

Discretized Approaches to Schematization*

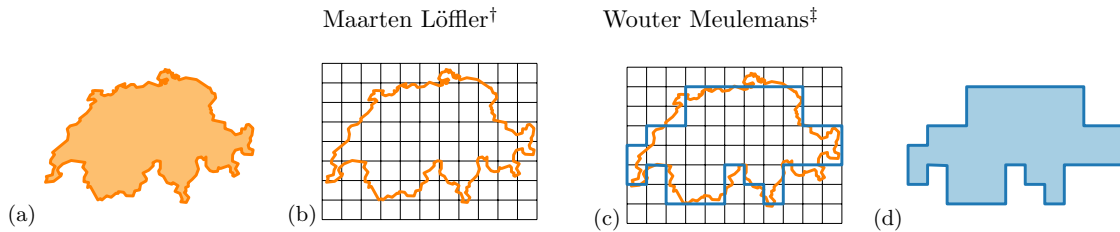


Figure 1: Discretized schematization. (a) Simple polygon P to be schematized (Switzerland). (b) Grid graph G placed on P . (c) Simple cycle S in G that best resembles P . (d) S is a rectilinear schematization of P .

Abstract

For both the Fréchet distance and the symmetric difference, we show that finding the simple polygon S restricted to a grid that best resembles a simple polygon P is NP-complete, even if: (1) we require that S and P have equal area; (2) we require turns to occur in a specified sequence for the Fréchet distance; (3) we permit S to have holes for the symmetric difference.

1 Introduction

Cartographic maps are an important tool for exploring, analyzing and communicating data in their geographic context. Effective maps show information as prominently as possible. In *schematic maps*, abstraction is taken to “extreme” levels, representing complex geographic elements with only few line segments. This highlights the primary aspects and avoids an “illusion of accuracy” [15]: the schematic appearance is a visual cue of distortion, imprecision or uncertainty. However, the low complexity must be balanced with recognizability.

Schematic maps tend to be stylized by constraining the permitted geometry. Orientations of line segments are often restricted to a small set \mathcal{C} . The typical example is a schematic transit map, in which all segments are horizontal, vertical or a 45-degree diagonal. A central problem in schematization is the following: given a simple polygon P , compute a simple \mathcal{C} -oriented polygon S with low complexity and high resemblance to P .

Here we investigate a discretized approach to schematization, characterized by placing a grid graph G over P that models our geometric style and requiring the boundary of S to coincide with a simple cycle in G (Fig. 1). Though it restricts the solution space, this approach readily offers some benefits.

- It can easily model a variety of constraints, even combining different geometry types.
- It promotes the use of collinear edges and provides a uniformity of edge lengths.
- Simplicity enforces a minimal width for narrow strips in P , leading to automated exaggeration—a main cartographic operator [16] for avoiding an undesirable visual collapse (see Fig. 2).
- It makes areas easy to assess [5] or subdivide [14].

Contributions. Focusing on grid graphs (a grid of unit squares), we consider two similarity metrics: the Fréchet distance and the symmetric difference. In Section 3 we prove that the problem is NP-complete under the Fréchet distance, even if we require area preservation and restrict valid solutions to those with a specific sequence of left and right turns. In Section 4 we prove that the problem is NP-complete also under the symmetric difference, even if we require area preservation and we permit the solution to be a polygon with holes. Though the problems are similar in setup, the very different natures of the metrics require different reductions.

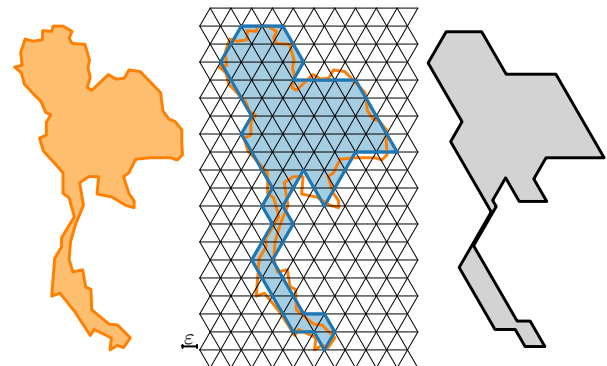


Figure 2: Discretization with simplicity (mid) exaggerates the narrow strip in Thailand (left). Result with a visual collapse, computed using [4] (right).

