.** *Lower Bound Continuous Fréchet (Problem 7) is NP-hard, even when the uncertain regions are all equal-length vertical segments with the same height and the same horizontal distance (to the left or right) between adjacent uncertain regions.*

3 Algorithms for Lower Bound Fréchet Distance

In the previous section, we have shown that the decision problem for $d_F^{\min}$ is hard, given a polygonal curve and an uncertain curve with line-segment-based imprecision model. Interestingly, the same problem is solvable in polynomial time for indecisive curves. The key idea is that we can use a dynamic programming approach similar to that for computing Fréchet distance [6] and only keep track of realisations of the last indecisive point considered so far. (Note that one can also reduce the problem to Fréchet distance between paths in *DAG complexes,* studied by Har-Peled and Raichel [31], but this yields a slower running time.)

Consider the setting with an indecisive curve $\mathcal{V} = \langle V_1, \ldots, V_n\rangle$ of n points and a precise curve $\pi = \langle p_1, \ldots, p_m\rangle$ with m points; each indecisive point has k possible realisations, $V_i = \{q_i^1, \ldots, q_i^k\}$. We can propagate reachability column by column. Define $\mathrm{Feas}(i, \ell)$ to be the *feasibility column* for realisation q_i^ℓ of U_i. This is a set of intervals on the vertical cell boundary line in the free-space diagram (see Figure 1), corresponding to the subintervals of one curve within distance δ from a point on the other curve. It is computed exactly the same way as for the precise Fréchet distance – it depends on the distance between a point and a line segment and gives a single interval on each vertical cell boundary.

Represent the standard dynamic program for computing Fréchet distance so that it operates column by column, grouping propagation of reachable intervals between vertically aligned cells. Call that procedure $\mathrm{Prop}(R)$, where R is the *reachability column* for point i and the result is the reachability column for point $i+1$ on one of the curves. The reachability column is a set of intervals on a vertical line, indicating the points in the free-space diagram that are reachable from the lower left corner with a monotone path.

Define $\mathrm{Reach}(i, s)$ to be the reachability column induced by q_i^s, where a point is in a reachability interval if it can be reached by a monotone path for some realisation of the previous points. Then we iterate over all the realisations of the previous column, getting precise cells, and propagate the reachable intervals as in the precise Fréchet distance algorithm:

$$\mathrm{Reach}(i+1, \ell) = \mathrm{Feas}(i+1, \ell) \cap \bigcup_{\ell' \in [k]} \mathrm{Prop}(\mathrm{Reach}(i, \ell')) \,.$$

For the column corresponding to U_1, we set one reachable interval of a single point at the bottom for all realisations p_1^s for which $\|q_1^s - p_1\| \le \delta$.

▶ **Theorem 21.** *Given an indecisive curve $\mathcal{V} = \langle V_1, \dots, V_n\rangle$ with k options per point, a precise curve $\pi = \langle p_1, \dots, p_m\rangle$, and a threshold $\delta > 0$, we can decide if $d_{\mathrm{F}}^{\min}(\pi, \mathcal{V}) \le \delta$ in time $\Theta(mnk^2)$ in the worst case, using $\Theta(mk)$ space. We can also report the realisation of $\mathcal{V}$ realising Fréchet distance at most δ, using $\Theta(mnk)$ space instead. Call the algorithm that solves the problem and reports a fitting realisation* Decider$(\delta, \pi, \mathcal{V})$.

We can extend this result to two indecisive curves. This result highlights a distinction between $d_{\mathrm{F}}^{\min}$ and $d_{\mathrm{F}}^{\max}$ and between different uncertainty models. To tackle $d_{\mathrm{F}}^{\min}$ with general uncertain curves, we develop approximation algorithms.

3.1 Approximation by Grids

Given a polygonal curve π and a general uncertain curve $\mathcal{U}$, in this section we show how to find a curve $\sigma \Subset \mathcal{U}$ such that $d_{\mathrm{F}}(\pi, \sigma) \le (1+\varepsilon) d_{\mathrm{F}}^{\min}(\pi, \mathcal{U})$. This is accomplished by carefully discretising the regions, in effect approximately reducing the problem to the indecisive case, for which we then can use Theorem 21. Missing proofs can be found in the full version [13].

For simplicity we assume the uncertain regions have constant complexity. Throughout, we assume $d_{\mathrm{F}}^{\min}(\pi, \mathcal{U}) > 0$, as justified by the following lemma.

