How To Take Short Cuts*

Claire Kenyon¹ and Richard Kenyon²

¹ LIP, CNRS URA 1398, Ecole Normale Supérieure de Lyon,
46, allée d'Italie, 69364 Lyon Cedex 07, France
kenyon@lip.ens-lyon.fr

² Institut Fourier, B.P. 74, 38402 Saint Martin d'Hères, France

Abstract. We present an elementary construction for adding short cuts on a rectifiable curve of the plane, so that any two points on the curve are linked by a path along the curve and short cuts, of total length at most a constant times their Euclidean distance. The total length of the short cuts is a constant times the perimeter of the curve.

1. Introduction

Given a smooth curve or polygon P of finite length in the plane, can we add “short cuts” to P of small total length so that any two points on P are within distance (along the curve and short cuts) at most a constant factor away from their Euclidean distance?

Jones has solved this problem in the affirmative [1]:

Theorem 1 [1]. *There are constants C_0, C_1 such that for every polygon P there is a rectifiable set Γ , $P \subset \Gamma$, with total length $L(\Gamma) \leq C_0 L(P)$ satisfying: if $x, y \in P$ there is a subarc $\gamma \subset \Gamma$ from x to y with total length $L(\gamma) \leq C_1 |x - y|$, where $|x - y|$ is the Euclidean distance.*

Definition 1. Such a γ is called a short path between x and y .

* This work was done while Claire Kenyon was at Ecole Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 05, France, and while Richard Kenyon was at IHES, 35 route de Chartres, 91440 Bures-sur-Yvette, France.

Actually he proves this (and more) for P any connected rectifiable set. A rectifiable curve is a curve which is the image of $[0, 1]$ under a Lipschitz map (in other words, it has a well-defined length, at least locally). His proof, however, uses the Riemann mapping from P to the circle, which is well known to be difficult to compute, so the construction in [1] is not necessarily computable.

We supply here a geometric proof of the above theorem. Our construction uses the *skeleton* of the polygon P , a subgraph of the Voronoi diagram of the edges. If there exists a finite set of “short cut” segments with the necessary length and path properties, our construction is an algorithm which finds and outputs such a finite set of segments. Otherwise, an infinite number of segments must be added, so our algorithm outputs a finite-length description of the set of segments.

This construction may necessarily add an infinite number of segments, but in practice when only finite precision is needed, it is finite. The construction also generalizes to smooth curves and planar graphs. It can be implemented easily, the two major subroutines being a convex hull algorithm and a Voronoi diagram algorithm (for edges).

In Section 2 we describe the construction of the short cuts across the polygon. In Section 3 we give an algorithm for finding a short path (using P and the short cuts) between any pair of points on the boundary of P . In Section 4 we analyze the total length of the short cuts added, and show that it is at most a constant times the length of P . In Section 5, given a pair of points on the boundary, we analyze the length of the path found by our algorithm, and show that it is at most a constant times their Euclidean distance. In Section 6 we extend the result to smooth curves and present some open problems.

The analysis initially depends on a few parameters a, α, β ; various expressions appear throughout which depend on these parameters. These expressions were chosen so as to minimize the resulting constants: at the end we choose values for the parameters a, α, β which again minimize the constant C_1 .

2. The Construction

We need to construct a family of segments linking together points of the polygon which are too far apart along the polygon. We first construct the convex hull of P . This divides the plane into several connected regions. All the finite regions (which we call R_i) are polygons.

2.1. Skeleton of a Polygonal Region

We construct short cuts in each region separately, using its skeleton.

Definition 2. The skeleton of the polygonal boundary of R_i is the set of centers of circles contained in R_i and which touch the boundary at two or more points. Such circles are called Voronoi circles.

The skeleton is a subgraph of the Voronoi diagram of the edges of the boundary. It is easy to check the following property:

Lemma 1. *The skeleton is a tree whose leaves are exactly the convex vertices of the boundary, and whose branching points have degree 3 exactly, except in degenerate cases.*

The edges of the skeleton are formed with pieces of straight lines and pieces of parabolas: straight lines for the centers of all the Voronoi circles which pass through two concave vertices of the boundary, and for the centers of all the Voronoi circles which are tangent to two edges of the boundary; parabolas for the centers of all the Voronoi circles which go through a concave vertex and are tangent to an edge. We first deal with all the cases when we need to add an infinite number of short cuts.

