Approximate Matching of Curves to Point Sets

Paul Accisano *   Alper Üngör *

Abstract

Let P be a polygonal curve in $\mathbb{R}^d$ of length n, and S be a point set of size k. The Curve/Point Set Matching problem consists of finding a polygonal curve Q on S such that the Fréchet distance from P is less than a given ε. We consider eight variations of the problem based on the distance metric used and the omittability and repeatability of the points. We provide closure to a recent series of complexity results for the case where S consists of precise points. We also formulate a more realistic version of the problem that takes into account measurement errors and attempts to match a given curve to a set of imprecise points. We show that all three variations of the problem that are in P when S consists of precise points become NP-complete when S consists of imprecise points. Finally, we present a 3-factor approximation algorithm for a version of the problem.

1 Introduction

We study the problem of curve and point set matching, using the Fréchet distance as the similarity metric. Given a point set and a polygonal curve, the goal is to connect the points into a new polygonal curve that is similar to the given curve. Formally, given a polygonal curve P of length n, a point set S of size k, and a real number $\varepsilon > 0$, determine whether there exists a polygonal curve Q on a subset of the points of S such that $\delta_{\mathcal{F}}(P, Q) \leq \varepsilon$.

The version of this problem in which the points to be matched are precise has been well studied in the literature [1, 11, 13], and we refer to it as the ***Curve/Point Set Matching (CPSM)*** problem. However, the limitations of modern scanner technology suggest that a more realistic version of this problem would be to consider the input points as imprecise regions. Here, we introduce this new version of the problem and refer to it as the ***Curve/Imprecise Point Set Matching (CIPSM)*** problem.

Eight versions of the original CPSM problem can be classified based on whether the use of all points is enforced, whether points are allowed to be visited more than once, and whether the Fréchet distance metric used is discrete or continuous. Table 1 summarizes the versions and their complexity classes.

*Dept. of Computer & Info. Sci. & Eng., University of Florida, `{accisano, ungor}@cise.ufl.edu`

		Discrete	Continuous
Subset	Unique	NP-C [13]	**NP-C**
	Non-Unique	P [13]	P [11]
All-Points	Unique	NP-C [13]	NP-C [1]
	Non-Unique	P [13]	NP-C [1]

Table 1: Complexity results for versions of the CPSM.

Our results. At an earlier workshop, we presented preliminary results showing that the CPSM problem is NP-complete when coverage of all points is enforced [1]. Here we extend this work by also proving the last remaining open question in Table 1, the Unique Subset version (bold), is also NP-complete (Section 4). Also, we formulate and present complexity results on matching a curve to imprecise points using Fréchet distance (CIPSM). Naturally, all the versions shown in Table 1 that are NP-complete are also NP-complete for their imprecise variations. However, we show that the other three versions are also NP-complete (Sections 5, 6). Finally, we present an approximation algorithm for one of the NP-complete CPSM variants.

2 Previous Work

The basic Fréchet distance problem asks, given two geometric objects of complexity n and m and a real number $\varepsilon > 0$, is the Fréchet distance $\delta_{\mathcal{F}}$ between the two objects less than ε? When the objects are curves, Alt and Godau [5] showed that problem can be solved in $O(nm)$. Later, Alt et al. [4] showed that if the two objects are a curve and a graph, the problem can be solved in $O(nm \log(m))$. When the input graph is a clique, their problem becomes the Continuous Non-unique Subset version of the CPSM problem. Maheshwari et al. presented an algorithm in [11] that decides this version of the problem in time $O(nk^2)$, improving on the result in [4] by a log factor. They also showed that the curve of minimal Fréchet distance can be computed in time $O(nk^2 \log(nk))$ using parametric search.

Wylie [13] also explored the CPSM problem from the perspective of *discrete* Fréchet distance, which only takes into account the distance at the vertices along the curves. They showed that the non-unique versions were solvable in $O(nk)$ time, and the unique versions were NP-complete, as listed in Table 1. However, as we later note in Section 6, the Discrete Non-Unique Subset version is equivalent to the directed Hausdorff distance problem between the curve vertices and the point set, which implies an $O((n+k)\log k)$ algorithm.

A typical scenario in geometric applications occurs when there exist measurement errors or finite precision computations. In such cases, it makes sense to integrate data imprecision into the formulation of the geometric problem [3, 10]. Related to our work, Ahn et al. [3] recently studied discrete Fréchet distance between two polygonal chains with imprecise vertices. Even when we limit ourselves to discrete Fréchet distance, this differs from our problem where only one curve is given. The CIPSM problem turns out to be hard, while their version of the problem admits a polynomial time solution.

