Measuring Length-Preserving Fréchet Correspondence for Graphs in $\mathbb{R}^2$

Kevin Buchin, Brittany Terese Fasy*,
Erfan Hosseini Sereshgi†, and Carola Wenk†

Abstract

Finding a comprehensive way to compare two road networks has been a challenge for quite a long time. One frequently-used comparison uses a discrete (easy to compute) sample from each graph to define a similarity measure based on a correspondence between the samples. In practice, this takes into account both the geometry and the topology of the graphs. However, this sampling approach introduces many variables along the way and can result in undesirable correspondences. In this paper, we introduce a continuous alternative to this method that uses Fréchet distance in a length-preserving setting.

1 Introduction

Finding a comprehensive way to compare two roadmaps or embedded graphs has been a challenge for quite a long time. One discrete way that has been used frequently in the literature is the graph sampling method. Graph sampling was originally defined in [3] as a sampling-based comparison between graphs embedded in the plane that takes into account both the geometry and the topology of the immersed graphs. Graph sampling follows two steps: sampling two embedded graphs with points and matching the corresponding points to each other in a one-to-one manner. The proportion of the matched pairs can be used later to explain the similarities and differences between the two said graphs. Despite its simple definition, Graph sampling introduces many variables along the way that can affect the final results. Sampling interval, sampling function, matching function and whether to take bearing into consideration are some examples of these variables. Furthermore, while graph sampling is a fairly effective approach for map comparison, it is still a discrete method and simply counting the number of matched samples is not a reliable measurement in many cases. A possible artifact of this method is that a single road can be matched to several roads (see Fig. 1). Another common problem with graph sampling on reconstructed maps is finding a suitable ground truth map that only contains roads that are covered in the GPS data. This particular issue results in inconsistent evaluations among experiments in the literature [2].

In this paper we introduce a continuous alternative to this method that uses Fréchet distance in a length-preserving setting and avoids the issues that are mentioned above. For this purpose, one needs to match (continuous) segments of two graphs. There is a related work that matches segments of two curves using Fréchet distance, the so-called partial Fréchet matching [4]. The authors give a polynomial-time dynamic programming algorithm if L_1 or L_∞ are used as the underlying metrics. Here, we put a length-preserving constraint on the matched segments and our goal is to maximize the total length of them. Furthermore, we generalize this definition for graphs in $\mathbb{R}^2$.

2 Measuring Length-Preserving Fréchet Correspondence

In this paper, we consider paths and graphs in $\mathbb{R}^2$, where we define the distance between two points x and y as the two-norm of their difference, denoted $||x - y||$. We denote the (open) metric ball centered at $x \in \mathbb{R}^2$ with radius $\delta \in \mathbb{R}_{\geq 0}$ by: $\mathbb{B}(x, \delta) := \{y \in \mathbb{R}^2 \mid ||x - y|| < \delta\}$.

*Partially supported by National Science Foundation grant 2046730.
†Partially supported by National Science Foundation grant CCF 2107434.

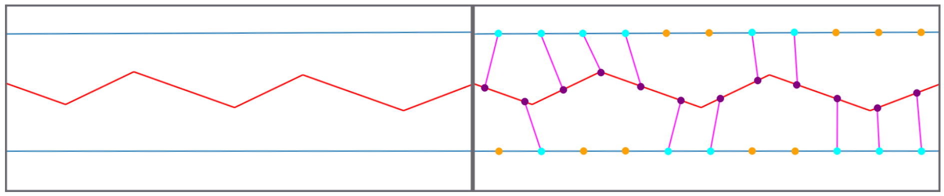

Figure 1: Graph sampling on two road maps G and H shown with blue and red respectively. Cyan and purple points are matched samples on G and H, while orange points indicate samples on G that are not matched. A magenta line shows a matching between a pair of samples.

Let $\mathcal{G}$ be the set of all pairs (G, ϕ), where G is an abstract graph and $\phi\colon G \to \mathbb{R}^2$ is a continuous map, up to the following equivalence: we say two pairs (G, ϕ_G) and (H, ϕ_H) are equivalent if there exists a homeomorphism $h\colon G \to H$ such that $\phi_G = \phi_H \circ h$. Throughout, we assume that each abstract graph G is comprised of a finite set of vertices, denoted $V(G)$, and a finite set of edges, denoted by $E(G)$. Given two points $x, y \in G$, we measure their distance by considering all paths in G and find the path whose Lebesgue measure (or, length) under ϕ is smallest:

$$L(x, y; (G, \phi)) := \inf_{p\colon [0,1] \to G} \operatorname{len}(\phi(p)),$$

where p ranges over all continuous maps such that $p(0) = x$ and $p(1) = y$. We measure the *length* of (G, ϕ) as the total Lebesgue measure of all edges in the graph: $\operatorname{len}(G) = \sum_{e \in E(G)} \operatorname{len} \phi(e)$.