*An early version of this work appeared on arXiv [13].

[†]Utrecht University, the Netherlands, m.loffler@uu.nl

[‡]Eindhoven University of Technology, the Netherlands, w.meulemans@tue.nl

Related work. We highlight the most relevant related work; for a more complete treatment refer to [13]. Recently, schematizing geographic regions has gained increasing attention, e.g. [4, 17]. Our discretized approach is similar in nature to the octilinear schematization technique of Cicerone and Cermignani [6], though simplicity is of no concern in their work. A grid graph always admits a solution with Hausdorff distance at most $3\sqrt{2}/2$ and Fréchet distance at most $(\beta + \sqrt{2})/2$, for β -narrow polygons [2]. Minimizing the Hausdorff distance is NP-complete even on a grid graph [2].

Our problem using the Fréchet distance closely resembles “map matching”, with applications in GIS [1, 11] Wylie and Zhu [18] prove independently that our problem is NP-hard under the *discrete* Fréchet distance, however, without requiring a simple input polygon nor a grid graph. A stronger result—with a simple input curve and a grid graph—follows directly from our proofs.

On the dual graph, the problem under the symmetric difference is a specialization of the known NP-hard maximum-weight-connected-subgraph problem [8, 12]. Our results readily imply that this dual problem remains NP-hard even in a constrained geometric setting.

2 Preliminaries

Polygons. A polygon P is defined by a cyclic sequence of vertices in \mathbb{R}^2 . We use $|P|$ to refer to the area of polygon P and ∂P for its boundary. A polygon is *simple* if no two edges intersect, except at common vertices.

Grid graphs. A grid graph $G = (V, E)$ is a plane graph with all vertices positioned at integer coordinates within a rectangular region, with edges being all unit-length segments connecting pairs of vertices at distance 1.

Cycles. A *cycle* in a graph is a (cyclic) sequence of adjacent vertices; a cycle is *simple* if the sequence does not contain a vertex more than once. A simple cycle in a grid graph corresponds to a simple rectilinear polygon.

Faces. G has two types of faces: *cells* (unit squares), and an *outer face*. A set of faces in G is said to be *connected* if the corresponding induced subgraph of the dual graph G^* is connected; it is *simply connected* if the remaining faces are also connected. A simply connected face set corresponds to a simple rectilinear polygon.

Fréchet distance. Let $B_P: S^1 \rightarrow \partial P$ continuously map the unit circle onto the boundary of P . Let Ψ denote the set of all orientation-preserving homeomorphisms on S^1 . The *Fréchet distance* between two polygons, $d_F(P, Q)$, is defined as $\inf_{\psi \in \Psi} \max_{t \in S^1} \|B_P(t) - B_Q(\psi(t))\|$, where $\|\cdot\|$ denotes the Euclidean distance.

Symmetric difference. The symmetric difference between two polygons P and Q is defined as the area covered by precisely one of the polygons: $d_{SD}(P, Q) = |(P \cup Q) \setminus (P \cap Q)| = |P \cup Q| - |P \cap Q| = |P| + |Q| - 2 \cdot |P \cap Q|$.

3 Using the Fréchet distance

Theorem 1 *Let G be a grid graph, let P be a simple polygon and let $\varepsilon > 0$. It is NP-complete to decide whether G contains a simple cycle C with $d_F(C, P) \leq \varepsilon$.*

We focus here on sketching a proof that it is indeed NP-hard. We assume $\varepsilon = 3.5$ throughout this proof.