▶ **Lemma 22.** *Let π be a polygonal curve with n vertices, and $\mathcal{U}$ an uncertain curve with m vertices. Then one can determine whether $d_{\mathrm{F}}^{\min}(\pi, \mathcal{U}) = 0$ in $O(mn)$ time.*

We call an algorithm a $(1+\varepsilon)$-*decider* for Problem 7, if when $d_{\mathrm{F}}^{\min}(\pi, \mathcal{U}) \le \delta$, it returns a curve $\sigma \Subset \mathcal{U}$ such that $d_{\mathrm{F}}(\pi, \sigma) \le (1+\varepsilon)\delta$, and when $d_{\mathrm{F}}^{\min}(\pi, \mathcal{U}) > (1+\varepsilon)\delta$, it returns False (in between either answer is allowed). In this section, we present a $(1+\varepsilon)$-*decider* for Problem 7. We make use of the following standard observation.

▶ **Observation 23.** *Given a curve $\pi = \langle \pi_1, \dots, \pi_n\rangle$, call a curve $\sigma = \langle \sigma_1, \dots, \sigma_n\rangle$ an r-perturbation of π if $\|\pi_i - \sigma_i\| \le r$ for all i. Since $\|\pi_i - \sigma_i\|, \|\pi_{i+1} - \sigma_{i+1}\| \le r$, all points of the segment $\sigma_i\sigma_{i+1}$ are within distance r of $\pi_i\pi_{i+1}$. For segments this implies that $d_{\mathrm{F}}(\pi_i\pi_{i+1}, \sigma_i\sigma_{i+1}) \le r$, which implies that $d_{\mathrm{F}}(\pi, \sigma) \le r$ by composing the mappings for all i.*

The high-level idea is to replace $\mathcal{U}$ with the set of grid points it intersects, however, as our uncertain regions may avoid the grid points, we need to include a slightly larger set of points.

▶ **Definition 24.** *Let U be a compact subset of $\mathbb{R}^d$. We now define the set of points $\mathrm{EG}_r(U)$ which we call the* expanded r-grid points *of U.*

Let $B(\sqrt{d}r)$ denote the ball of radius $\sqrt{d}r$, centred at the origin. Let $\mathrm{Thick}(U, r) = U \oplus B(\sqrt{d}r)$, where $\oplus$ denotes Minkowski sum. Let G_r be the regular grid of side length r, and let $\mathrm{GT}_r(U)$ be the subset of grid vertices from G_r that fall in $\mathrm{Thick}(U, r)$. Finally, define

$$\mathrm{EG}_r(U) = \{p \mid p = \arg\min_{q \in U} \|q - x\| \text{ for } x \in \mathrm{GT}_r(U)\}\,.$$

The following lemma argues that one can build a decider by using grids as hinted above. Using this decider, we can solve the corresponding optimisation problem.

▶ **Lemma 25.** *There is a $(1 + \varepsilon)$-decider for Problem 7 with running time $O(mn \cdot (1 + (\Delta/(\varepsilon\delta))^{2d}))$, for $1 \geq \varepsilon > 0$, where $\Delta = \max_i \operatorname{diam}(U_i)$ is the maximum diameter of an uncertain region.*

▶ **Theorem 26.** *Let π be a polygonal curve with n vertices, $\mathcal{U}$ an uncertain curve with m vertices, and $\delta = d_{\mathrm{F}}^{\min}(\pi, \mathcal{U})$. Then for any $1 \geq \varepsilon > 0$, there is an algorithm which returns a curve $\sigma \Subset \mathcal{U}$ such that $d_{\mathrm{F}}(\pi, \sigma) \leq (1 + \varepsilon)\delta$, whose running time is $O(mn(\log(mn) + (\Delta/(\varepsilon\delta))^{2d}))$, where $\Delta = \max_i \operatorname{diam}(U_i)$ is the maximum diameter of an uncertain region.*

If the polygonal curve π is replaced with an uncertain curve $\mathcal{W}$, it easy to argue that this approach extends to approximating $d_{\mathrm{F}}^{\min}(\mathcal{W}, \mathcal{U})$.

3.2 Greedy Algorithm

Here we argue that there is a simple 3-decider for Problem 7, running in near-linear time in the plane. The idea is to greedily and iteratively pick $\sigma_i \in U_i$ so as to allow us to get as far as possible along π. Without any assumptions on $\mathcal{U}$, this greedy procedure may walk too far ahead and get stuck. Thus, here we assume that consecutive U_i are separated to ensure that optimal solutions do not lag too far behind. Here we also assume the U_i are convex, i.e. imprecise, and have constant complexity, as it simplifies certain definitions. In this section, let $\pi = \langle \pi_1, \ldots, \pi_n \rangle$ be a polygonal curve and let $\mathcal{U} = \langle U_1, \ldots, U_m \rangle$ be an imprecise curve.