3 Preliminaries

Let the continuous function $P : [0,1] \to \mathbb{R}^d$ be a polygonal curve composed of n segments in $\mathbb{R}^d$, denoted by $(P_1, P_2, \ldots, P_n)$. Let Q be another curve defined analogously. The *Fréchet distance* between P and Q is defined as $\delta_{\mathcal{F}}(P,Q) = \inf_{\sigma,\tau} \max_{t\in[0,1]} \|P(\sigma(t)), Q(\tau(t))\|$, where $\sigma, \tau : [0,1] \to [0,1]$ range over all continuous non-decreasing surjective functions [8]. The region of points at most ε distance away from a segment P_i is referred to as the *cylinder* of P_i, following the notation of [11]. Given a point s within the cylinder of P_i, we define $\mathrm{L}_i^\varepsilon(s)$ to be the earliest occurring point on P_i which is at most ε distance away from s, and $\mathrm{R}_i^\varepsilon(s)$ to be the latest.

Let Q be a polygonal curve whose vertices are in S and whose Fréchet distance from P is at most ε. For some vertex $s \in S$, Q is said to *visit* a point $s \in S$ at segment i if there exist subcurves P' and Q' such that Q' ends at s, P' ends at some point $p \in P_i$, and $\delta_{\mathcal{F}}(P', Q') \le \varepsilon$. A point $s \in S_i$ is said to be *reachable* at i if there exists a curve that visits it at i, and the pair (s, p) is called a *feasible pair*.

As in [9], we use $\tilde{S}$ to denote a set of *imprecise points*, which are regions in $\mathbb{R}^d$. A *realization* S of $\tilde{S}$ is a set of points such that there exists a surjective function $R : \tilde{S} \to S$ with $R(\tilde{s}) \in \tilde{s}$ for all $\tilde{s} \in \tilde{S}$.

4 Continuous Unique Subset CPSM Complexity

We now show that the Continuous Unique Subset version of the CPSM problem is NP-complete. Our reduction is from the (3,B2)-SAT problem, a variant of the famous 3-SAT problem in which each literal, positive and negative, is restricted to occur exactly twice. This variant was shown to be NP-complete in [6]. We note that our reduction would also work from the standard 3-SAT problem, but we use (3,B2)-SAT to simplify the construction. Let Φ be a formula given as input to the (3,B2)-SAT problem. We construct a polygonal curve P and a point set S such that Φ is satisfiable if and only if there exists a vertex-unique polygonal curve Q with Fréchet distance at most ε from P. A curve is *vertex-unique* if it has no shared vertices.

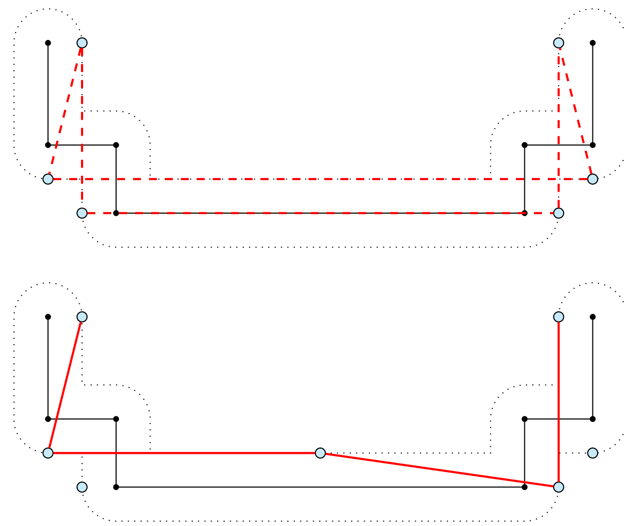


Figure 1: Normally, once the curve starts down one path, changing to the other is impossible (top) unless an extra point on the cylinder boundaries exists (bottom).

Our reduction makes use of a gadget we call *separation corners*, a special version of which was introduced in [1]. These corner constructs force a choice between two possible paths in S, allowing the effects of binary choice to be propagated to other parts of the construction (Figure 1 Top). Ordinarily, there exist only two path possibilities, and once the first corner point is decided, the curve is fully determined until the end of the loop. However, an extra point on the cylinder boundary allows the curve to change tracks to the other path possibility (Figure 1 Bottom). We will use this property when constructing the clause section of the construction.

We will first create a series of small chains, each consisting of two separation corners, laid out horizontally. These chains will represent the variables of Φ, and the four corner points used in the separation corners will represent the four literal instances of the variable. Then, we will create a separation corner *loop* for each clause. However, instead of allowing both possible paths, we will force one of the two to be chosen. At the end of the loop, we will force the chosen path to terminate in a dead end. The loop will be arranged so that the literal points corresponding to the literals used in the clause provide an opportunity for the curve to "change tracks" and avoid the dead end. Since points cannot be used more than once, a literal point will only be available for use to change tracks if it was not already used in the initial variable assignment path. Thus, there exists a path that can traverse the entire curve if and only if Φ has a satisfying assignment.

