Uncertain Curve Simplification

Citation for published version (APA):

Document status and date:
Published: 07/04/2021

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 13. Jun. 2022
Uncertain Curve Simplification

Kevin Buchin\textsuperscript{1}, Maarten Löffler\textsuperscript{*2}, Aleksandr Popov\textsuperscript{†1}, and Marcel Roeloffzen\textsuperscript{‡1}

\textsuperscript{1} Department of Mathematics and Computer Science, TU Eindhoven, Netherlands
\{k.a.buchin, a.popov, m.j.m.roeloffzen\}@tue.nl

\textsuperscript{2} Department of Information and Computing Sciences, Utrecht University, Netherlands
m.loffler@uu.nl

Abstract

We study polygonal curve simplification under uncertainty, where instead of a sequence of exact points, each uncertain point is represented by a region, which contains the (unknown) true location of the vertex. The regions we consider are discrete sets of points, line segments, and convex polygons. We are interested in finding the shortest subsequence of uncertain points such that no matter what the true location of each point is, the resulting polygonal curve is a valid simplification of the original curve under the Hausdorff distance. We present polynomial-time algorithms for this problem.

Related Version An extended version with complete proofs, in which we also discuss the uncertainty modelled with disks and show the set of results for the Fréchet distance, is available on arXiv: Full Version: https://arxiv.org/abs/2103.09223 [6]

1 Introduction

Curve simplification is the problem of replacing a polygonal curve by another one which has a similar shape, but fewer vertices. Classical solutions [1], such as those by Ramer and by Douglas and Peucker [7, 15] and by Imai and Iri [8], constrain the new curve to use a subset of the vertices of the old curve and use the Hausdorff distance to measure similarity. Curve simplification is still an active field of study [2, 3, 5, 16].

Typically, it is assumed that the locations of the input curve vertices are known precisely, which is often not the case in real-life data, for instance, when locations are measured using GPS technology. There have been some advances in the study of uncertainty in computational geometry [9, 10, 11, 12, 13], and more recently of uncertain curves [4, 14]. However, to our knowledge, there is no such previous work tackling curve simplification under uncertainty.

An uncertain curve consists of a sequence of uncertain points, each modelled as a convex polygon, a line segment, or a discrete set of points that contains the true location of the point. When faced with the task of simplifying an uncertain curve, one must consider what is the expected output. We investigate the following practical point of view: we wish to select a subsequence of the (uncertain) input points which is guaranteed to be a valid simplification; see Figure 1. Our results are based on adapting the approach by Imai and Iri [8]: we construct a shortcut graph which contains all shortcuts that are guaranteed to be valid. In this abstract, we show that we can check a single shortcut in Section 4, using the procedure of Section 3 for several pairs of points.

\textsuperscript{*} Partially supported by the Dutch Research Council (NWO) under project no. 614.001.504.

\textsuperscript{†} Supported by the Dutch Research Council (NWO) under project no. 612.001.801.

\textsuperscript{‡} Supported by the Dutch Research Council (NWO) under project no. 628.011.005.

This is an extended abstract of a presentation given at EuroCG'21. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.
Uncertain Curve Simplification

![Figure 1](a) An uncertain curve and a potential realisation. (b) A valid simplification: for every realisation, the subsequence is within Hausdorff distance ε from the full sequence. (c) An invalid simplification: there is a realisation for which the subsequence is not within Hausdorff distance ε.

**Theorem 1.** We can find the shortest vertex-constrained simplification of an uncertain curve modelled with convex polygons or discrete sets of points, such that for any realisation the simplification is valid under the Hausdorff distance, in time $O(n^3k^3)$, where $k$ is the size of the point sets or the complexity of the polygons and $n$ is the length of the input curve; and in time $O(n^3)$ for line segments.

### 2 Preliminaries

Denote $[n] \defeq \{1, 2, \ldots, n\}$ for any $n \in \mathbb{N}^>0$. Given two points $p, q \in \mathbb{R}^2$, denote their Euclidean distance with $\|p - q\|$. Denote a sequence of points in $\mathbb{R}^2$ with $\pi = \langle p_1, \ldots, p_n \rangle$. For only two points $p, q \in \mathbb{R}^2$, we also use $pq$ instead of $\langle p, q \rangle$. Denote a subsequence of a sequence $\pi$ from index $i$ to $j$ with $\pi[i : j] = \langle p_i, p_{i+1}, \ldots, p_j \rangle$. This notation can also be applied if we interpret $\pi$ as a polygonal curve on $n$ vertices (of length $n$). It is defined by linearly interpolating between the successive points in the sequence and can be seen as a continuous function, for $i \in [n - 1]$ and $\alpha \in [0, 1]$: $\pi(i + \alpha = (1 - \alpha)p_i + \alpha p_{i+1}$.