▶ **Definition 27.** *Call $\mathcal{U}$* γ-separated *if for all $1 \leq i < m$, $\|U_i - U_{i+1}\| > \gamma$ and each U_i is convex. Define an* r-visit *of U_i to be any maximal-length contiguous portion of $\pi \cap (U_i \oplus B(2r))$ which intersects $U_i \oplus B(r)$, where $\oplus$ denotes Minkowski sum. If $\mathcal{U}$ is γ-separated for $\gamma \geq 4r$, then any r-visit of U_i is disjoint from any r-visit of U_j for $i \neq j$, in which case define the* true r-visit *of U_i to be the first visit of U_i which occurs after the true r-visit of U_{i-1}. (For U_1 it is the first r-visit.)*

▶ **Lemma 28.** *If $\mathcal{U}$ is γ-separated for $\gamma \geq 4r$, then for any curve $\sigma \Subset \mathcal{U}$ and any reparametrisations f and g such that $\mathrm{width}_{f,g}(\pi, \sigma) \leq r$, σ_i must map to a point on the true r-visit of U_i for all i.*

For two points α and β on π, let $\alpha \leq \beta$ denote that α occurs before β, and for any points $\alpha \leq \beta$ let $\pi(\alpha, \beta)$ denote the subcurve between α and β.

▶ **Definition 29.** *The* δ-greedy sequence *of π with respect to $\mathcal{U}$, denoted $\mathrm{gs}(\pi, \mathcal{U}, \delta)$, is the longest possible sequence $\alpha = \langle \alpha_1, \ldots, \alpha_k \rangle$ of points on π, where $\alpha_1 = \pi_1$, and for any $i > 1$, α_i is the point furthest along π such that $\|\alpha_i - U_i\| \leq \delta$ and $d_{\mathrm{F}}(\alpha_{i-1}\alpha_i, \pi(\alpha_{i-1}, \alpha_i)) \leq 2\delta$.*

▶ **Observation 30.** *For any $i \leq k$, let $\alpha^i = \langle \alpha_1, \ldots, \alpha_i \rangle$ be the ith prefix of $\mathrm{gs}(\pi, \mathcal{U}, \delta)$. Then $d_{\mathrm{F}}(\alpha^i, \pi(\alpha_1, \alpha_i)) \leq 2\delta$, and $\alpha^i \Subset \mathcal{U}_i \oplus B(\delta)$, where $\mathcal{U}_i \oplus B(\delta) = \langle U_1 \oplus B(\delta), \ldots, U_i \oplus B(\delta) \rangle$.*

The following is the main lemma used to argue the correctness of our greedy approach, and it makes use of helper Lemma 28.

▶ **Lemma 31.** *If $\mathcal{U}$ is 10δ-separated and $d_{\mathrm{F}}^{\min}(\pi,\mathcal{U}) \le \delta$, then $\mathrm{gs}(\pi,\mathcal{U},\delta)$ has length m and $\alpha_m = \pi_n$.*

The following lemma is the only place where we require the points to be in $\mathbb{R}^2$. The proof is interesting and uses a result from Guibas et al. [30].

▶ **Lemma 32.** *For π and $\mathcal{U}$ in $\mathbb{R}^2$, where $\mathcal{U}$ is 10δ-separated, $\mathrm{gs}(\pi,\mathcal{U},\delta)$ is computable in $O(m + n\log n)$ time.*

▶ **Theorem 33.** *Let $\mathcal{U}$ be $10r$-separated for some $r > 0$. There is a 3-decider for Problem 7 with running time $O(m + n\log n)$ in the plane that works for any query value $0 < \delta \le r$.*

Proof. Compute $\mathrm{gs}(\pi,\mathcal{U},\delta)$. If it has length m, then let $\sigma = \langle \sigma_1, \dots, \sigma_m \rangle$ be any curve in $\mathrm{Real}(\mathcal{U})$ such that $\|\sigma_i - \alpha_i\| \le \delta$ for all i. If this occurs and if $\alpha_m = \pi_n$, we output σ as our solution, and otherwise we output False. Thus, the running time follows from Lemma 32.