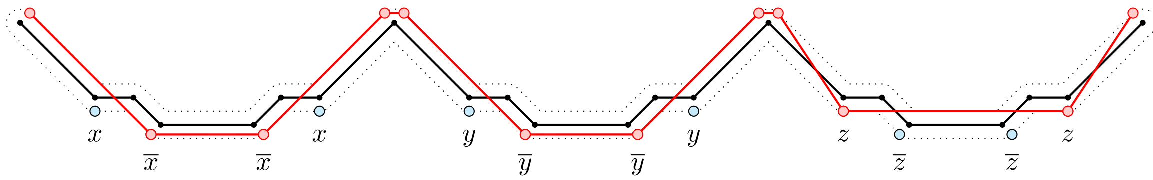


Figure 2: The variable section of the construction for a formula with three variables. The curve corresponding to the assignment TRUE, TRUE, FALSE is shown.

4.1 Construction: Variable Section

The construction is composed of two sections: one for the variables and one for the clauses. We begin with the variable section. The construction starts with a set of separation corners for each variable laid out as in Figure 2, creating two path alternatives for each variable. The two path possibilities will correspond to TRUE or FALSE assignments to that variable. Note that in order to traverse this part of the construction, either the inner or outer corner points *must* be visited. We refer to these points as literal points, as they will represent the literals of Φ. The outer corner points of each variable construct will be referred to as the positive-points, and the inner corner points will be the negative-points.

The purpose of the variable section is to "use up" the literal points corresponding to whichever TRUE/FALSE value is *not* assigned to the variable, leaving the points corresponding to the actual variable value for later use by the clause section. Figure 2 shows how an assignment to the variables of Φ maps to a traversal of the variable section in the construction. Variables assigned to TRUE take the inner path, leaving the outer points available for use later, while variables assigned to FALSE take the outer path, leaving the inner points available.

4.2 Construction: Clause Section

We next create the clause section of the construction, appending it to the variable section. We begin by adding a separation corner loop. However, we leave out one of the two corner points in the first separation corner. This will force the curve to pick a specific possibility and remove the option to pick the other. Next, we place more separation corners, arranging the loop so that the three literal points corresponding to the clause's literals are exactly on the cylinder boundaries of the segments. Once this is done, we remove another point from the next separation corner in the loop corresponding to the path that was forced earlier, creating a dead end. The only way to proceed will be to use one of the literal points to change tracks before the dead end is reached. If no literal point is available for use, then the clause is not satisfied and there will be no way for the curve to proceed without increasing the Fréchet distance.

Figure 3 demonstrates a single clause loop. Note that the first and last separation corners of the clause loop are missing a corner point. Since the corresponding literal points are already used, there is no place to switch, and the solid curve cannot continue because of the missing point at the end. However, if the value of variable y is changed from TRUE to FALSE, the corresponding literal point is free to be used by the clause loop and escape the dead end. The dashed curve shows this configuration.

This process is then repeated for every clause, with a dead end separation corner between each clause loop. The full construction is therefore only traversable if every clause loop has a point at which it can switch tracks, which corresponds to a satisfying assignment. Figure 6 in the Appendix shows an example of a completed construction.

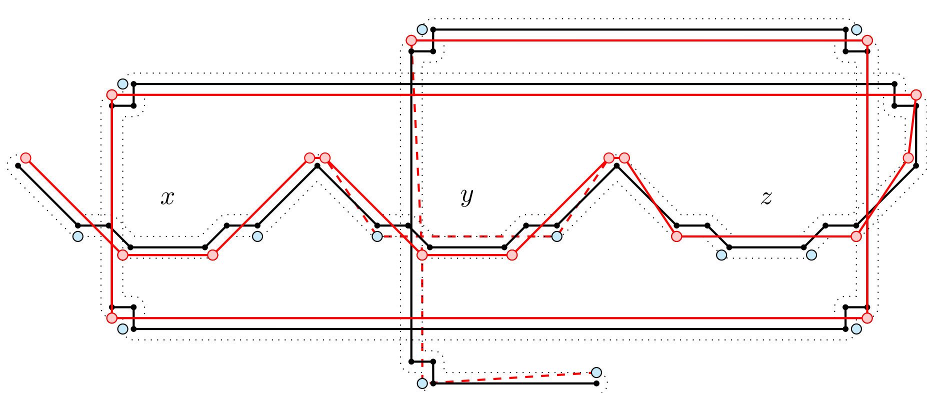


Figure 3: A clause loop for the clause $(\overline{x} \vee \overline{y} \vee z)$.