We reduce from *planar monotone 3-SAT* [7]: the problem to decide whether a 3CNF formula F with clauses of either fully positive or fully negative literals is satisfiable, where F is embedded such that all variables lie on a single horizontal line, and clauses are positioned above (positive) or below (negative) this line and are connected to the variables using 1-bend orthogonal leaders (Fig. 3). We construct a simple polygon P and a grid graph G such that G contains a simple cycle C with $d_F(C, P) \leq 3.5$ if and only if F is satisfiable. The construction, and therefore G , has polynomial complexity. Below, we sketch the necessary gadgets which lead to the construction shown in Fig. 4.

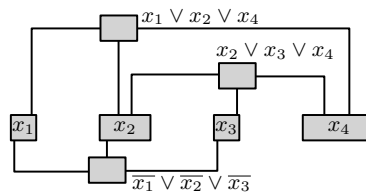


Figure 3: Instance of the constrained 3CNF formula F .

3.1 Gadgets

We first review the shared general idea of all gadgets and their visual encoding in the corresponding figures. Graph G is shown by a grid of thin light-gray lines. Each gadget contains a part of P called the *local curve* (red), ending in two *gates* (red dots). To reason about gadgets, we also use a corresponding cycle in G which we call the *path boundary* (black) and a *curve area* containing the local curve (gray-filled). The gadgets interact via vertices and edges on shared path boundaries. There is no interaction based on the local curve: it is used only to force choices in using edges of G .

Pressure. If a cycle exists in the complete graph, a *local path* within or on the path boundary must have Fréchet distance at most 3.5 to the local curve. The local path “claims” its vertices: these can no longer be used by another gadget. This results in *pressure* on the other gadget to use a different path, if shared vertices (on the path boundary) are used. To support reasoning about interaction, a gadget has *pressure ports* (green): a sequence of edges on the path boundary that may be shared with another gadget. A port may *receive* pressure, indicating that the shared vertices may not be used in the gadget for its local path. Similarly, it may

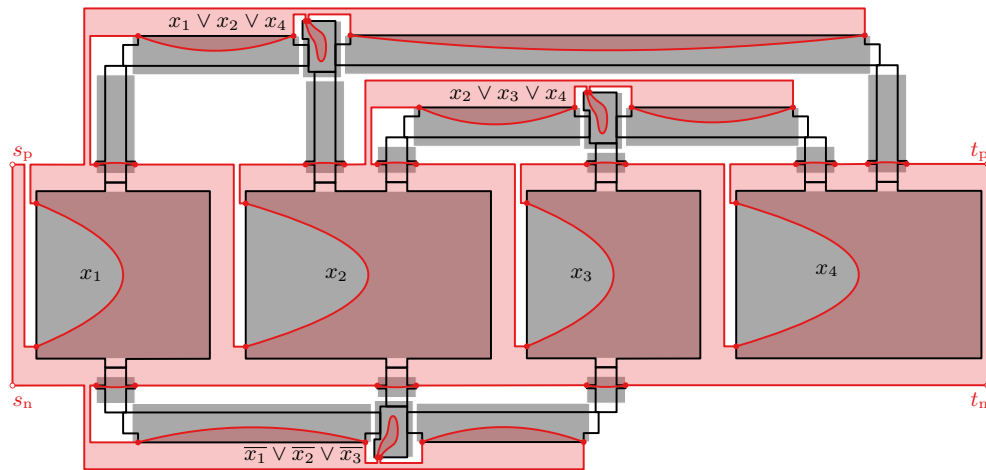


Figure 4: Construction sketch with gadgets for the formula in Fig. 3. Gadgets specify an area in which part of P is located (gray rectangle) and in which part of C must lie (black polygon). They interact via shared boundaries of the black polygons. The red lines connect the various gadgets to obtain a simple polygon, its interior is slightly shaded; the curves within the gadgets are abstracted.

give pressure, indicating that the shared vertices may not be used by an adjacent gadget.