Observe that if we output a curve σ, then $d_{\mathrm{F}}(\sigma,\pi) \le 3\delta$, using the triangle inequality:

$$d_{\mathrm{F}}(\sigma,\pi) \le d_{\mathrm{F}}(\sigma,\alpha) + d_{\mathrm{F}}(\alpha,\pi) \le \delta + 2\delta = 3\delta\,.$$

Thus, we only need to argue that when $d_{\mathrm{F}}^{\min}(\pi,\mathcal{U}) \le \delta$, a curve is produced, which is immediate from Lemma 31. ◀

4 Algorithms for Upper Bound and Expected Fréchet Distance

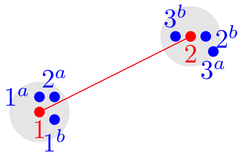
As shown in Section 2.1, finding the upper bound and expected discrete and continuous Fréchet distance is hard even for simple uncertainty models. However, restricting the possible couplings between the curves makes the problem solvable in polynomial time. In this section, we use *indecisive* curves. Define a Sakoe–Chiba time band [44] in terms of reparametrisations of the curves: for a band of width w and all $t \in [0,1]$, if $\phi_1(t) = x$, then $\phi_2(t) \in [x-w, x+w]$. In the discrete case we only couple point i on one curve to points $i \pm w$ on the other curve.

4.1 Upper Bound Discrete Fréchet Distance

First of all, let us discuss a simple setting. Suppose we are given a curve $\sigma = \langle q_1, \dots, q_n \rangle$ of n precise points and $\mathcal{U} = \langle U_1, \dots, U_n \rangle$ of n indecisive points, each of them having ℓ options, so for all $i \in [n]$ we have $U_i = \{p_i^1, \dots, p_i^\ell\}$. We would like to answer the following decision problem: *"If we restrict the couplings to a Sakoe–Chiba band of width w, is it true that $d_{\mathrm{dF}}^{\max}(\mathcal{U},\sigma) \le \delta$ for some given threshold $\delta > 0$?"* So, we want to solve the decision problem for the upper bound discrete Fréchet distance between a precise and an indecisive curve.

In a fully precise setting the discrete Fréchet distance can be computed using dynamic programming [22]. We create a table where the rows correspond to vertices of one curve, say σ, and columns correspond to vertices of the other curve, say π. Each table entry (i,j) then contains a True or False value indicating if there is a coupling between $\sigma[1:j]$ and $\pi[1:i]$ with maximum distance at most δ. We use a similar approach.

Suppose we position $\mathcal{U}$ to go horizontally along the table, and σ to go vertically. Consider an arbitrary column in the table and suppose that we fix the realisation of $\mathcal{U}$ up to the previous column. Then we can simply consider the new column ℓ times, each time picking a different realisation for the new point on $\mathcal{U}$, and compute the resulting reachability. As we do this for the entire column at once, we can ensure consistency of our choice of realisation.



F F	F T	T T
T T	T F	F F

F
T

→

F	T
T	F

→

T
F

Figure 8 Left: An indecisive and a precise curve. Middle: Distance matrix. "T T" in the bottom left cell means $\|1 - 1^a\| \leq \delta$ and $\|1 - 1^b\| \leq \delta$. Right: Computing reachability matrix, column by column. Note two reachability vectors for the second column.

This procedure will give us a set of binary reachability vectors for the new column, each vector corresponding to a realisation. The *reachability vector* is a boolean vector that, for the cell (i, j) of the table, states whether for a particular realisation π of $\mathcal{U}[1 : i]$ the discrete Fréchet distance between π and $\sigma[1 : j]$ is below some threshold δ.

An important observation is that we do not need to distinguish between the realisations that give the same reachability vector: once we start filling out the next column, all we care about is the existence of some realisation leading to that particular reachability vector. So, we can keep a *set* of binary vectors corresponding to reachability in the column.

This procedure was suggested for a specific realisation. However, we can also repeat this for each previous reachability vector, only keeping the unique results. As all the realisation choices happen along $\mathcal{U}$, by treating the table column-by-column we ensure that we do not have issues with inconsistent choices. Therefore, repeating this procedure n times, we fill out the last column of the table. At that point, if any vector has FALSE in the top right cell, then there is some realisation $\pi \Subset \mathcal{U}$ such that $d_{\mathrm{dF}}(\pi, \sigma) > \delta$, and hence $d_{\mathrm{dF}}^{\max}(\mathcal{U}, \sigma) > \delta$.