4.3 Hardness Result

Lemma 1 *There exists a vertex-unique polygonal path Q on S with $\delta_{\mathcal{F}}(P,Q) \le \varepsilon$ if and only if the formula Φ is satisfiable.*

Proof. ($\rightarrow$) Assume Φ has a satisfying assignment. The variable portion of the construction always has a vertex-unique polygonal path Q on S with $\delta_{\mathcal{F}}(P,Q) \le \varepsilon$; assume each variable gadget is chosen according to the satisfying assignment. Then each clause loop will have at least one literal point that can be used to change tracks before its dead end is reached. After all clause loops have been traversed, the path will have Fréchet distance less than ε from P.

($\leftarrow$) Assume Φ does not have a satisfying assignment. Then, no matter how the initial variable portion of the construction is traversed, there will be at least one clause loop which will not be able to use any literal point it passes. Once the end of the clause loop is reached, there will be no way to continue the path without increasing the Fréchet distance beyond ε. □

The variable section contains two separation corners for each variable, and the clause section contains six separation corners for each clause, so the construction is clearly of polynomial size. This leads to the following result.

Theorem 2 *The Unique Subset Continuous CPSM Problem is NP-complete.*

5 Continuous Non-unique Subset CIPSM

In the CIPSM problem, we are given a curve P and an imprecise point set $\tilde{S}$, and the goal is to find a realization S on which there exists a curve with Fréchet distance at most ε from P. For simplicity, we will treat the imprecise points as line segments, but we observe that all results trivially extend to other regions. The CIPSM problem has eight versions corresponding to the eight versions of the CPSM problem. However, note that the CPSM is a special case of the CIPSM in which the diameter of the imprecise points happens to be zero. This shows that the five NP-complete versions of the CPSM imply the NP-completeness of their corresponding CIPSM versions. As such, we focus on the remaining versions.

Here, we show that the Continuous Non-unique Subset CIPSM problem is NP-complete, using a reduction similar to the one presented in the previous section. A key property of the construction in Section 4 is that after the variable section has been traversed, exactly two of the four corner points of each variable remain usable by the clause section. This is due to the fact that points cannot be reused, and two of the four points must be used to traverse the variable section. To adapt the reduction to the Non-unique CIPSM problem, we simply connect the two points of each corner into a single imprecise segment (Figure 4). Since each imprecise segment must resolve to a single point, and since points can be used more than once in this version of the problem, the end result after traversing the variable section is exactly the same: two points for each variable are available for the clause section to choose, either in the positive literal positions or the negative literal positions. Thus, instead of "using up" corner points, we are "making them available" for the clause loops to use to escape their dead ends.

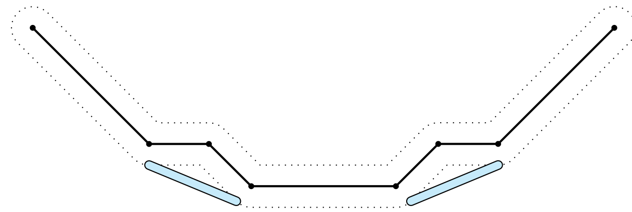


Figure 4: The two corner points of the variable section's corners are replaced with a single segment joining them.

We have modeled imprecise points as line segments for simplicity. However, the model can easily be extended to disks by placing them so that they are tangent to the appropriate cylinders at the appropriate locations. Other shapes can also be positioned to correctly intersect the cylinders by aligning the cylinder boundaries at their extremal points. Since the entire construction is scalable, there is no danger of being forced to place imprecise points close enough to interfere with each other. This leads to the following theorem.

Theorem 3 *The Continuous Non-unique Subset CIPSM is NP-complete.*

6 Discrete CIPSM Problem

In this section, we study the CIPSM under discrete Fréchet distance, a variation of the standard Fréchet distance that only takes into account distance at the curve vertices [7]. For two curves P and Q of lengths n and m respectively, a *paired walk* or *coupling sequence* is a pair of integer sequences $(a_1,b_1),\ldots,(a_k,b_k)$, $k \ge \max(n,m)$, with the properties that $(a_1,b_1) = (1,1), (a_k,a_k) = (n,m)$, and for all i, $(a_{i+1},b_{i+1}) \in \{(a_i+1,b_i),(a_i,b_i+1),(a_i+1,b_i+1)\}$. Let W be the set of all paired walks for P and Q. Then the *discrete Fréchet distance* can be defined as $\delta_{\mathrm{F}} = \min_{(a,b)\in W} \max_i \|P_{a_i}, Q_{b_i}\|$.