Propagation gadget. A propagation gadget has exactly two ports. The gadget does not admit a path if both ports receive pressure. If one port receives pressure, the other must give pressure: in other words, it propagates pressure. The gadgets can be constructed with any length that is at least 12. Fig. 5 shows one of length (height) 12. The length of the gadget determines where its gates appear: on opposite sides for odd-length and on the same side for even-length gadgets. The complexity of the local curve is linear in its length.

The propagation works due to a local curve that zigzags back and forth, with a distance over $2\epsilon = 7$ between the endpoints of the zigzags. Hence, the middle of the gadget must be crossed for each zigzag. By placing the right number of zigzags, and having the first and last at exactly distance 3.5 from the ports, we achieve that one of the ports must be overlapped by a local path. The positioning of the first and last further ensure that *all* edges of one port must be used. This is illustrated in Fig. 5: since the dotted variants do not work, the solid ones must be used. Taking an extra grid cell upwards to reach the endpoint leads to pressure on the other side, in which case the entire other port must be covered.

Clause gadget. A clause gadget is illustrated in Fig. 6. It has fixed dimensions. The gadget admits a local path only if one of its ports does *not* receive pressure. Any local path causes pressure on at least one port; for each port there is a path that causes pressure only on that port. The lack of external pressure on a port indicates that the value of the corresponding variable satisfies the clause. There is no local path that avoids all three ports:

if all ports receive pressure, none of the variables satisfy the clause and the gadget does not admit a local path.

The construction roughly consists of three zigzags, with one extra spike nested inside the middle one. This extra spike is the crucial element: it can be positioned inside any of the other three zigzags (since ϵ is 3.5), mimicking that (at least) one of three variables satisfies the clause. The middle zigzag and the spike reach down to approximately the same y -coordinate. In particular, to cover the endpoints of the zigzag or spike, the local

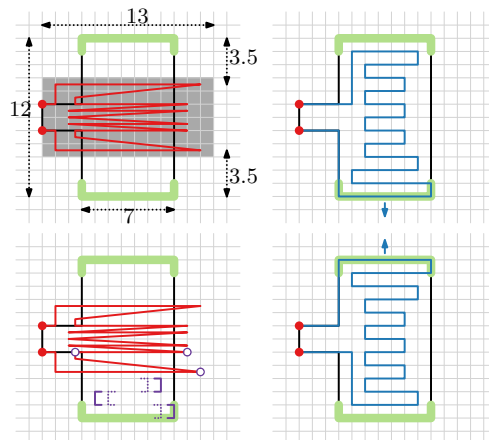


Figure 5: A propagation gadget. Specification and local curve (top left). Local paths with pressure on exactly one port (right). The first three endpoints (open dots) can be reached with the solid subpaths (purple), but not with the dotted variants (bottom left).

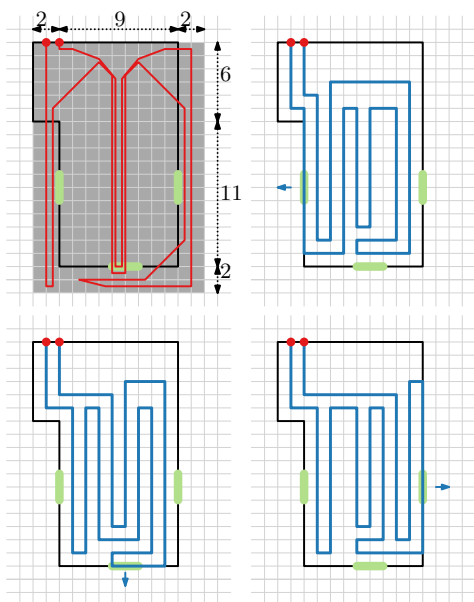


Figure 6: The specification and local curve (top left) of a clause gadget. The other three figures illustrate local paths; each gives pressure on at least one port.

path must reach down to distance 3 above the port. Without the spike inside, the zigzag can be reached by going to exactly this distance. However, if the spike is inside it, the middle zigzag is forced down one more space. The extra cover from the right zigzag propagates this towards the port. This extra cover is necessary due to the convexity of the Euclidean distance: without it, the outer zigzags must reach at least as far down as the middle one, when the spike is inside it; this would cover any potential port intended for the middle zigzag.