In more detail, we use two tables, distance matrix D and reachability matrix R. First of all, we initialise the distance matrix D and the reachability of the first column for all possible locations of U_1. Then we fill out R column-by-column. We take the reachability of the previous column and note that any cell can be reached either with the horizontal step or with the diagonal step. We need to consider various extensions of the curve $\mathcal{U}$ with one of the ℓ realisations of the current point: the distance matrix should allow the specific coupling. Assume we find that a certain cell is reachable; if allowed by the distance matrix, we can then go upwards, marking cells above the current cell reachable, even if they are not directly reachable with a horizontal or diagonal step. Then we remember the newly computed vector; we only add distinct vectors. The computation is illustrated in Figure 8; missing details can be found in the full version [13]. We use the following loop invariant to show correctness.

▶ **Lemma 34.** *Consider column i. Every reachability vector of this column corresponds to at least one realisation of $\mathcal{U}[1 : i]$ and the discrete Fréchet distance between that realisation and $\sigma[1 : \min(n, i + w)]$; and every realisation corresponds to some reachability vector.*

▶ **Theorem 35.** *Problem UPPER BOUND DISCRETE FRÉCHET restricted to a Sakoe–Chiba time band of width w on a precise curve and an uncertain curve on indecisive points with ℓ options, both of length n, can be solved in time $\Theta(4^w \ell n \sqrt{w})$ in the worst case.*

Now we extend our previous result to the setting where both curves are indecisive, so instead of σ we have $\mathcal{V} = \langle V_1, \ldots, V_n \rangle$, with, for each $j \in [n]$, $V_j = \{q_j^1, \ldots, q_j^\ell\}$. Suppose we pick a realisation for curve $\mathcal{V}$. Then we can apply the algorithm we just described. We cannot run it separately for every realisation; instead, note that the part of the realisation that matters for column i is the points from $i - w$ to $i + w$, since any previous or further points are outside the time band. So, we can fix these $2w + 1$ points and compute the column. We do so for each possible combination on these $2w + 1$ points.

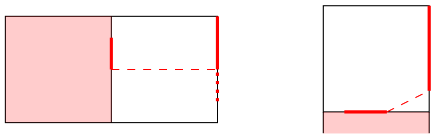


Figure 9 Reachability adjustments. Left: Although the dotted interval is free according to the distance matrix, only the solid interval is reachable from the cell on the left with a monotone path, assuming the cell on the left is free. Right: The full interval that is marked as free is reachable.

▶ **Theorem 36.** *Suppose we are given two indecisive curves of length n with ℓ options per indecisive point. Then we can compute the upper bound discrete Fréchet distance restricted to a Sakoe–Chiba band of width w in time $\Theta(4^w \ell^{2w+1} n \sqrt{w})$.*

4.2 Expected Discrete Fréchet Distance

To compute the expected discrete Fréchet distance with time bands, we need two observations:

1. For any two precise curves, there is a single threshold δ where the answer to the decision problem changes – a *critical value*; it is the distance between two points on the curves.
2. We can modify our algorithm to store associated counts with each reachability vector, obtaining the fraction of realisations that yield the answer TRUE for a given threshold δ.

We can execute our algorithm for each critical value and get the cumulative distribution function $\mathbb{P}(d_{\text{dF}}(\pi, \sigma) > \delta)$ for $\pi, \sigma \Subset_{\mathbb{U}} \mathcal{U}, \mathcal{V}$. Using the fact that the cumulative distribution function is a step function, we compute $d_{\text{dF}}^{\mathbb{E}}$.

▶ **Theorem 37.** *Suppose we are given two indecisive curves of length n with ℓ options per indecisive point. Then we can compute the expected discrete Fréchet distance when constrained to a Sakoe–Chiba band of width w in time $\Theta(4^w \ell^{2w+3} n^2 w^2)$ in the worst case.*

4.3 Continuous Fréchet Distance

We can adapt our time band algorithms to handle continuous Fréchet distance. Instead of the boolean reachability vectors, we use vectors of *free space* cells, introduced by Alt and Godau [6, 27]. We need to now store reachability intervals on cell borders (see Figure 9). The number of these intervals is limited: for any cell, the upper value of the interval is defined by the distance matrix, so yielding at most ℓ^2 values; the lower value of the interval is defined by the distance matrix or by one of the cells from the same row, yielding exponential dependency on w. However, the algorithm is still polynomial-time in n.

We can also store the associated counts. We then find critical values, in line with those arising in precise curve Fréchet distance [6]. This way we adapt our algorithm for computing expected distance to continuous case, and it runs in time polynomial in n for fixed w and ℓ, as desired. Further details are provided in the full version [13].