Since the edges of the given curve have no impact on the discrete Fréchet distance, and since we are allowed to visit the points of $\tilde{S}$ in any order, the discrete Fréchet distance between P and a given realization of $\tilde{S}$ is the same for any curve with the same vertex set as P. Thus, the Discrete Subset CPSM can be restated as follows: does every ε-ball around the vertices of P contain at least one point from S? This is equivalent to the directed Hausdorff distance problem, which is to determine if $\max_{p\in V(P)} \min_{s\in S} \|p - s\| \leq \varepsilon$. This problem has already been studied under imprecision in [9], and all the results apply. Namely, the Discrete Subset CIPSM is NP-complete, but the optimization version can be 4-approximated in $O((n + k)^3 \log^2(n + k))$. The problem can also be solved exactly $O((n + k)^3)$ if the imprecise points are circular, disjoint, and sufficiently large relative to the optimal discrete Fréchet distance.

Furthermore, the All-points version can be reduced to the Subset version. For a given curve and point set, if there is no solution to the Subset version, then there is clearly no solution to the All-points version either. But, given a solution to the Subset version, one can easily determine if there is a solution to the All-points version. Every vertex of P must have at least one associated point of $\tilde{S}$ from the output curve Q within distance ε. As long as the unused imprecise points each overlap with the ε-ball around some vertex of P, the output curve Q can jump from that vertex's associated point in $\tilde{S}$ and visit the unused points close to that vertex before continuing. Thus, the All-points version is also NP complete.

Theorem 4 *The Discrete Non-Unique CIPSM problems, both Subset and All-Points, are NP-Complete, even when the given curve P is simple.*

7 An Approximation Algorithm for CPSM

In this section, we consider the optimization version of the Continuous Non-Unique All-Points CPSM and detail an approximation algorithm for it. To do so, we develop an exact algorithm for a restricted version of the problem, which also serves as a 3-approximation to the unrestricted version.

The main combinatorial challenge of the All-Points version of the problem stems from the fact that a point in cylinders of multiple segments can be visited at any one of the segments. To remove this challenge, we introduce an additional restriction to the problem; we enforce that each point in S be visited at its closest segment. Visiting points at other segments is also allowed, but each point must be visited at its nearest segment even if it is also visited at another one. We call a curve that respects this condition *NS-compliant* (Nearest Segment compliant).

7.1 NS-Compliant Algorithm

We follow the parametric search paradigm by first developing an algorithm for the decision version of the problem. Let S_i be S intersected with the cylinder of P_i. For convenience, let S_0 and S_{n+1} be the members of S that are within ε of the start and end points of P respectively. For a given $s \in S$, let P^s be the segment of P closest to s. Let the *essential points* of P_i, denoted by S_i^*, be the set $\{s \in S \mid P^s = P_i\}$.

An obvious preprocessing step is to confirm that all points of S are a member of some S_i. Another is to confirm that S_0 and S_{n+1} are non-empty. If either of these conditions are false, we can stop and return FALSE immediately. Note that, under this assumption, $S_i^* \subset S_i$.

Per our restriction of NS-compliance, every point must be visited at its closest cylinder. However, it may be necessary to visit points in other cylinders as well. For example, even if $S_i^* = \emptyset$, some point $s \in S_i$ may need to be visited in order to stay close to the given curve and reach future points. If we think of single points in multiple cylinders as if they were separate points, then there are two types: points we must visit, and points we may skip. In this way, the problem is very similar to the Subset version, in which all points are the latter type.

In order to visit every point in a segment's essential set, care must be taken regarding the first and last points visited for a given segment. Let $s \in S_i$ be the first point visited in P_i, which may not be an essential point of P_i. If there exists an essential point s' for which $\mathrm{R}_i^\varepsilon(s')$ comes before $\mathrm{L}_i^\varepsilon(s)$, then it will not be possible to visit s'; the curve has already gone too far and cannot backtrack far enough. By the same token, if $t \in S_i$ is the last point visited in P_i and there exists an essential point t' for which $\mathrm{L}_i^\varepsilon(t')$ comes after $\mathrm{R}_i^\varepsilon(t)$, then t' must not have been visited, because it is too far ahead to have been backtracked from.

To formalize this notion, we say a point t is an *entry point* for P_i if $\mathrm{L}_i^\varepsilon(t)$ comes before $\mathrm{R}_i^\varepsilon(s)$ for all $s \in S_i^*$. Analogously, we say t is an *exit point* if $\mathrm{R}_i^\varepsilon(t)$ comes after $\mathrm{L}_i^\varepsilon(s)$ for all $s \in S_i^*$. Note that, if $S_i^* = \emptyset$, then every point in S_i is an entry and exit point for that segment. In order to ensure that every point in S_i^* can be visited, we must enter each cylinder via an entry point and leave it through an exit point. As long as this is enforced, we can simply visit all the essential points in monotonic order along the segment.