We introduce the notation for order along a curve. Let $p := \pi(a)$ and $q := \pi(b)$ for some $a, b \in [1, n]$. Then $p \prec q$ iff $a < b$, $p \preceq q$ iff $a \leq b$, and $p \equiv q$ iff $a = b$. Finally, given points $p, q, r \in \mathbb{R}^2$, define the distance from $p$ to the segment $qr$ as $d(p, qr) \defeq \min_{t \in [0, 1]} \|p - tq + (1-t)r\|$. An uncertainty region $U \subset \mathbb{R}^2$ describes a possible location of a point: it has to be inside the region, but we do not know where. Call a sequence of uncertainty regions an uncertain curve: $U = \langle U_1, \ldots, U_n \rangle$. Picking a point from each uncertainty region of $U$, we get a polygonal curve $\pi$ called a realisation of $U$, denoted $\pi \in U$. That is, if for some $n \in \mathbb{N}^>0$ we have $\pi = \langle p_1, \ldots, p_n \rangle$ and $U = \langle U_1, \ldots, U_n \rangle$, then $\pi \in U$ if and only if $p_i \in U_i$ for all $i \in [n]$.

Suppose we are given a polygonal curve $\pi = \langle p_1, \ldots, p_n \rangle$, a threshold $\varepsilon \geq 0$, and a curve built on the subsequence of vertices of $\pi$ for some set $I = \{i_1, \ldots, i_\ell\} \subseteq [n]$, i.e. $\sigma = \langle p_{i_1}, \ldots, p_{i_\ell} \rangle$ with $i_j < i_{j+1}$ for all $j \in [\ell - 1]$ and $\ell \leq n$. We call $\sigma$ an $\varepsilon$-simplification of $\pi$ if for each segment $\langle p_{i_j}, p_{i_{j+1}} \rangle$, we have $\delta((p_{i_j}, p_{i_{j+1}}), \pi[i_j : i_{j+1}]) \leq \varepsilon$, where $\delta$ denotes some distance measure, e.g. the Hausdorff distance $d_H$.

---

1. We use $\equiv$ and $\defeq$ for equivalent quantities in definitions or to point out equality by earlier definition, and $\equiv$ in other contexts. We also use $\pi$, but its usage is always explained.
2. Note that we can have $p = q$ for $a \neq b$ if the curve intersects itself.
Define a **polygonal closed convex set (PCCS)** as a closed convex set with bounded area that can be described as the intersection of a finite number of closed half-spaces. Note that this includes both convex polygons and line segments. Given a PCCS \( U \), let \( V(U) \) denote the set of vertices of \( U \). An indecisive point is a discrete set of points: \( U = \{ p_1, \ldots, p_k \} \), with \( k \in \mathbb{N}^{>0} \) and \( p_i \in \mathbb{R}^2 \) for all \( i \in [k] \). We solve the following problem to check each shortcut.

**Problem 2.** Given an uncertain curve \( U = \{ U_1, \ldots, U_n \} \) on \( n \in \mathbb{N}^{>3} \) with \( U_i \subset \mathbb{R}^2 \) for all \( i \in [n] \) modelled as PCCSs or indecisive points, find the maximum possible distance between the curve and its one-segment simplification for any realisation, i.e. find \( \max_{\pi \in \mathcal{P}} d_H(\pi, \langle \pi(1), \pi(n) \rangle) \).

### 3 Shortcut Testing: Intermediate Points

In this section, we start with the start and end points fixed, so we are given some \( p_1 \in U_1 \) and \( p_n \in U_n \) and consider realisations \( \pi \in \mathcal{U} \) with \( \pi(1) \equiv p_1 \) and \( \pi(n) \equiv p_n \). We make use of the following intuitive statements.

**Lemma 3.** Given four points \( a, b, c, d \in \mathbb{R}^2 \) forming segments \( ab \) and \( cd \), the largest distance from one segment to the other is achieved at an endpoint:

\[
\max_{p \in \{a,b\}} d(p, cd) = \max \{d(a, cd), d(b, cd)\}.
\]

**Lemma 4.** Given \( n \in \mathbb{N}^{>0} \), for any precise curve \( \pi = \langle p_1, \ldots, p_n \rangle \) with \( p_i \in \mathbb{R}^2 \) for all \( i \in [n] \), we have \( d_H(\pi, p_1p_n) = \max_{i \in [n]} d(p_i, p_1p_n) \).