2.2. Trimming the Acute Vertices

The first operation is called “trimming the acute leaves of the skeleton.” Let a be a constant greater than 1. A vertex v of P is called convex if the interior angle at v is at most π . We define an *acute* vertex to be a convex vertex v of the boundary such that the angle at v is smaller than $2 \operatorname{Arcsin}(1/a)$. For each such vertex, we consider the straight edge of the skeleton leading to that vertex, and choose some point c_0 on that edge. It is the center of a circle tangent at points p_0 and q_0 to the two edges adjacent to v . We add segment p_0q_0 to the (initially empty) list of short cuts. We move from c_0 on the skeleton toward v , adding to the list of short cuts the segments p_iq_i defined inductively as follows: given c_{i-1} , p_{i-1} , q_{i-1} , we move from c_{i-1} on the skeleton toward v . Any point c corresponds to two points p and q on the boundary; let $\gamma(p, q)$ be the path

$$p \rightarrow p_{i-1} \rightarrow q_{i-1} \rightarrow q$$

going from p to p_{i-1} along the boundary, along the short cut $p_{i-1}q_{i-1}$, and from there along the boundary to q . As we move toward v , $|\gamma(p, q)|$ increases and $|p - q|$ decreases. We define c_i to be the point such that $|\gamma(p_i, q_i)| = a|p_i - q_i|$. See Fig. 1 for an illustration.

Fig. 1. Trimming the acute vertices.

Thus we add an infinite sequence of cuts which get shorter and shorter as we get closer to v . We can now erase the part of the skeleton from c_0 to v . After this is done for every acute vertex of the skeleton, we are finished with trimming. (Note that any construction must necessarily add an infinite number of short cuts in the neighborhood of acute vertices.)

2.3. Main Construction

In the rest of the construction we only add a finite number of short cuts. We choose to root the trimmed skeleton at an arbitrary point on an edge leading to a leaf. The main part of the construction of short cuts is in bottom-up order from the leaves to the root of the trimmed skeleton, erasing the skeleton as we move up it. In other words, we only deal with a branching point once all its descendants have been dealt with (and so the skeleton below the branching point is all erased), and we deal with an edge, once its endpoint furthest from the root has been erased, by walking along the edge from that endpoint toward the root, erasing the edge as we move along.

Let p and q be two points of the original polygon.

Definition 3. A hop $\gamma(p, q)$ between p and q is a path along the boundary and the short cuts, with total length at most $a|p - q|$.

As we perform our construction, we keep the following invariant property:

Invariant. If c , a point on the skeleton, has been erased, then there are hops joining the boundary points on the Voronoi circle of center c .

We observe that this property is indeed true immediately after the original trimming: in fact, the convex vertices with small angle have been dealt with precisely so that the property would be satisfied.

We now need to define how we deal with an edge or with a branching point of the skeleton. If c is our current position, two cases arise: either the circle associated to c touches the boundary at exactly two points p and q , or c is a vertex of the skeleton and touches the boundary at three points p , q , and r . (We assume that the polygon is not degenerate. If it is degenerate, take a nondegenerate polygon sufficiently close; the construction is then the same as in Section 6).

Case 1: dealing with an arc between vertices c_1 and c_2 in the skeleton. Let c_1 be the endpoint furthest from the root. c_1 is already erased, so there are hops between the boundary points p_1, q_1 on its Voronoi circle. As we move from c_1 to c_2 , let p, q be our current boundary points, and let $\gamma(p, q)$ be the path from p to q which goes through p_1 and q_1 . If $\gamma(p, q)$ is a hop, we erase c and keep going. If $\gamma(p, q)$ equals $a|p - q|$, we add segment pq to the list of short cuts, and trim the boundary further by replacing the part followed by $\gamma(p, q)$ by the segment pq . See Fig. 2.