The curve area extends slightly outside of the path boundary. Hence, local paths exist with Fréchet distance at most 3.5 that lie outside the path boundary. However, this would claim more beyond the indicated ports, only further restricting any other nearby gadgets.

Variable gadget. Variable gadgets are obtained by making a cycle of propagation gadgets, such that they must all find a local path in the same direction. This direction readily represents the truth value of the variable, which can be transferred to other propagation gadgets.

3.2 Construction with gadgets

We now construct polygon P based on formula F (Fig. 4). First, we place all variable gadgets next to one another, in the order determined by F , with a distance of 11 between consecutive variables. This placement ensures that the ports of variables start on an even and end on an odd x -coordinate.

Using the y -coordinates in the embedding of F , we

sort the positive clauses to define a *positive order* $\langle c_1, \dots, c_k \rangle$. We place the gadget for clause c_j such that the bottom side of its path boundary is at a distance $13 + 24(j - 1)$ above the variables. Analogously, we use a *negative order* to place the negative clauses below the variables. Horizontally, the clause gadgets are placed such that the right side of the path boundary lines up with the right side of the appropriate port on the variable gadget of the middle literal. Finally, we place propagation gadgets for each link in F to connect the clause and variable gadgets.

Any overlap in the curve areas would imply that the provided embedding for F —which structures the layout—is not planar. Thus, all gadgets have disjoint curve areas: local curves do not intersect.

Connecting gadgets. We have composed the various gadgets in polynomial time. However, we do not yet have a simple polygon. We must “stitch” the local curves together (in any order) to create polygon P . To this end we first create two subcurves: P_p for the variables, positive clause gadgets and their propagation gadgets; and P_n for the negative clause gadgets and their propagation gadgets. Fig. 4 visually illustrates the construction of these three subcurves, each with endpoints s_* and t_* , and how to connect them.

Proving the theorem. We now have a simple polygon P ; with G implicitly defined as a large enough grid graph. We must argue that the complexity is polynomial and that F is satisfiable if and only if a simple cycle C exists in G with $d_F(C, P) \leq 3.5$. We sketch the main argument of the proof.

A satisfying assignment is derived from C by inspecting the local paths of the variable gadgets: is the pressure clockwise or counterclockwise? Similarly, a satisfying assignment leads to a cycle in G , by concatenating the appropriate local paths given with the gadgets.

Let n denote the number of variables, and m the number of clauses in formula F . Variable gadgets have width linear in their degree and are placed with $O(1)$ distance between them: the entire width is bounded by $O(n+m)$. Variable and clause gadgets have constant height and are placed with $O(1)$ distance between them: the entire height is bounded by $O(m)$. Hence, the polygon’s coordinates remain polynomial.

4 Using the symmetric difference

Theorem 2 *Let G be a grid graph, let P be a simple polygon and let $D > 0$. It is NP-complete to decide whether G contains a connected face set S with $d_{SD}(S, P) \leq D$.*

Here we focus on sketching a proof to show that the problem is NP-hard.

We reduce from the *rectilinear Steiner tree problem* [9]: given a set X of n points in \mathbb{R}^2 , is there a tree T

of total edge length at most L that connects all points in X , using only horizontal and vertical line segments? Vertices of T are not restricted to X . An optimal result must be contained in graph $H(X)$ corresponding to the arrangement of horizontal and vertical lines through each point in X [10]; $H(X)$ is illustrated in Fig. 7(a). We call a vertex of $H(X)$ a *node* if it corresponds to a point in X and a *junction* otherwise. As the problem is scale invariant, we assume $L = 1$ and thus all edges in $H(X)$ must be shorter than 1.