We can show the following lemma, allowing us to test the shortcut with fixed realisations of the endpoints of length \( n \) in time \( \mathcal{O}(nk) \) for PCCSs with at most \( k \) vertices. We can obtain the same equality for indecisive points, replacing \( V(U_i) \) with \( U_i \).

**Lemma 5.** Given \( n \in \mathbb{N}^{>3} \), for any uncertain curve modelled with PCCSs \( U = \{ U_1, \ldots, U_n \} \) with \( U_i \subset \mathbb{R}^2 \) for all \( i \in [n] \) and \( V(U_1) = \{ p_{1i}, \ldots, p_{ki} \} \) for all \( i \in [n], k \in \mathbb{N}^{>0} \), and given some \( p_1 \in U_1 \) and \( p_n \in U_n \), we have

\[
\max_{\pi \in \mathcal{U}} \max_{\pi(1) \equiv p_1, \pi(n) \equiv p_n} d_H(\pi, p_1p_n) = \max_{i \in \{2, \ldots, n-1\}} \max_{v \in V(U_i)} d(v, p_1p_n).
\]

**Proof.** Assume the setting of the lemma. Derive \( \max_{\pi \in \mathcal{U}} \max_{\pi(1) \equiv p_1, \pi(n) \equiv p_n} d_H(\pi, p_1p_n) = \max_{\pi \in \mathcal{U}, \pi(1) \equiv p_1, \pi(n) \equiv p_n} \max_{i \in [n]} d(\pi(i), p_1p_n) = \max_{i \in \{2, \ldots, n-1\}} \max_{p \in U_i} d(p, p_1p_n) \), where the steps hold by Lemma 4 and the definition of \( \langle \rangle \). It is trivial to show that for any PCCS \( U \) and a line segment \( ab \) it holds that \( \max_{p \in \mathcal{U}} d(p, ab) = \max_{v \in V(U)} d(v, ab) \).

### 4 Shortcut Testing: All Points

In the previous section, we have covered testing a shortcut, given that the first and last points are fixed. Here we generalise the problem: we check if \( \max_{\pi \in \mathcal{U}} d_H(\pi, \langle \pi(1), \pi(n) \rangle) \leq \varepsilon \).

First of all, for indecisive curves, we can just use the approach of the previous section for each pair of realisations from \( U_1 \times U_n \). This way we test a single shortcut in time \( \mathcal{O}(nk^3) \).

#### 4.1 Non-intersecting Polygonal Closed Convex Sets

Consider first the case where the interiors of convex polygons \( U_1 \) and \( U_n \) do not intersect.
Observation 6. Given an uncertain curve modelled by convex polygons \( U = \{U_1, \ldots, U_n\} \) with the interiors of \( U_1 \) and \( U_n \) not intersecting, note:

- There are two outer tangents to the polygons \( U_1 \) and \( U_n \), and the convex hull of \( U_1 \cup U_n \) consists of a convex chain from \( U_1 \), a convex chain from \( U_n \), and the outer tangents.
- Let \( C_i \) be the convex chain from \( U_i \) that is not part of the convex hull for \( i \in \{1, n\} \); then

\[
\max_{\pi \in \mathcal{U}} d_H(\pi, (\pi(1), \pi(n))) \leq \varepsilon \iff \max_{\pi \in \mathcal{U}, \pi(1) \in C_1, \pi(n) \in C_n} d_H(\pi, (\pi(1), \pi(n))) \leq \varepsilon.
\]

To see that the second observation is true, note that one direction is trivial. In the other direction, note that any line segment \( pq \) with \( p \in U_1, q \in U_n \) crosses both \( C_1 \) and \( C_n \), say, at \( p' \in C_1 \) and \( q' \in C_n \). We know that there is a valid alignment for \( p'q' \), both for the Hausdorff and the Fréchet distance; we can then use this alignment for \( pq \). See Figure 2.

We claim that we can use the following procedure to check \( \max_{\pi \in \mathcal{U}} d_H(\pi, (\pi(1), \pi(n))) \leq \varepsilon \).

1. Triangulate the region \( R \) bounded by two convex chains \( C_1 \) and \( C_n \) and the outer tangents.
2. Check that \( \max_{\pi \in \mathcal{U}, \pi(1) = t, \pi(n) = s} d_H(\pi, st) \leq \varepsilon \) for each line segment \( st \) of the triangulation with \( s \in C_1, t \in C_n \).