Case 2: dealing with a branching point. The Voronoi circle at c touches the boundary at p, q, r . Suppose that the root of the skeleton is “between”

Fig. 2. Constructing the short cuts: case 1.

p and r . We know that there are already hops between p and q and between q and r . Let $\gamma(p, r)$ be the concatenation of those two hops. As in the first case, if $\gamma(p, r) \geq a|p - r|$ we add segment pr to the list of short cuts and trim the boundary by replacing $\gamma(p, r)$ by segment pr ; otherwise we do nothing. Then we erase the branching point. See Fig. 3.

2.4. Termination of the Construction

For the construction to be well defined, we only need to show that the process in the main part of the construction is finite.

Lemma 2. *The process is finite.*

Proof. If we remove small neighborhoods around each vertex, any two points on the polygon are either on the same edge or at distance at least ε apart, $\varepsilon > 0$. Thus the segments we add during the main part of the construction all have length at least ε . Every time we add a segment, the length of the boundary of the remaining region is at most the length of the boundary of the previous region, minus $(a - 1)\varepsilon$. Since the original boundary has finite length, we can only add a finite number of segments.

As for the vertices which are a nontrimmed leaf of the skeleton, they must be convex with angle larger than $2 \arcsin(1/a)$, and so any two points on the same Voronoi circle in the neighborhood of such a vertex are joined by a hop (obtained

Fig. 3. Constructing the short cuts: case 2.

by simply following the boundary). Thus no short cuts are added in the neighborhood of that vertex. $\square$

We cannot bound the number of segments added independently of the polygon; in fact no such bound holds for *any* short cut construction: a polygon which has two parallel edges very close to each other must have a large number of short cuts. If we are operating under the assumption of finite resolution, however, we ignore edges of length less than ε , and this lemma gives a bound on the number of other edges needed since the total length is bounded: namely,

$$\# \text{ short cuts} \leq \frac{|\partial P|}{(a-1)\varepsilon}.$$

After this construction is done, all the skeleton is erased, and so, for any two points p and q which are on the same Voronoi circle, there is a hop between p and q . In the next section we forget about the actual construction of the short cuts and only use that property:

Whenever we are on a Voronoi circle, we can hop across to any other point on the circle with cost at most equal to a times the Euclidean distance spanned.

3. Finding the Short Path

Once we have constructed the short cuts of P , we need to describe how to get quickly from one point on P to another.

3.1. Definitions and Basic Properties

Let l be the straight line segment between two points p and q of the polygon.

Remark 1. We can assume that l does not cross the boundary of P .

In fact, if it does cross the boundary, consider the short paths between successive intersections of P with l ; the concatenation of these short paths gives a short path from p to q .

Let $c(p)$ be the center of the Voronoi circle $C(p)$ which touches p (in the region R containing l). Similarly define $c(q)$. Let $V(p, q)$ denote the unique shortest path in the skeleton between $c(p)$ and $c(q)$. The following property holds:

Lemma 3. $V(p, q)$ is monotone increasing from $c(p)$ to $c(q)$, in direction $q - p$.

Proof. Since l does not cross the boundary of P , points p and q see each other. Every Voronoi circle on $V(p, q)$ must touch ∂P at least at one point above line

(pq) and at one point below line (pq) , and its interior must be empty. Those conditions cannot be simultaneously fulfilled by two Voronoi circles whose centers project to the same point of line (pq) . $\square$

Definition 4. A path is called legal if it follows the boundary and short cuts of R .

3.2. Correspondence Between General Paths and Legal Paths

For any unconstrained path δ from p to q , contained in region R , there is a corresponding legal path γ_δ from p to q ; we define γ_δ in the following way.

The skeleton partitions R into (closed) regions which we call *skeleton cells*. We have the initial points $\gamma_\delta(0) = \delta(0) = p$. Let c_0 be the first point at which δ leaves the skeleton cell $S(p)$ containing p ; let C_0 be the Voronoi circle with center c_0 . Then C_0 contains at least two points of the boundary of R , and exactly one point x_0 in the cell $S(p)$. Define the initial segment of γ_δ to be the unique path on ∂R which goes from p to that point x_0 , and is contained in the region $S(p)$.

The path δ moves from c_0 into a new skeleton cell S' , and C_0 has exactly one point x'_0 on R in this cell.