▶ **Theorem 38.** *Suppose we are given two indecisive curves of length n with ℓ options per indecisive point. Then we can compute upper bound Fréchet distance and expected Fréchet distance restricted to a Sakoe–Chiba band of fixed width w in time polynomial in n.*

References

1 Manuel Abellanas, Ferran Hurtado, Christian Icking, Rolf Klein, Elmar Langetepe, Lihong Ma, Belén Palop, and Vera Sacristán. Smallest color-spanning objects. In *Algorithms – ESA 2001*, volume 2161 of *Lecture Notes in Computer Science*, pages 278–289, Berlin, Germany, 2001. Springer Berlin Heidelberg. doi:10.1007/3-540-44676-1_23.

2 Pankaj K. Agarwal, Boris Aronov, Sariel Har-Peled, Jeff M. Phillips, Ke Yi, and Wuzhou Zhang. Nearest-neighbor searching under uncertainty II. *ACM Transactions on Algorithms (TALG)*, 13(1):3:1–3:25, December 2016. doi:10.1145/2955098.

3 Pankaj K. Agarwal, Rinat Ben Avraham, Haim Kaplan, and Micha Sharir. Computing the discrete Fréchet distance in subquadratic time. *SIAM Journal on Computing*, 43(2):429–449, 2014. doi:10.1137/130920526.

4 Pankaj K. Agarwal, Alon Efrat, Swaminathan Sankararaman, and Wuzhou Zhang. Nearest-neighbor searching under uncertainty I. *Discrete & Computational Geometry*, 58(3):705–745, July 2017. doi:10.1007/s00454-017-9903-x.

5 Hee-Kap Ahn, Christian Knauer, Marc Scherfenberg, Lena Schlipf, and Antoine Vigneron. Computing the discrete Fréchet distance with imprecise input. *International Journal of Computational Geometry & Applications*, 22(01):27–44, 2012. doi:10.1142/S0218195912600023.

6 Helmut Alt and Michael Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and Applications*, 5(1):75–91, 1995. doi:10.1142/S0218195995000064.

7 Esther M. Arkin, Aritra Banik, Paz Carmi, Gui Citovsky, Matthew J. Katz, Joseph S.B. Mitchell, and Marina Simakov. Selecting and covering colored points. *Discrete Applied Mathematics*, 250:75–86, December 2018. doi:10.1016/j.dam.2018.05.011.

8 Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pages 359–370, Palo Alto, CA, USA, 1994. AAAI Press. doi:10.5555/3000850.3000887.

9 Karl Bringmann. Why walking the dog takes time: Fréchet distance has no strongly subquadratic algorithms unless SETH fails. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 661–670, Piscataway, NJ, USA, August 2014. IEEE. doi:10.1109/FOCS.2014.76.

10 Karl Bringmann, Marvin Künnemann, and André Nusser. Fréchet distance under translation: Conditional hardness and an algorithm via offline dynamic grid reachability. In *Proceedings of the Thirtieth Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2019)*, pages 2902–2921. Society for Industrial and Applied Mathematics, January 2019. doi:10.5555/3310435.3310615.

11 Kevin Buchin, Maike Buchin, and Joachim Gudmundsson. Constrained free space diagrams: A tool for trajectory analysis. *International Journal of Geographical Information Science*, 24(7):1101–1125, July 2010. doi:10.1080/13658810903569598.

12 Kevin Buchin, Maike Buchin, Wouter Meulemans, and Wolfgang Mulzer. Four Soviets walk the dog: Improved bounds for computing the Fréchet distance. *Discrete & Computational Geometry*, 58(1):180–216, 2017. doi:10.1007/s00454-017-9878-7.

13 Kevin Buchin, Chenglin Fan, Maarten Löffler, Aleksandr Popov, Benjamin Raichel, and Marcel Roeloffzen. Fréchet distance for uncertain curves, April 2020. arXiv:2004.11862.

14 Kevin Buchin, Maarten Löffler, Pat Morin, and Wolfgang Mulzer. Preprocessing imprecise points for Delaunay triangulation: Simplified and extended. *Algorithmica*, 61(3):674–693, November 2011. doi:10.1007/s00453-010-9430-0.

15 Kevin Buchin, Tim Ophelders, and Bettina Speckmann. SETH says: Weak Fréchet distance is faster, but only if it is continuous and in one dimension. In *Proceedings of the Thirtieth Annual ACM–SIAM Symposium on Discrete Algorithms (SODA '19)*, pages 2887–2901. Society for Industrial and Applied Mathematics, January 2019. doi:10.5555/3310435.3310614.