We transform point set X into a grid graph G , a polygon P and a value $D > 0$. We construct G such that each cell corresponds to a vertex (*node-cell* or *junction-cell*), an edge (*edge-cell*), or a bounded face (*face-cell*) of $H(X)$; see Fig. 7(b). Polygon P is constructed to partially overlap all cells, except for the face-cells. To structure P we use a *skeleton* ς , a tree spanning the non-face-cells in the dual of G .

Weights. The symmetric difference between P and a face set $S = \{c_1, c_2, \dots\}$ may be computed as $|P| + \sum_{c \in S} (|c| - 2 \cdot |P \cap c|)$, since the faces in S are interior-disjoint. As $|c| = 1$, we define the *weight* of a cell c in G as $w(c) = 1 - 2 \cdot |P \cap c|$. Hence, the symmetric difference is $|P| + \sum_{c \in S} w(c)$. We set the desired weight $w(c)$ for cell c to: $-\frac{3}{4}$ if c is a node-cell; 0 if c is a junction-cell; $\|e\|/2$ if c is an edge-cell, where $\|e\|$ is the length of the corresponding edge e in $H(X)$; and 1 if c is a face-cell.

Given $w(c)$ for cell c , the area of overlap $A(c)$ is $|P \cap c| = \frac{1-w(c)}{2}$. Every cell is covered by $A(c)$ area of P ; $|P|$ equals $\sum_{c \in G} A(c)$. We call $P \cap c$ the *local polygon* of c . We set $D = |P| - \frac{3}{4}n + \frac{1}{2} = (\sum_{c \in G} A(c)) - \frac{3}{4}n + \frac{1}{2}$.

Designing cells. We design every cell such that the desired weight is achieved. For a face-cell, $w(c) = 1$ and $A(c) = 0$: P does not overlap this cell. For all other cells the local polygon covers a fraction of its interior, as determined by $A(c)$. Skeleton ς dictates how to connect the local polygons; we ensure that at least the middle 25% of the shared edge (the *connector*) is covered. A local polygon never touches the corners of its cell.

Node- and junction-cells may have up to four neigh-

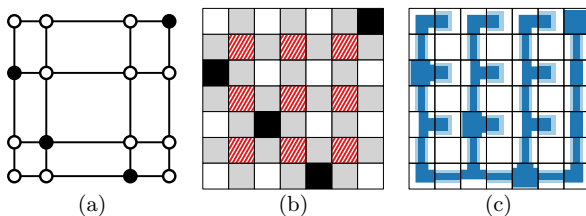


Figure 7: Sketch of the reduction. (a) Graph $H(X)$ where X is given by dark points. (b) Grid graph G : node-cells are black, junction-cells white, edge-cells gray; face-cells are hatched. (c) Constructed polygon P with respect to G .

bors in ς . A node-cell has weight $-\frac{3}{4}$; $A(c) = \frac{7}{8}$. A junction-cell has weight 0; $A(c) = \frac{1}{2}$. The local polygon can easily touch the connectors while covering exactly the prescribed area, see Fig. 8(a-b).

An edge-cell has weight $\|e\|/2$ and thus should cope with weights between 0 and 0.5; $A(c)$ lies between $\frac{1}{4}$ and $\frac{1}{2}$. Any edge-cell has degree 1 or 2 in ς ; if it has degree 2, the neighboring cells are on opposite sides. The local polygon for $A(c) = \frac{1}{4}$ is a rectangular shape that touches exactly the necessary connectors; we widen this shape to cover the precise area needed if $A(c) > \frac{1}{4}$. This is illustrated in Fig. 8(c).

Proving the theorem. The reduction is polynomial, as P has $O(1)$ complexity in each of the $O(n^2)$ cells. What remains is to prove equivalence of the answers.

