On the Curve/Point Set Matching Problem

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. We consider the problem of finding a polygonal curve Q on S such that all points in S are visited and the Fréchet distance between Q and P is at most a given ε. We show that this problem is NP-complete, regardless of whether or not the points of S are allowed to be visited more than once.

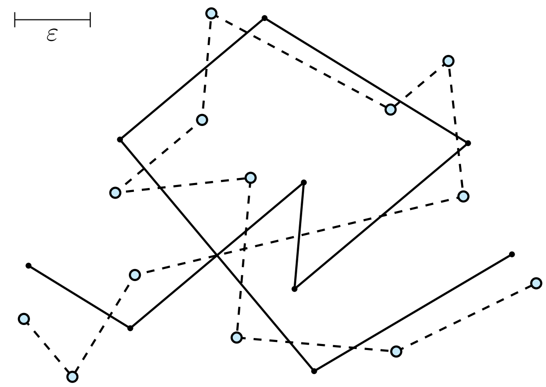

Figure 1: A problem instance and its solution. The input is the solid line and the circle points, and the solution is the dotted line.

1 Introduction

Measuring the similarity between two geometric objects is a fundamental problem in many fields of science and engineering. However, to perform such comparisons, a good metric is required to formalize the intuitive concept of "similarity." Among the many metrics that have been considered, Fréchet distance has emerged as a popular and powerful choice, especially when the geometric objects are curves. Shape matching with Fréchet distance has been applied in many different fields, including handwriting recognition [7], protein structure alignment [5], and vehicle tracking [3].

In this abstract, we consider the basic problem of measuring the similarity of two polygonal curves. However, in our problem, the input is only partially defined. Instead of being given both curves, we are given only one polygonal curve P as well as a point set S. Our problem is to complete this partial input by constructing a polygonal curve Q that best matches the given curve, under the restriction that the constructed curve's vertices are exactly S. We show that, under the Fréchet distance metric, this problem is NP-complete. Figure 1 shows an example problem instance and its solution.

2 Previous Work and New Results

Given two curves $P, Q : [0,1] \to \mathbb{R}^d$, the *Fréchet distance* between P and Q is defined as $\delta_F(P,Q) = \inf_{\sigma,\tau} \max_{t\in[0,1]} \|P(\sigma(t)), Q(\tau(t))\|$, where $\sigma, \tau : [0,1] \to [0,1]$ range over all continuous non-decreasing surjective functions [4].

The decision version of the Fréchet distance problem asks, given two geometric objects and a real number $\varepsilon > 0$, is the Fréchet distance between the two objects less than or equal to ε? Alt and Godau [1] showed that, when the objects in question are polygonal curves of length n and m, this problem can be solved in $O(nm)$ time. They also showed that finding the exact Fréchet distance between the two curves can be done in $O(nm \log(nm))$ time using parametric search.

Maheshwari et al. [6] examined the following variant of the Fréchet distance problem, which we refer to as the Curve/Point Set Matching (CPSM) problem. Given a polygonal curve P of length n, a point set S of size k, and a number $\varepsilon > 0$, determine whether there exists a polygonal curve Q on a subset of the points of S such that $\delta_F(P,Q) \le \varepsilon$. They gave an algorithm that decides this problem in time $O(nk^2)$. They also showed that the curve of minimal Fréchet distance can be computed in $O(nk^2 \log(nk))$ time using parametric search.

Wylie and Zhu [8] also explored the CPSM problem from the perspective of discrete Fréchet distance. In contrast to the continuous Fréchet distance defined above, the discrete Fréchet distance only takes into account the distance at the vertices along the paths. They formulated four versions of the CPSM problem depending on whether or not points in S were allowed to be visited more than once (Unique vs. Non-unique) and whether or not Q was required to visit all points in S at least once (All-Points vs. Subset) They showed that, under the discrete Fréchet distance metric, both non-unique versions were solvable in $O(nk)$ time, and both unique versions were NP-complete.