By the construction of our short cuts, there is a hop from x_0 to x'_0 ; extend γ_δ by this hop. This defines γ_δ from p to x'_0 , and by repeating the same construction from cell to cell we define γ_δ for the entire path δ .

3.3. Construction of the General Path

To define a short path from p to q , we consider first the segment l from p to q . Unfortunately, the corresponding path γ_l around R constructed in the above manner may be very inefficient (see Fig. 4). However, the inefficiencies are only caused by l intersecting $V(p, q)$ a large number of times at a small angle; we perturb l slightly, creating a new path $\tilde{l}$ for which $\gamma_{\tilde{l}}$ is more efficient.

Let α be a given angle, a constant of our construction. We define $\tilde{l}$ piece by piece, so that it follows l as closely as possible, but is constrained to lie on one side of $V(p, q)$, unless $V(p, q)$ diverges from l too steeply. We use here the fact that

Fig. 4. An inefficient path from p to q may be obtained by following l exactly.

$V(p, q)$ is monotone in direction l (hence the terms *diverges from* and *approaches l* make sense).

More precisely, $\tilde{l}$ starts at p , and goes straight toward q . Let c_0 be any intersection point of $V(p, q)$ with the part of $\tilde{l}$ defined up to this point. We can assume (by perturbing p or q slightly) that c_0 is not a vertex of the skeleton; since $V(p, q)$ at c_0 is then smooth, we can measure the angle of l and $V(p, q)$ at c_0 . If this angle is larger than α , our predefined constant (that is, $V(p, q)$ *diverges from* l at angle $\geq \alpha$), we let $\tilde{l}$ cross $V(p, q)$ and continue straight toward q into the new region.

On the other hand, if this angle is smaller than α or if $V(p, q)$ approaches l , we let $\tilde{l}$ follow $V(p, q)$, so as to remain on the same side of $V(p, q)$ (remaining just slightly to one side), until we either arrive at $c(q)$, or $V(p, q)$ bends back toward q , or $V(p, q)$ starts diverging from l with angle larger than α . In each case we then continue straight toward q , crossing $V(p, q)$ in the last case.

This completely defines the path $\tilde{l}$ from p to q , and (by the construction of Section 3.2) a path $\gamma_{\tilde{l}}$ (the short path) along R and the short cuts of R from p to q .

4. Length of the Construction

Let $L(P)$ be the length of the polygon P . The convex hull of P has length at most $L(P)$. Let $L(R)$ be the length of the boundary of a region R .

Let us first analyze the length of the segments added during the preliminary trimming. The length of the short cuts $p_0q_0, p_1q_1, \dots$ (See Section 2) form a decreasing geometric progression. Since the angle at the vertex v is at most $2 \operatorname{Arcsin}(1/a)$, it is easy to see that

$$\sum_{i \geq 0} |p_i - q_i| < \frac{1}{a-1} (|p_0 - v| + |v - q_0|).$$

After trimming the boundary, the boundary of the trimmed region R' has length

$$\begin{aligned} L(R') &= L(R) + \sum_{v \text{ trimmed}} (|p_0 - q_0| - |p_0 - v| - |v - q_0|) \\ &\leq L(R) + \sum_{v \text{ trimmed}} (|p_0 - v| + |v - q_0|) \left(\frac{1}{a} - 1 \right). \end{aligned}$$

In the main construction, whenever we add the short cut pq , the boundary gets trimmed, and its length is reduced by at least $(a-1)|p-q|$. In the end it is still nonnegative. Thus,

$$L(R') \geq (a-1) \sum_{\text{cuts } pq} |p-q|.$$

The total length of the short cuts added in region R is overall at most

$$\begin{aligned} & \frac{1}{a-1} L(R') + \sum_{v \text{ trimmed}} \frac{1}{a-1} (|p_0 - v| + |v - q_0|) \\ & \leq \frac{1}{a-1} \left(1 + \frac{1}{a}\right) L(R). \end{aligned}$$

As we sum over all the regions, every edge of P is counted at most twice, and every edge of the convex hull is counted only once. Thus the total length of the construction is at most

$$\frac{3(a+1)}{a(a-1)} L(P).$$

Property 1. *The short cuts built in Section 2 have total length at most*

$$\frac{3(a+1)}{a(a-1)} L(P).$$

5. Analysis of Path Length

Let p and q be two points of polygon P which are visible across region R . In Section 3 we defined a legal path going from p to q . We must now show that this path γ really is a *short path* between p and q , i.e., its length is at most a constant times $|p - q|$.