Let $(A, \phi_A), (B, \phi_B) \in \mathcal{G}$. Let $f\colon A \to B$ be a function that is homeomorphic onto its image. We say that f is *length-preserving* if for each $x, y \in A$, the length of the shortest path in (A, ϕ_A) from $\phi_A(x)$ to $\phi_A(y)$ is equal to the length of the shortest path in (B, ϕ_B) from $\phi_B(f(x))$ to $\phi_B(f(y))$. More formally, f is length-preserving iff for all $x, y \in A$, $L(x, y; (A, \phi)) = L(f(x), f(y); (B, \phi_B))$.

Let $(G, \phi_G), (H, \phi_H) \in \mathcal{G}$. Let $\varepsilon > 0$. Let C be a connected subgraph of G, and let $h\colon C \to H$ be a continuous map that is homeomorphic onto its image. We measure the length of the subgraph of C that maps within ε of $h(C)$. More formally, let

$$C_h^\varepsilon := \{x \in C \mid ||\phi_G(x) - \phi_H(h(x))|| \leq \varepsilon \text{ and } \exists \delta > 0\colon h|_{\mathbb{B}(x,\delta)} \text{ is length-preserving}\}. \tag{1}$$

We can interpret this as a subgraph of (G, ϕ_G). Specifically, $(C_h^\varepsilon, \phi_G|_{C_h^\varepsilon})$ is in $\mathcal{G}$, but we note that it may not

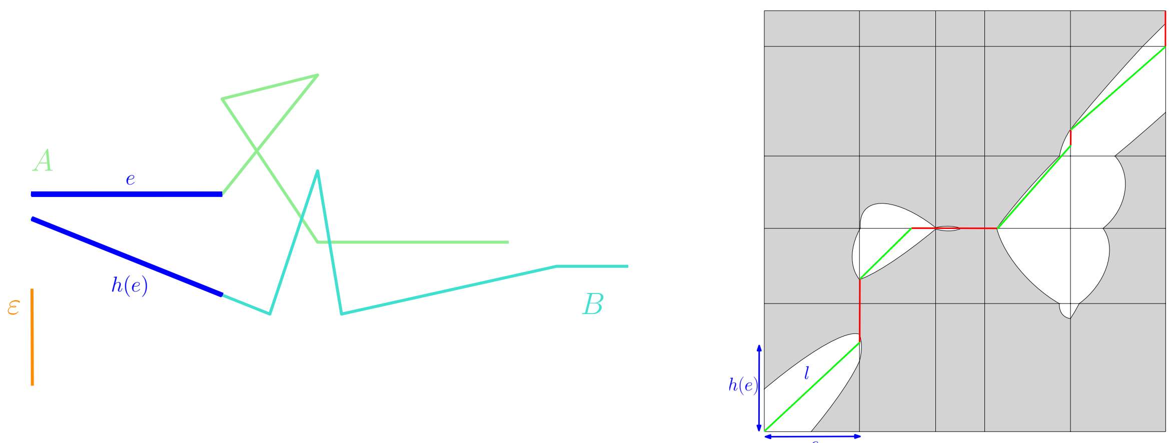

Figure 2: Free-space diagram of two curves A and B for a given ε. The green (slope-one) line segments correspond to length-preserving matchings. e is a line segment on A that was matched with $h(e)$ on B.

be a path-connected graph. To highlight the relation between length-preserving and paths in the free space diagram, consider Fig. 2. Similar to [4], in this example, we are looking for a monotone path from bottom-left to top-right but only maximizing the length of slope-one segments in the white space. A line segment l

indicates a length-preserving matching because the corresponding matched line segments, e and $h(e)$ on the two curves have the same length if and only if l is slope-one. Fig. 3 shows an example of C_h^ε. Note that C_h^ε itself is not necessarily a connected graph.