To turn this idea into an algorithm, we adapt the algorithm for the Subset version of the problem given in [11]. We provide a small review here. The first step of the Subset algorithm is to precompute a reachability function $r_i(s, t)$. Let $s \in S_i$ be a point that is reachable at P_i by some feasible curve Q ending in s. Given a point $t \in S$, $r_i(s, t)$ is defined as the largest index $j \geq i$ such that the curve $Q + \overline{st}$ visits t at P_j, or 0 if $Q + \overline{st}$ is not feasible. As proven in [11], t is reachable at P_j for all $i \leq j \leq r_i(s, t)$. Therefore, this value provides reachability information for all pairs of points in S from any segment to any other. To ensure that no essential points are skipped, we must modify $r_i(s, t)$ to obtain a new function $r'_i(s, t)$ with the following properties:

- $r'_i(s, t) = 0$ or i if s is not an exit point for P_i.
- $r'_i(s, t) > i$ only if t is an entry point for $P_{r'_i(s,t)}$.
- For all $i < j < r'_i(s, t)$, $S_j^* = \emptyset$.

The third item ensures that previously stated property of t being reachable at P_j if $t \in P_j$ for $i \le j \le r'_i(s,t)$ still holds; recall that every point in a cylinder with an empty essential set is an entry point. Given $r_i(s,t)$, we define $r'_i(s,t)$ to be the smaller of $r_i(s,t)$ and the index of the first segment after i with a non-empty essential set. It is straightforward to verify that this definition respects the three rules above. Note that $r'_i(s,t)$ can be computed in $O(nk^2)$ time by pre-computing which segments have non-empty essential sets.

Under this modified reachability function, the Subset algorithm decides the NS-compliant problem. Note that, even though the actual curve returned by the Subset algorithm is not guaranteed to visit all points, it will return a curve that enters each cylinder via an entry point and exits via an exit point, which is sufficient to guarantee the existence of an NS-compliant curve. The algorithm pseudo-code is shown in Algorithm 1 below.

Algorithm 1 NS-compliant CPSM (P, S, ε)

```
1: Compute S_i and S_i^* for all i
2: If any point is outside all S_i, return no
3: Compute r_i(s,t) for all 1 ≤ i ≤ n and s,t ∈ S
4: Compute the entry and exit sets for each segment.
5: Modify r_i(s,t) to obtain r'_i(s,t)
6: Apply the Subset algorithm using r'_i(s,t)
7: Return the result
```

Time Complexity. Lines 1 and 2 of Algorithm 1 takes $O(nk)$ time. Lines 3 and 6 take $O(nk^2)$ time [4, 11]. Computing the entry and exit sets on Line 4 requires comparing $O(k)$ candidates with $O(k)$ other points, repeated for each of the n cylinders, so this step takes $O(nk^2)$ time. Finally, computing $r'_i(s,t)$ in Line 5 takes $O(nk^2)$ time, as previously discussed. Thus, the complexity of the algorithm is $O(nk^2)$.

Theorem 5 *Algorithm 1 correctly decides the NS-compliant version of the Continuous Non-unique Subset CPSM in $O(nk^2)$ time.*

With an algorithm for the decision version in hand, the technique of parametric search is employed to find the optimal curve. By analyzing the so-called *free space diagram* of P and each of the $S \times S$ possible segments of Q, as done in [4] and [11], $O(nk^2)$ critical values of ε can be identified. These values can then be sorted, and the decision version of the algorithm can be used to binary search for the smallest value. This technique yields an algorithm with running time $O(nk^2 \log(nk))$.

Theorem 6 *Given a polygonal curve P and a point set S, a polygonal curve Q whose vertices are exactly S with $\delta_{\mathcal{F}}(P,Q)$ at most 3 times that of the optimal can be computed in $O(nk^2 \log(nk))$ time.*

Not only does the algorithm as described compute the NS-compliant problem exactly, but it also 3-approximates the unrestricted version. This is due to the fact that any curve that visits all points in S can be transformed into an NS-compliant curve with Fréchet distance from P at most 3 times the original. We present the proof of this statement in the Appendix. However, if the algorithm yields a Fréchet distance for which no point in S belongs to more than one cylinder, then this solution must also be optimal for the unrestricted version. In a related study [2], we make use of this very algorithm in combination with another algorithm by Wenk [12] to solve the CPSM problem under affine transformations.

References

[1] P. Accisano and A. Üngör. Hardness results on curve/point set matching with Fréchet distance. In *Proc. 29th Euro. Workshop on Comp. Geom.*, 2013.

[2] P. Accisano and A. Üngör. Finding a curve in a point set. *arXiv preprint arXiv:1405.0762*, 2014.

[3] H. Ahn, C. Knauer, M. Scherfenberg, L. Schlipf, and A. Vigneron. Computing the discrete Fréchet distance with imprecise input. *Int. J. of Comp. Geom. & Appl.*, 22(01):27–44, 2012.