16 Maike Buchin, Anne Driemel, and Bettina Speckmann. Computing the Fréchet distance with shortcuts is NP-hard. In *Proceedings of the Thirtieth Annual Symposium on Computational Geometry (SoCG 2014)*, pages 367–376, New York, NY, USA, June 2014. Association for Computing Machinery. doi:10.1145/2582112.2582144.

17 Sandip Das, Partha P. Goswami, and Subhas C. Nandy. Smallest color-spanning object revisited. *International Journal of Computational Geometry & Applications*, 19(5):457–478, October 2009. doi:10.1142/S0218195909003076.

18 Thomas Devogele, Laurent Etienne, Maxence Esnault, and Florian Lardy. Optimized discrete Fréchet distance between trajectories. In *Proc. 6th ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data*, pages 11–19, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3150919.3150924.

19 Anne Driemel and Sariel Har-Peled. Jaywalking your dog: Computing the Fréchet distance with shortcuts. *SIAM Journal on Computing*, 42(5):1830–1866, October 2018. doi:10.1137/120865112.

20 Anne Driemel, Sariel Har-Peled, and Carola Wenk. Approximating the Fréchet distance for realistic curves in near linear time. *Discrete & Computational Geometry*, 48(1):94–127, July 2012. doi:s00454-012-9402-z.

21 Anne Driemel, Herman Haverkort, Maarten Löffler, and Rodrigo I. Silveira. Flow computations on imprecise terrains. *Journal of Computational Geometry (JoCG)*, 4(1):38–78, 2013. doi:10.20382/jocg.v4i1a3.

22 Thomas Eiter and Heikki Mannila. Computing discrete Fréchet distance. Technical Report CD-TR 94/64, Technishe Universität Wien, April 1994. URL: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf [cited 2019-04-23].

23 Chenglin Fan, Jun Luo, and Binhai Zhu. Tight approximation bounds for connectivity with a color-spanning set. In *Algorithms and Computation (ISAAC 2013)*, volume 8283 of *Lecture Notes in Computer Science*, pages 590–600, Berlin, Germany, 2013. Springer Berlin Heidelberg. doi:10.1007/978-3-642-45030-3_55.

24 Chenglin Fan and Benjamin Raichel. Computing the Fréchet gap distance. In *33rd International Symposium on Computational Geometry (SoCG 2017)*, volume 77 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 42:1–42:16, Dagstuhl, Germany, 2017. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.SoCG.2017.42.

25 Chenglin Fan and Binhai Zhu. Complexity and algorithms for the discrete Fréchet distance upper bound with imprecise input, February 2018. arXiv:1509.02576v2.

26 Omrit Filtser and Matthew J. Katz. Algorithms for the discrete Fréchet distance under translation. In *16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2018)*, volume 101 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.SWAT.2018.20.

27 Michael Godau. A natural metric for curves: Computing the distance for polygonal chains and approximation algorithms. In *STACS 91: Proceedings of 8th Annual Symposium on Theoretical Aspects of Computer Science*, volume 480 of *Lecture Notes in Computer Science*, pages 127–136, Berlin, Germany, 1991. Springer Berlin Heidelberg. doi:10.1007/BFb0020793.

28 Chris Gray, Frank Kammer, Maarten Löffler, and Rodrigo I. Silveira. Removing local extrema from imprecise terrains. *Computational Geometry*, 45(7):334–349, 2012. doi:10.1016/j.comgeo.2012.02.002.

29 Joachim Gudmundsson, Majid Mirzanezhad, Ali Mohades, and Carola Wenk. Fast Fréchet distance between curves with long edges. *International Journal of Computational Geometry & Applications*, 29(2):161–187, 2019. doi:10.1142/S0218195919500043.

30 Leonidas J. Guibas, John E. Hershberger, Joseph S. B. Mitchell, and Jack S. Snoeyink. Approximating polygons and subdivisions with minimum-link paths. *International Journal of Computational Geometry & Applications*, 3(4):383–415, 1993. doi:10.1142/S0218195993000257.

31 Sariel Har-Peled and Benjamin Raichel. The Fréchet distance revisited and extended. *ACM Transactions on Algorithms (TALG)*, 10(1):3:1–3:22, January 2014. doi:10.1145/2532646.