Suppose we have a rectilinear Steiner tree T of length at most 1 in $H(X)$. We construct a face set S as the union of all cells corresponding to vertices and edges in T . By definition of T , this must contain all node-cells and cannot contain face-cells. As junction-cells have no weight, the total weight of S is $-\frac{3}{4}n + \sum_{e \in T} w(c_e) = -\frac{3}{4}n + \frac{1}{2} \sum_{e \in T} \|e\|$ where c_e is the cell of G corresponding to edge e . By assumption $\sum_{e \in T} \|e\| \leq 1$: the total weight is at most $-\frac{3}{4}n + \frac{1}{2}$. Thus, the symmetric difference for S is at most $|P| - \frac{3}{4}n + \frac{1}{2} = D$.

Suppose we have a connected face set S in G such that $d_{SD}(S, P) \leq D$. The total weight is thus $D - |P| = -\frac{3}{4}n + \frac{1}{2}$. Since face-cells have weight 1 and only node-cells have negative weight, being $-\frac{3}{4}$, this can be achieved only if S contains all node-cells and no face-cells. In particular, the sum of the weights over all edge-cells is at most $\frac{1}{2}$. Thus, the subgraph of $H(X)$ described by the selected cells is connected, contain all nodes of X , and have total length at most 1. If this subgraph is not a tree, we can make it a tree, by leaving out edges (further reducing the total length), until the subgraph is a tree.

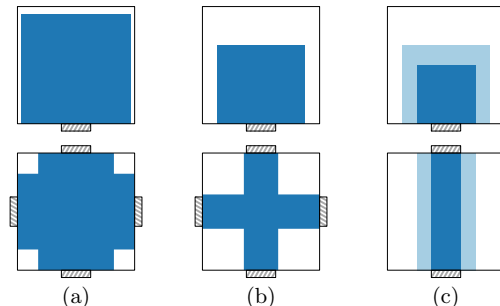


Figure 8: Local polygons, with connectors as hatched rectangles. (a) Node-cell, covered for 87.5%. (b) Junction-cell, covered for 50%. (c) Edge-cell, covered for 25% (dark) up to 50% (dark and light).

5 Conclusions

We studied discretized approaches to the construction of schematic maps, by restricting solutions to grid graph. This has several advantages, such as promoting alignment and uniformity of edge lengths and avoiding the risk of a visual collapse. We considered two similarity metrics: the Fréchet distance and the symmetric difference. Unfortunately, both turn out to be NP-complete.

Implications. Our reductions imply several further results; refer to [13] for further details. Relevant for e.g. area-preserving schematization [4] and cartograms [5], computing the best *area-equivalent* shape is also NP-complete, under both similarity metrics. The Fréchet-distance variant admits no PTAS and its reduction extends to the *discrete* Fréchet distance, as well as to polygons with a given *bend profile*: the sequence of left and right bends in counterclockwise order along its boundary. The symmetric difference reduction also works for a *simply* connected face set. Finally, the problem remains NP-complete for graphs representing hexagonal and triangular tilings.

Open problems. In contrast to *partial* grid graphs [13], strict monotonicity in the Fréchet distance is crucial for this reduction with full grid graphs; the problem under the *weak* Fréchet distance on full grid graphs remains open. Moreover, it remains open whether the problem (general or grid-based) is fixed-parameter tractable—in e.g. the number of bends in the output polygon—for either similarity metric.

Though our problem under the Fréchet distance is now proven NP-complete, the construction requires that ε is 3.5. For $\varepsilon < 0.5$, the problem becomes trivial to solve: there is only one feasible sequence of vertices in G . For partial grid graphs hardness can be proven for $\varepsilon = 1$ [13]; but what is the complexity with full grid graphs for ε in between 0.5 and 3.5?

Do realistic input assumptions help to obtain efficient algorithms? A first result is known [2], finding a cycle that has Fréchet distance bounded in the “narrowness” of the input polygon. Can we obtain a complementary result, where the algorithm’s running time rather than its Fréchet distance depends on the realism parameter?