[4] H. Alt, A. Efrat, G. Rote, and C. Wenk. Matching planar maps. In *Proc. of the 14th annual ACM-SIAM Symp. on Discr. Algo.*, pages 589–598, 2003.

[5] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *Int. J. of Comp. Geom. & Appl.*, 5(01n02):75–91, 1995.

[6] P. Berman, M. Karpinski, and A. Scott. Approximation hardness of short symmetric instances of MAX-3SAT. *El. Coll. on Comp. Complex.*, (049), 2003.

[7] T. Eiter and H. Mannila. Computing discrete Fréchet distance. Technical Report CD-TR 94/65, Christian Doppler Laboratory, 1994.

[8] G. Ewing. *Calculus of variations with applications.* Dover Pub., New York, 1985.

[9] C. Knauer, M. Löffler, M. Scherfenberg, and T. Wolle. The directed Hausdorff distance between imprecise point sets. *Theoretical Computer Science*, 412(32):4173–4186, 2011.

[10] M. Löffler. *Data imprecision in computational geometry.* PhD thesis, Utrecht University, 2009.

[11] A. Maheshwari, J. Sack, K. Shahbaz, and H. Zarrabi-Zadeh. Staying close to a curve. In *Canadian Conf. on Comp. Geom.*, pages 55–58, 2011.

[12] C. Wenk. *Shape matching in higher dimensions.* PhD thesis, Freie Universität Berlin, 2002.

[13] T. Wylie. Discretely following a curve. In *Comb. Opt. and Appl.*, pages 13–24. Springer, 2013.

Appendix A: Approximation Proof

We now show that any curve can be transformed into an NS-compliant curve while only increasing its Fréchet distance by a factor of 3. This will show that the NS-compliant algorithm given in Section 7 is a 3-factor approximation algorithm for the unrestricted version.

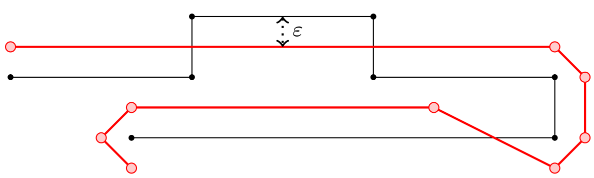
Theorem 7 *Let $\varepsilon = \delta_{\mathcal{F}}(P, Q)$, where P and Q are curves. There exists an NS-compliant curve Q' with the same vertex set as Q such that $\delta_{\mathcal{F}}(P, Q') \leq 3\varepsilon$.*

Proof. Since Q might not be NS-compliant, there may be some vertices of Q that are not visited at their closest segment of P; let $u_1, \ldots, u_m$ be the set of such vertices.

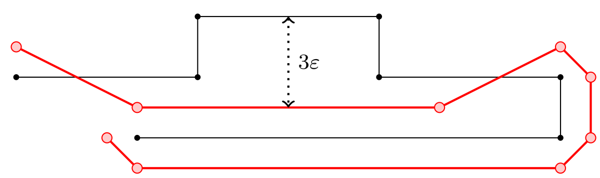
Since each u_i is visited at a segment other than its closest, each can be no further than ε distance away from P^{u_i}. Consider a given u_i and let p be a point in P^{u_i} that is within ε of u_i. Since P and Q have Fréchet distance ε, there must be at least one feasible pair for any point on P; let u'_i be a point (not necessarily a vertex) in Q such that (u'_i, p) is feasible. Then, add u'_i as a new vertex of Q. Note that the distance between u_i and u'_i is at most 2ε. Repeating this process for every u_i yields a new curve Q^*. Since each new vertex has been added along an existing segment, $\delta_{\mathcal{F}}(P, Q) = \delta_{\mathcal{F}}(P, Q^*)$.

Now, merge each u'_i with u_i by translating the former to the position of the latter, yielding a new curve Q' with a potentially different Fréchet distance from P. Let σ and τ be reparameterizations of P and Q^*, and consider the point $Q^*(\tau(t))$ for some $t \in [0, 1]$, which lies on some segment of Q^*. The endpoints of the corresponding segment in Q' may have been displaced up to 2ε, and thus the point $Q'(\tau(t))$ may be up to 2ε away from $Q^*(\tau(t))$. Therefore, $\|P(\sigma(t)), Q'(\tau(t))\|$ can be at most 2ε larger than $\|P(\sigma(t)), Q^*(\tau(t))\|$. Finally, since the Fréchet distance is the infimum of the maximum distance over all reparameterizations, we have that $\delta_{\mathcal{F}}(P, Q') \leq \delta_{\mathcal{F}}(P, Q^*) + 2\varepsilon = 3\varepsilon$. As Figure 5 shows, this bound is realizable.