First of all, observe that we can compute a triangulation (see Figure 2), and that every triangle has two points from one convex chain and one point from the other chain. If all three points were from the same chain, then the triangle would lie outside of \( R \). Now consider some line segment \( pq \) with \( p \in C_1, q \in C_n \). To complete the argument, it remains to show that the checks in step 2 mean that also \( \max_{\pi \in \mathcal{U}, \pi(1) = p, \pi(n) = q} d_H(\pi, pq) \leq \varepsilon \). Observe that the triangles span across the region \( R \), so when going from one tangent to the other within \( R \) we cross all the triangles. Therefore, we can number the edges of the triangles that go from \( C_1 \) to \( C_n \), in the order of occurrence on such a path, from \( 1 \) to \( k \). We can establish a Hausdorff alignment, as shown intuitively in Figure 3; denote the alignment established on line \( j \in [k] \) with the sequence of \( a_i^j \), for \( i \in \{2, \ldots, n - 1\} \). We can establish polygonal curves \( A_i := (a_i^1, \ldots, a_i^k) \); they all stay within \( R \) (see Figure 4). We claim that for any line segment \( pq \) defined above, we can establish a valid alignment from intersection points of \( pq \) and \( A_i \).
Figure 4 An example set of curves $A = \{A_2, A_3\}$ discussed in Lemma 7.

Lemma 7. Given a set of curves $A := \{A_2, \ldots, A_{n-1}\}$ in $R$ described above and a line segment $pq$ with $p \in C_1$, $q \in C_n$, we have $\max_{\pi \in \mathcal{U}, \pi(1) = p, \pi(n) = q} d_H(\pi, pq) \leq \varepsilon$.

Proof. Note that $pq$ crosses each $A_i$ at least once. We can take any one crossing for each $i$ and establish the alignment. Consider such a crossing point $p_i'$. It falls in some triangle bounded by a segment from either $C_1$ or $C_n$ and two line segments that contain points $a_i^j$ and $a_i^{j+1}$ for some $j \in [k-1]$. We know, using Lemma 5, that $\max_{w \in U_i} \|a_i^j - w\| \leq \varepsilon$ and $\max_{w \in U_i} \|a_i^{j+1} - w\| \leq \varepsilon$. Consider any point $w' \in U_i$. Then, using Lemma 3 with $c := d := w'$, we find that $\|w' - p_i'\| \leq \varepsilon$. Therefore, also $\max_{w \in U_i} \|p_i' - w\| \leq \varepsilon$; using Lemma 5, we conclude that indeed $\max_{\pi \in \mathcal{U}, \pi(1) = p, \pi(n) = q} d_H(\pi, pq) \leq \varepsilon$. ▷

The proof of Lemma 7 shows how to solve the problem for two convex polygons with non-intersecting interiors. We can also use it directly for the case of line segments that do not intersect except at endpoints. Furthermore, in this particular case it is not necessary to use a triangulation, so we can get rid of one of the three resulting line segments.

Lemma 8. Given $n \in \mathbb{N}^{\geq 3}$, for any uncertain curve modelled with line segments $U = \{U_1, \ldots, U_n\}$ with $U_i = p_i^1p_i^2 \subset \mathbb{R}^2$ for all $i \in [n]$, given a threshold $\varepsilon \in \mathbb{R}^{> 0}$, and given that $U_1 \cap U_n \subset \{p_i^1, p_i^2\}$, and assuming that the triangles $p_i^1p_i^2p_n^1$ and $p_i^2p_n^1p_n^2$ form a triangulation of the convex hull of $U_1 \cup U_n$, we have $\max_{\pi \in \mathcal{U}} \delta(\pi, (\pi(1), \pi(n))) \leq \varepsilon$ if and only if

$$\max \left\{ \max_{\pi \in \mathcal{U}, \pi(1) = p_i^1, \pi(n) = p_n^1} \delta(\pi, p_i^1p_n^1), \max_{\pi \in \mathcal{U}, \pi(1) = p_i^2, \pi(n) = p_n^2} \delta(\pi, p_i^2p_n^2) \right\} \leq \varepsilon.$$

4.2 Intersecting Polygonal Closed Convex Sets

We proceed to discuss the situation where the interiors of $U_1$ and $U_n$ intersect, or where line segments $U_1$ and $U_n$ cross.

Line segments. Assume line segments $U_1 \triangleq p_1^1p_1^2$ and $U_n \triangleq p_n^1p_n^2$ cross; call their intersection point $s$. Then we can use Lemma 8 separately on pairs of $\{p_i^1s, sp_n^2\} \times \{p_i^1s, sp_n^2\}$. Clearly, together this will cover the entire set of realisations of $pq$ with $p \in U_1$, $q \in U_n$.