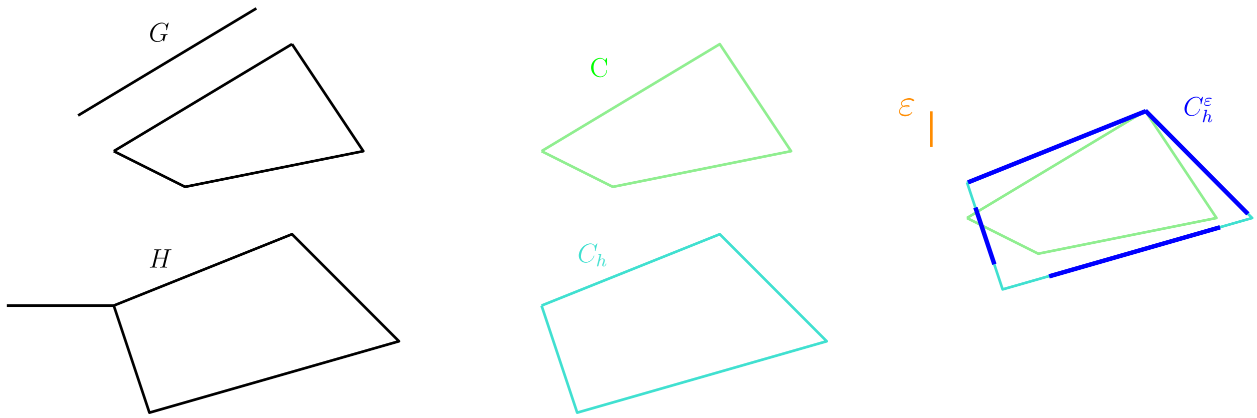

Figure 3: Computing C_h^ε on two given graphs, G and H. C is a connected subgraph of G. $C_h := h(C)$ is a subgraph of H such that C and C_h are homeomorphic. For a given ε, dark blue segments are C_h^ε.

Given this set-up, the length of the maximum *length-preserving* Fréchet correspondence is

$$L_{FC}(G, H; \varepsilon) \mapsto \sup_{C \subset G} \sup_{h : C \mapsto H} \operatorname{len}(C_h^\varepsilon).$$

While C_h^ε only requires small portions of h to be length-preserving, we show that indeed all connected components are length-preserving.

Lemma 1 (Path-Connected Components Are Length-Preserving)**.** *Let C_h^ε be defined as in Equation (1). Then, each restriction of h to a path-connected component of C_h^ε is a length-preserving map.*

Proof. Let $\widetilde{C}$ be a path-connected component of C_h^ε. Let $x, y \in \widetilde{C}$. Since $\widetilde{C}$ is a path-connected space, let $p\colon [0,1] \to \widetilde{C}$ be a path that starts at x and ends at y. Then, by the definition of C_h^ϵ, for each $t \in [0,1]$, there exists a $\delta_t > 0$ such that h restricted to $\mathbb{B}(p(x), \delta_t)$ is length-preserving. Let $\mathcal{U} := \{\mathbb{B}(p(x), \delta_t)\}_{t \in [0,1]}$. Since $\operatorname{Im}(p)$ is a compact subspace of $\mathbb{R}^2$, there exists a finite subcover $\widehat{\mathcal{U}}$ of $\mathcal{U}$. Then, there exists a decomposition of p such that each subpath lies entirely in at least one open set in $\widehat{\mathcal{U}}$. Let $\{p_i\}_{i=1}^n$ be one such decomposition. Then, we know that $\operatorname{len}(p) = \sum_i \operatorname{len} p_i$. Since each $\operatorname{Im} p_i$ is contained in some $U \in \widehat{\mathcal{U}}$ and since h restricted to U is length-preserving, we know that $\operatorname{len} p_i = \operatorname{len} h(p_i) = \operatorname{len} h(p)$. Hence, $\operatorname{len}(p) = \operatorname{len}(h(p))$ for each path from x to y, which means that $L(x, y; (G, \phi_G)) = L(h(x), h(y); (H, \phi_H))$. Thus, we have shown that h restricted to $\widetilde{C}$ is length-preserving, as was to be shown. □

3 NP-Hardness

Unfortunately, deciding whether an optimal Length-Preserving Fréchet Correspondence is above a given threshold is NP-hard.

Theorem 2 (Maximum Length-Preserving Fréchet Correspondence is NP-hard)**.** *Deciding $L_{FC}(G, H; \varepsilon) > L$ is NP-hard, even if G consists of only one edge and H is a plane graph.*

Proof. We reduce from the Hamiltonian path problem in grid graphs, which is known to be NP-hard [5], even for induced grid graphs of degree at most three [6]. The vertex set of such a grid graph is a finite subset of $\mathbb{Z}^2$, and there is an edge between two vertices u, v if and only if $\|u - v\| = 1$.