The path γ can be decomposed into several parts: sections where the path follows the boundary of R between hops, sections where it hops across $V(p, q)$ (when $V(p, q)$ diverges from l with an angle larger than α), and sections where it hops across other Voronoi circles (When $\tilde{l}$ crosses skeleton edges which are not in $V(p, q)$).

In our analysis we use the fact that $\tilde{l}$ is not too long compared with $|p - q|$.

Lemma 4. *The length of $\tilde{l}$ is at most*

$$|p - q| \frac{1 + \sin \alpha}{\cos \alpha}.$$

Proof. $\tilde{l}$ is a path from p to q , monotone in direction l , and can never diverge from l at an angle steeper than α . $\square$

5.1. Analysis of the Boundary Sections

Consider a section where the path γ follows the boundary. Except near p and q , the corresponding section of $\tilde{l}$ goes between two successive intersections c_1 and

Fig. 5. A concave path as shown cannot be longer than $|c_1 - c_2|$.

c_2 of $\tilde{l}$ with the skeleton. Thus the boundary section is all within one skeleton cell, hence concave.

Case 1: the Voronoi circles of centers c_1 and c_2 intersect. The concave path must stay outside the circles, and so its length is at most equal to the distance from c_1 to c_2 (see Fig. 5).

Thus the length of the boundary section is at most equal to the length of the part of $\tilde{l}$ between c_1 and c_2 .

Case 2: the circles of centers c_1 and c_2 are disjoint.

Lemma 5. *In that case, the length of the boundary section is at most equal to*

$$2\left(1 + \frac{1}{\cos \alpha}\right)|c_1 - c_2|.$$

Proof. The concave path must not intersect the section of $\tilde{l}$ between c_1 and c_2 . The path $\tilde{l}$ is not necessarily a straight line between c_1 and c_2 , but it is constrained with respect to direction l ; thus $\tilde{l}$ has to lie inside a certain quadrilateral, and so does the concave path. Elementary geometric arguments (project the path outward to the boundary of the parallelogram) give the bound, see Fig. 6. $\square$

Case 3: boundary sections near p and q . Initially, we follow the boundary of the region from p to the point on the first hopping circle. Let c_1 be the center of the circle, and let p_1 be the first hopping point. The boundary between p and p_1 must be concave, and cannot intersect either l or the first hopping circle, see Fig. 7. Therefore this initial length is at most $|p - c_1|$. The section near q is dealt with similarly.

Fig. 6. A concave path between disjoint circles must lie within a quadrilateral.

(this follows from elementary geometric considerations: the region is starred around c_i , has radius at most $(1 + \beta)r_i$, and none of the tangents to the boundary has slope greater than $1/\sqrt{1 - \beta^2}$ with respect to the circle C_i). Thus the boundary of R has length at most $c(\beta)\sum_{\text{batches}} r_i$. On the other hand, we have

$$L(\tilde{l}) \geq \sum |c_i - c_{i+1}| \geq \sum_{\text{batches}} \beta r_j.$$

Thus we have

$$L(\partial R) \leq \frac{c(\beta)}{\beta} L(\tilde{l}).$$

This is minimized for $\beta = (\sqrt{5} - 1)/2$, and we arrive at the constant stated. $\square$

We now look at the path γ . For every hop between two points above $V(p, q)$, we map the hopping segment onto the arc of circle on its left, and for every hop between two points below $V(p, q)$, we map the hopping segment onto the arc of circle on its right. See Fig. 8. Two different hops which do not cross $V(p, q)$ must have disjoint images under this map, and the image is on the boundary of R . Thus the total length of the hops which are not across $V(p, q)$ is at most equal to the boundary of R , hence $\leq 24L(\tilde{l})$.