Lemma 9. Given $n \in \mathbb{N}^{\geq 3}$, for any uncertain curve modelled with line segments $U = \{U_1, \ldots, U_n\}$ with $U_i = p_i^1p_i^2 \subset \mathbb{R}^2$ for all $i \in [n]$, given a threshold $\varepsilon \in \mathbb{R}^{> 0}$, we can check that $\max_{\pi \in \mathcal{U}} d_H(\pi, (\pi(1), \pi(n))) \leq \varepsilon$.

Convex polygons. Convex polygons whose interiors intersect can be partitioned along the intersection lines, so into a convex polygon $R := U_1 \cap U_n$ and two sets of polygons $\mathcal{P}_1 := \{P_1^1, \ldots, P_k^1\}$ and $\mathcal{P}_n := \{P_n^1, \ldots, P_\ell^1\}$ for some $k, \ell \in \mathbb{N}^{> 0}$. We can look at pairs from $\mathcal{P}_1 \times \mathcal{P}_n$ separately. The pairs where $R$ is involved are treated later. Consider some $(P, Q) \in \mathcal{P}_1 \times \mathcal{P}_n$. Note that $P$ and $Q$ are convex polygons with a convex cut-out, so the
boundary forms a convex chain, followed by a concave chain. We need to compute some convex polygons $P'$ and $Q'$ with non-intersecting interiors that are equivalent to $P$ and $Q$, so that we can apply the approaches from Section 4.1.

We claim that we can simply take the convex hull of $P$ and $Q$ to obtain $P'$ and $Q'$. Clearly, the resulting polygons will be convex. Also, the concave chains of $P$ are bounded by points $s$ and $t$ and are replaced with the line segment $st$; same happens for $Q$ with point $u$ and $v$. The points $s,t,u,v$ are points of intersection of original polygons $U_1$ and $U_n$, so they lie on the boundary of $R$, and their order along that boundary can only be $s,t,u,v$ or $s,t,v,u$. Thus, it cannot happen that $st$ crosses $uv$, and it cannot be that $uv$ is in the interior of the convex hull of $P$, as otherwise $R$ would not be convex. Hence, the interiors of $P'$ and $Q'$ cannot intersect, so they satisfy the necessary conditions.

Finally, we need to show that the solution for $(P',Q')$ is equivalent to that for $(P,Q)$. One direction is trivial, as $P \subseteq P'$ and $Q \subseteq Q'$; for the other direction, consider any line segment that leaves $P$ through the concave chain. In our approach, we test the lines starting in $s$ and $t$: the established alignments are connected into paths. The paths $A_i$ do not cross $st$. So, any alignment in the region of $\text{CH}(P \cup Q) \setminus (P \cup Q)$ can also be made in the region $\text{CH}(P' \cup Q') \setminus (P' \cup Q')$. So, this approach yields valid solutions for all pairs not involving $R$.

Now consider the pair $(R,R)$. A curve may now consist of a single point, so all the points of $U_i$ need to be close enough to all the points of $R$. To check that, observe that the pair of points $p \in U_i$ and $q \in R$ that has maximal distance has the property that $p$ is an extreme point of $U_i$ in direction $qp$ and $q$ is an extreme point of $R$ in direction $pq$. So, it suffices, starting at the rightmost point of $U_i$ and leftmost point of $R$ in some coordinate system, to then rotate clockwise around both regions keeping track of the distance between tangent points. Note that only vertices need to be considered, as the extremal point cannot lie on an edge. Finally, any other pair that involves $R$ is covered by the stronger case of $(R,R)$: for any line we can align every intermediate object with any point in $R$.

**Lemma 10.** Given $n \in \mathbb{N}^\geq 3$, for any uncertain curve modelled with convex polygons $U = (U_1, \ldots, U_n)$ with $U_i \subseteq \mathbb{R}^2$ for all $i \in [n]$ and $V(U_i) = \{p_1^i, \ldots, p_k^i\}$ for all $i \in [n]$, $k \in \mathbb{N}^>0$, given a threshold $\varepsilon \in \mathbb{R}^>0$, we can check that $\max_{\pi \in U} d_{\text{H}}(\pi, (\pi(1), \pi(n))) \leq \varepsilon$.

## 5 Combining Steps

In the previous sections, we have shown how to check if a shortcut of length $n \geq 3$ is valid under the Hausdorff distance, for indecisive points and polygonal closed convex sets. It is easy to see that a shortcut of length $n = 2$ is always valid. Therefore, we can use the previously described procedures to construct a shortcut graph; any path in such a graph from the vertex 1 to vertex $n$ corresponds to a valid simplification, so the shortest path gives us the result we need. The graph has $O(n^2)$ edges. These observations lead to Theorem 1.