32 Allan Jørgensen, Jeff Phillips, and Maarten Löffler. Geometric computations on indecisive points. In *Algorithms and Data Structures (WADS 2011)*, volume 6844 of *Lecture Notes in Computer Science*, pages 536–547, Berlin, Germany, 2011. Springer Berlin Heidelberg. doi:10.1007/978-3-642-22300-6_45.

33 Eamonn Keogh and Chotirat Ann Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3):358–386, 2005. doi:10.1007/s10115-004-0154-9.

34 Christian Knauer, Maarten Löffler, Marc Scherfenberg, and Thomas Wolle. The directed Hausdorff distance between imprecise point sets. *Theoretical Computer Science*, 412(32):4173–4186, 2011. doi:10.1016/j.tcs.2011.01.039.

35 John Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, August 2009. doi:10.1007/s00779-008-0212-5.

36 Maarten Löffler. *Data Imprecision in Computational Geometry*. PhD thesis, Universiteit Utrecht, October 2009. URL: https://dspace.library.uu.nl/bitstream/handle/1874/36022/loffler.pdf [cited 2019-06-15].

37 Maarten Löffler and Wolfgang Mulzer. Unions of onions: Preprocessing imprecise points for fast onion decomposition. *Journal of Computational Geometry (JoCG)*, 5(1):1–13, 2014. doi:10.20382/jocg.v5i1a1.

38 Maarten Löffler and Jack Snoeyink. Delaunay triangulations of imprecise points in linear time after preprocessing. *Computational Geometry: Theory and Applications*, 43(3):234–242, 2010. doi:10.1016/j.comgeo.2008.12.007.

39 Maarten Löffler and Marc van Kreveld. Largest and smallest tours and convex hulls for imprecise points. In *Algorithm Theory – SWAT 2006*, volume 4059 of *Lecture Notes in Computer Science*, pages 375–387, Berlin, Germany, 2006. Springer Berlin Heidelberg. doi:10.1007/11785293_35.

40 Anil Maheshwari, Jörg-Rüdiger Sack, Kaveh Shahbaz, and Hamid Zarrabi-Zadeh. Fréchet distance with speed limits. *Computational Geometry*, 44(2):110–120, 2011. doi:10.1016/j.comgeo.2010.09.008.

41 Jian Pei, Bin Jiang, Xuemin Lin, and Yidong Yuan. Probabilistic skylines on uncertain data. In *Proceedings of the 33rd International Conference on Very Large Data Bases*, pages 15–26. VLDB Endowment, September 2007. doi:10.5555/1325851.1325858.

42 Dieter Pfoser and Christian S. Jensen. Capturing the uncertainty of moving-object representations. In *Advances in Spatial Databases*, volume 1651 of *Lecture Notes in Computer Science*, pages 111–131, Berlin, Germany, June 1999. Springer Berlin Heidelberg. doi:10.1007/3-540-48482-5_9.

43 A. Prasad Sistla, Ouri Wolfson, Sam Chamberlain, and Son Dao. Querying the uncertain position of moving objects. In *Temporal Databases: Research and Practice*, volume 1399 of *Lecture Notes in Computer Science*, pages 310–337. Springer Berlin Heidelberg, Berlin, Germany, 1998. doi:10.1007/BFb0053708.

44 Hiroaki Sakoe and Seibi Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, February 1978. doi:10.1109/TASSP.1978.1163055.

45 Jeff Sember and William Evans. Guaranteed Voronoi diagrams of uncertain sites. In *Proceedings of the 20th Canadian Conference on Computational Geometry (CCCG 2008)*, pages 203–206, 2008. URL: http://cccg.ca/proceedings/2008/paper50full.pdf.

46 Subhash Suri, Kevin Verbeek, and Hakan Yıldız. On the most likely convex hull of uncertain points. In *Algorithms – ESA 2013*, volume 8125 of *Lecture Notes in Computer Science*, pages 791–802, Berlin, Germany, 2013. Springer Berlin Heidelberg. doi:10.1007/978-3-642-40450-4_67.

47 Marc van Kreveld, Maarten Löffler, and Joseph S. B. Mitchell. Preprocessing imprecise points and splitting triangulations. *SIAM Journal on Computing*, 39(7):2990–3000, May 2010. doi:10.1137/090753620.

48 Man Lung Yiu, Nikos Mamoulis, Xiangyuan Dai, Yufei Tao, and Michail Vaitis. Efficient evaluation of probabilistic advanced spatial queries on existentially uncertain data. *IEEE Transactions on Knowledge and Data Engineering*, 21(1):108–122, 2009. doi:10.1109/TKDE.2008.135.