Given such a grid graph $H' = (V', E')$, we construct the graph H as follows: for every vertex we add an edge to a new degree-1 vertex at distance > 1, see Figure 4. Formally, let $V'' = V' + (3/4, 3/4)$ be the set V' translated by $(3/4, 3/4)$ and $E'' = \{(v', v'') \in V' \times V'' \mid v'' = v' + (3/4, 3/4)\}$. We note that the edges in E''

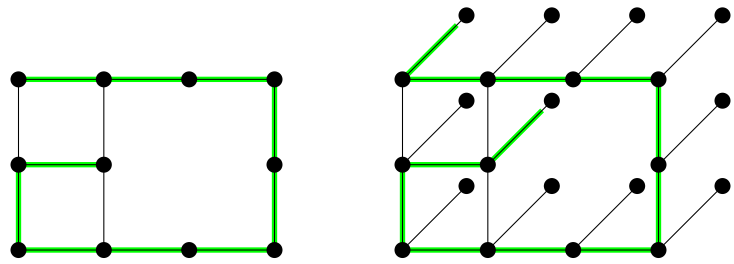

Figure 4: A grid graph H' with Hamiltonian path in green. The graph H and the image of G corresponding to the Hamiltonian path.

have length $\sqrt{2} \cdot 3/4 \approx 1.06$. We choose $H = (V' \cup V'', E' \cup E'')$. Without loss of generality, we can assume that the coordinates of the vertices of H are between 0 and $n = |V'| > 1$, since we can assume that H' is connected. Let G consist of only one edge $(0, n)$. We choose $\varepsilon = n$. We claim that if H' has a Hamiltonian path then $L_{FC}(G, H; \varepsilon) = n + 1$, and otherwise $L_{FC}(G, H; \varepsilon) < n + 1/5$. This then implies the theorem.

We have chosen ε sufficiently large such that h can map G onto any simple path in H (not necessarily ending at vertices). If H' has a Hamiltonian path, then we can map G length-preserving onto the corresponding path in H extended by parts (of length 1) of edges in E'' at the beginning and end. Thus, $L_{FC}(G, H; \varepsilon)$ is the length of the edge $(0, n)$, i.e, $n + 1$. If H' does not have a Hamiltonian path, then the longest simple path in H' that starts and ends at vertices has length at most $n - 2$. This implies that the longest simple path in H, not necessarily ending at vertices, has length at most $n - 2 + 2\sqrt{2} \cdot 3/4 \approx n + 0.12 < n + 1/5$. □

4 Discussion

For the simpler setting of comparing two curves, we expect that the framework from [4] can be applied to develop a dynamic programming algorithm for computing the length of an optimal correspondence for a given $\varepsilon > 0$. Similar to [4], we also need to compute a partial matching of maximum length inside the free space, just that we only measure the length-preserving portions. If we have two points (x, y) and $(x + \Delta x, y + \Delta y)$ on boundaries of the white space within a cell, then the the length-preserving portion that we can achieve between those two points is $\min(\Delta x, \Delta y)$. Thus, the main difference to [4] is that we have a piecewise linear function in Δx and Δy rather than just $\Delta x + \Delta y$. We therefore expect that their algorithm can be modified to our setting for two curves under the L_1 or L_∞ distances.

Assuming that we can compute $L_{FC}(G, H; \varepsilon)$ by dynamic programming for two curves, we expect that such an algorithm then generalizes to the case that H is a tree (or even the case that H has small treewidth). Another interesting case is when G also is a tree; it is related to the subtree isomorphism problem for which efficient algorithms exist; see e.g., [1].

The original aim of this research was to avoid issues with discontinuities in the graph sampling method. As an intermediate measure, we could also consider a version of L_{FC} where vertices have to be mapped to vertices, i.e., a length-preserving discrete Fréchet correspondence. For this measure NP-hardness follows even more directly, but also the algorithms become much simpler. The hardness proof assumes that there is a large number of vertices within a distance ε, and it would be interesting to develop algorithms for the case where the number of points in any ε-ball is constant.

References

[1] Amir Abboud, Arturs Backurs, Thomas Dueholm Hansen, Virginia Vassilevska Williams, and Or Zamir. Subtree isomorphism revisited. *ACM Transactions on Algorithms (TALG)*, 14(3):1–23, 2018.

[2] Jordi Aguilar, Kevin Buchin, Maike Buchin, Erfan Hosseini Sereshgi, Rodrigo I. Silveira, and Carola Wenk. Graph sampling for map comparison. *In submission*, 2021.

[3] James Biagioni and Jakob Eriksson. Inferring road maps from global positioning system traces. *Transportation Research Record: Journal of the Transportation Research Board*, 2291(1):61–71, 2012.

[4] Kevin Buchin, Maike Buchin, and Yusu Wang. Exact algorithms for partial curve matching via the Fréchet distance. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, page 645654, USA, 2009. Society for Industrial and Applied Mathematics.

[5] Alon Itai, Christos H. Papadimitriou, and Jayme Luiz Szwarcfiter. Hamilton paths in grid graphs. *SIAM Journal on Computing*, 11(4):676–686, 1982.

[6] Christos H. Papadimitriou and Umesh V. Vazirani. On two geometric problems related to the travelling salesman problem. *Journal of Algorithms*, 5(2):231–246, 1984.