Since each u'_i was visited at $P^{u'_i}$ in Q^*, the same logic used to show that the Fréchet distance has increased by at most a factor of 3 can again be used to show that each u_i is visited at P^{u_i} in Q'. Thus, we conclude that Q' is NS-compliant. □



(a) Unrestricted optimal solution



(b) NS-compliant optimal solution

Figure 5: If the two middle points are slightly closer to the top segments than the bottom segment, the optimal solution for the NS-compliant version has Fréchet distance 3 times that of the unrestricted version.

Appendix B: Complete Construction Figure

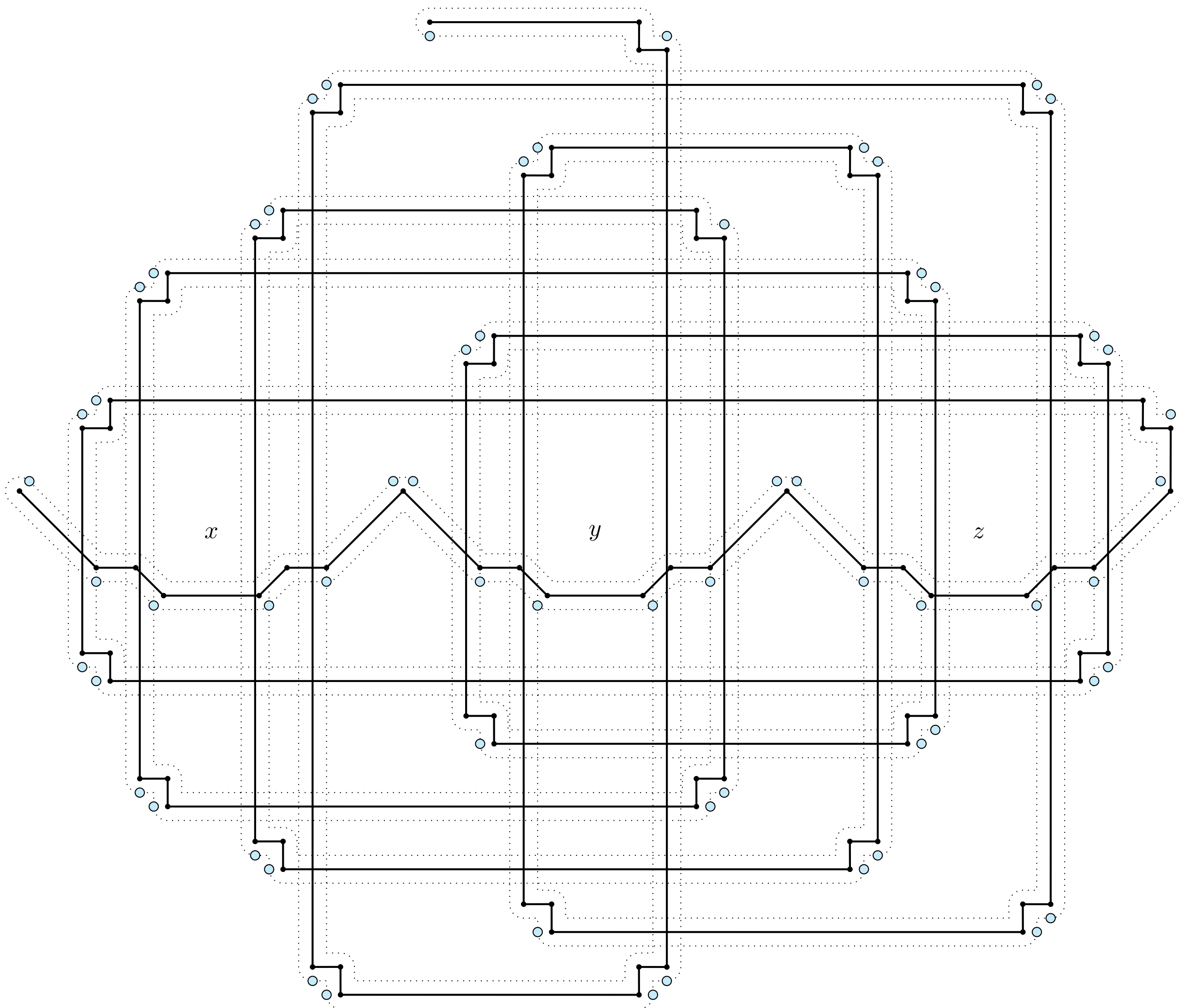


Figure 6: A completed construction for the formula $\Phi = (x \vee y \vee z) \wedge (\overline{x} \vee y \vee \overline{z}) \wedge (\overline{x} \vee \overline{y} \vee z) \wedge (x \vee \overline{y} \vee \overline{z})$.