Results obtained via the Fréchet distance may locally deviate more than necessary. Can we extend *locally correct Fréchet matchings* [3] to our setting?

Acknowledgments. The authors would like to particularly thank: Kevin Buchin, Bart Jansen, Arthur van Goethem, Marc van Kreveld, Aidan Slingsby, and Bettina Speckmann for inspiring discussions on the topic of this paper. W. Meulemans was partially supported by Marie Skłodowska-Curie Action MSCA-H2020-IF-2014 656741, the Netherlands Organisation for Scientific Research (NWO) project 639.023.208 and the Netherlands eScience Center (NLeSC) project 027.015.G02.

References

- [1] H. Alt, A. Efrat, G. Rote, and C. Wenk. Matching planar maps. *J. Algo.*, 49:262–283, 2003.
- [2] Q. W. Bouts, I. Kostitsyna, M. van Kreveld, W. Meulemans, W. Sonke, and K. Verbeek. Mapping polygons to the grid with small Hausdorff and Fréchet distance. In *Proc. 24th ESA, LIPIcs 57*, Art. 22, 2016.
- [3] K. Buchin, M. Buchin, W. Meulemans, and B. Speckmann. Locally correct Fréchet matchings. In *Proc. 20th ESA, LNCS 7501*, pages 229–240, 2012.
- [4] K. Buchin, W. Meulemans, A. van Renssen, and B. Speckmann. Area-preserving simplification and schematization of polygonal subdivisions. *ACM TSAS*, 2(1):Art. 2, 2016.
- [5] R. G. Cano, K. Buchin, T. Castermans, A. Pieterse, W. Sonke, and B. Speckmann. Mosaic drawings and cartograms. *CGF*, 34(3):361–370, 2015.
- [6] S. Cicerone and M. Cermignani. Fast and simple approach for polygon schematization. In *Proc. 12th ICCSA, LNCS 7333*, pages 267–279, 2012.
- [7] M. de Berg and A. Khosravi. Optimal binary space partitions in the plane. In *Proc. 16th COCOON, LNCS 6196*, pages 216–225, 2010.
- [8] M. El-Kebir and G. W. Klau. Solving the maximum-weight connected subgraph problem to optimality. *CoRR*, abs/1409.5308, 2014.
- [9] M. R. Garey and D. S. Johnson. The rectilinear Steiner tree problem is NP-complete. *SIAM J. Appl. Math.*, 32(4):826–834, 1977.
- [10] M. Hanan. Steiner’s problem with rectilinear distance. *SIAM J. Appl. Math.*, 14(2):255–265, 1966.
- [11] J.-H. Haurert and B. Budig. An algorithm for map matching given incomplete road data. In *Proc. 20th ACM SIGSPATIAL GIS*, pages 510–513, 2012.
- [12] D. Johnson. The np-completeness column: an ongoing guide. *J. Alg.*, 6(1):145–159, 1985.
- [13] W. Meulemans. Discretized approaches to schematization. *CoRR*, abs/1606.06488, 2016.
- [14] W. Meulemans, J. Dykes, A. Slingsby, C. Turkay, and J. Wood. Small multiples with gaps. *IEEE TVCG*, 23(1):381–390, 2017.
- [15] M. Monmonier. *How to Lie with Maps*. University of Chicago Press, 1996.
- [16] N. Regnaud and R. B. McMaster. A synoptic view of generalisation operators. In *Generalisation of Geographic Information: Cartographic Modelling & Applications*, pages 37–66, 2007.
- [17] A. van Goethem, W. Meulemans, B. Speckmann, and J. Wood. Exploring curved schematization of territorial outlines. *IEEE TVCG*, 21(8):889–902, 2015.
- [18] T. Wylie and B. Zhu. Intermittent map matching with the discrete Fréchet distance. *CoRR*, abs/1409.2456, 2014.