In this abstract, we show that the Continuous All-

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

Points versions of the CPSM problem, both Unique and Non-unique, are NP-complete. Table 1 shows the eight versions of the problem, with our results highlighted.

		Discrete	Continuous
Subset	Unique	NP-C [8]	Open
	Non-Unique	P [8]	P [6]
All-Pts	Unique	NP-C [8]	NP-C*
	Non-Unique	P [8]	NP-C*

Table 1: Eight versions of the CPSM problem and their complexity classes. New results starred.

3 Reduction Outline

The well-known 3SAT problem takes as input a Boolean formula with clauses of size 3, and asks whether there exists an assignment to the variables that makes the formula evaluate to TRUE. If we restrict the input to formulas in which each literal occurs exactly twice, the problem becomes the (3,B2)-SAT problem. This may seem to be a rather extreme restriction, and, indeed, formulas of this type with less than 20 clauses are always satisfiable. However, despite this restriction, the problem was shown to be NP-complete in [2], and an example of an unsatisfiable formula with 20 clauses was presented.

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 polygonal curve Q whose vertices are exactly S with Fréchet distance at most ε from P. Our construction is somewhat lengthy and involves the construction of a complex gadget, as well as a number proofs about its properties. For this reason, we provide only a brief summary of the construction in this abstract.

First, we construct a gadget consisting of components of P and S that will force any algorithm to choose between two possible polygonal path constructions. The gadget is constructed in such a way that these two choices are the only possible polygonal paths along the gadget's component of S with Fréchet distance at most ε from P. These two path possibilities will correspond to TRUE and FALSE assignments for a given variable.

Then, we create a series of points in S to represent the clauses in Φ, one point for each clause. For each variable, a gadget will be placed so that the pair clause points representing the clauses in which the variable's positive instances occur are only reachable along one of the two curve possibilities, and likewise for the negative instances. Once this has been done for each variable in Φ, any polygonal curve Q whose vertices are exactly S with Fréchet distance at most ε from P will correspond to an assignment to the variables of Φ in which every clause is satisfied, thus making the formula evaluate to TRUE. Furthermore, if no such curve exists, then there can be no such satisfying assignment for Φ.

Given that the problem of determining the Fréchet distance between two given polygonal curves is in P, the CPSM problem is clearly in NP. This leads to our main result.

Theorem 1 *The All-points Continuous CPSM Problem is NP-complete.*

There are still a number of details that have been omitted for the sake of brevity, including the proof that placing the gadgets in the necessary positions is always geometrically possible. These will be included in our full paper. We also plan to explore various generalizations of this problem. For example, the given geometric object could perhaps be a tree or graph, or the point set could be given as imprecise points or regions.

References

[1] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 5(01n02):75–91, 1995.

[2] Piotr Berman, Marek Karpinski, and Alex D. Scott. Approximation hardness of short symmetric instances of MAX-3SAT. *Electronic Colloquium on Computational Complexity (ECCC)*, (049), 2003.

[3] Sotiris Brakatsoulas, Dieter Pfoser, Randall Salas, and Carola Wenk. On map-matching vehicle tracking data. In *Proceedings of the 31st international conference on Very large data bases*, VLDB '05, pages 853–864. VLDB Endowment, 2005.

[4] George Ewing. *Calculus of variations with applications.* Dover Publications, New York, 1985.

[5] M. Jiang, X. Ying, and B. Zhu. Protein structure–structure alignment with discrete frechet distance. *Journal of bioinformatics and computational biology*, 6(01):51–64, 2008.

[6] Anil Maheshwari, Jörg-Rüdiger Sack, Kaveh Shahbaz, and Hamid Zarrabi-Zadeh. Staying close to a curve. In *CCCG*, 2011.

[7] R. Sriraghavendra, K. Karthik, and C. Bhattacharyya. Fréchet distance based approach for searching online handwritten documents. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 461–465. IEEE, 2007.

[8] T. Wylie and B. Zhu. Discretely following a curve (short abstract). In *Computational Geometry: Young Researchers Forum (CG:YRF)*, 2012.