For the hops in the batch which do cross $V(p, q)$, we project them onto direction l ; if α is not too small (in particular, if $\sin(\alpha/2) > \beta/(1 - \beta)$), the projections of these hops are disjoint. Thus we have a contribution from these hops of

$$\sum_{\text{batches}} \frac{\text{diam}(\text{batch})}{\sin \alpha} < \frac{2(1 + \beta)L(\tilde{l})}{\beta \sin \alpha}.$$

Finally, if we consider all the contributions to γ , we find that the total length is $O(|p - q|)$, hence the theorem. The constants we obtain by minimizing over β and α ($\alpha = 0.18\pi$ and $\beta = 0.22$) are

$$C_1 = 81.5a \quad \text{when} \quad C_0 = 3 \frac{(a + 1)}{a(a - 1)}.$$

Property 2. *The short path constructed in Section 3 has length at most $81.5a|p - q|$.*

There is, however, a lot of room for more careful geometric analysis for the constant C_1 , which is in practice much smaller.

6. Smooth Curves and Further Generalizations

Having solved the short cut problem for polygons, we can solve it for smooth curves as follows.

Given a smooth curve S , define $\varepsilon > 0$ so that any disk of diameter less than ε intersects S in a connected set, and in any disk of diameter ε centered on a point of S , the curve S does not deviate much from a straight line.

A map is (c, c') -bi-Lipschitz if

$$c|x - y| \leq |\pi(x) - \pi(y)| \leq c'|x - y|$$

for all x and y . Take a polygon P approximating S to within $\varepsilon/10$, such that the vertices of the polygon are not too acute, and let π be a map which is a one-to-one bi-Lipschitz projection of the polygon onto the curve (in which π moves points by at most $\varepsilon/10$, and the Lipschitz constants are close to 1). We can find such a polygon and π since the curve S looks locally flat.

When we construct short cuts for the polygon P , there are no trimmed vertices, since the vertices of P are sufficiently obtuse. Given a short cut pq for P of length at least ε , we construct the short cut $\pi(p)\pi(q)$ on S . Note that this short cut has length at most 1.2 times the length of pq .

These are the only short cuts we use on S ; their total length is at most 1.2 times the total length of the short cuts on P . We claim that they allow us to construct short paths for pairs of points on S .

To construct a short path from s_1 to s_2 if $|s_1 - s_2| < \varepsilon$, simply follow the boundary, which is approximately a straight line from s_1 to s_2 . If $|s_1 - s_2| > \varepsilon$, first construct a short path from the corresponding points $\pi^{-1}(s_1)$ to $\pi^{-1}(s_2)$ on P ; and then project the path using π to the set S and its short cuts. The length of the projection is at most $\max\{1.2, \text{Lip}_\pi\}$ times the length of the short path on P .

The preceding construction works for arbitrary rectifiable sets in the plane; these are the most general sets which are not locally "twisted," which is the condition we need to make this approximation by polygonal shortcuts work.

There are several open problems related to this work:

1. Lipschitz partitions. Given a polygon (or rectifiable set) in the plane, is it possible to partition its interior into Lipschitz regions, i.e., regions such that, for any two points p and q on the boundary of the region, following the boundary gives a short path between p and q ? Note that our construction does not answer that question, since it does not give short paths between points on the short cuts. Actually, Jones [1] solves the Lipschitz partition problem, but with no simple computable solution.
2. Disconnected sets. In our construction the connectedness of ∂P guarantees that the skeleton is a tree. How do we deal with a set of segments which do not necessarily form a polygon? Is there a construction with optimal length? In that case the total length of the input segments is irrelevant: a cloud of points, for example (segments of length 0), requires many short cuts (in fact, a minimum spanning tree at least), while the input segments have total length 0.
3. Higher dimensions. Is it possible to generalize the construction to polyhedra? Is the length of the short cuts within a constant of optimal? Here the optimal

length of the short cuts cannot be related to the length of the edges or to the area of the two-dimensional facets of the polyhedron, as easy counter-examples show.

References

- [1] P. Jones, Rectifiable sets and the traveling salesman problem, *Invent. Math.*, **102**, 1–15 (1990).

Received July 29, 1991.