Approximation of an open polygonal curve with a minimum number of circular arcs

R. L. Scot Drysdale* Günter Rote†§ Astrid Sturm‡§

Abstract

An algorithm for approximating a given open polygonal curve with a minimum number of circular arcs is introduced. In computer-aided manufacturing environments the paths of cutting tools are usually described with circular arcs and straight line segments. Greedy algorithms for approximating a polygonal curve with curves of higher order can be found in the literature. Without theoretical bounds it is difficult to prove anything about the quality of these algorithms. We present an algorithm which allows us to build a directed graph of all possible arcs and look for the shortest path from the start point to the end point of the polygonal curve. We can prove a runtime of $O(n^2 \log n)$, for n the number of vertices of the original polygonal chain.

1 Introduction

In computer-aided manufacturing enviroments tool paths are usully made of line segments and circular arcs [3, 4, 5]. Approximating data by curves of higher order [1, 2, 3, 4, 5, 6, 7, 8] has been investigated extensively in the past. The theoretical bounds of these problem are not as well studied.

We wish to approximate a polygonal chain $P = (p_1, \dots, p_n)$ by a series of circular arcs (which could include straight lines, as circles of infinite radius). The endpoints of the arcs are vertices of P. We want our approximating curve to have distance at most ε from P. As a first approximation to this problem, one can look at a region formed from strips of width ε centered at the polygon edges. However, in the vicinity of sharp corners, this does not guarantee that the curve remains close to the given points. Figure 1 shows a circular piece of a hypothetic curve that can shortcut the bend at p_4 if it is only required to remain in the strips. (Also, it might overshoot the bend, as indicated in the vicinity of p_6, although this looks like a theoretical possibility only.) To avoid this, we introduce a *gate* at every vertex. The approximating curve is required to pass through all gates in succession, and the curves are not allowed to pass through a gate twice. This will guarantee that any curve into a point p_i can be joined with any curve out of p_i without danger of an intersection other than at p_i.

For our problem, we assume that we are given a polygonal "tolerance region" R and a sequence of gates $g_1, g_2, \dots, g_n$, which are segments through the points p_i. We will refer to endpoints of gates lying to the left of P as we walk from p_1 to p_n as left endpoints and the other endpoints as right endpoints. We require that the gates do not cross. We require that R is a simple polygon passing through all gate endpoints, that R does not intersect the interior of gates or cross the segments connecting corresponding endpoints of successive gates, and that the sections of R connecting g_i with g_{i+1} do not cross the lines that extend g_i and g_{i+1}. (The line extending a segment is the line that contains that segment.)

Ideally, the gate g_i at vertex p_i is a line segment of length 2ε centered at p_i that bisects the angle $p_{i-1}p_ip_{i+1}$. For a convolved curve with sharp bends close together, we might have to shorten the gates and to reduce the ε-strip around the edges, as shown in the right part of Figure 1.

Modeling the curve approximation problem by an appropriate tolerance region with gates is a problem of its own, which we do not discuss here. In Figure 1, we have chosen to approximate the "ideal" circular boundary at the outer angle of each vertex by a single edge of P. One can use more edges to get a finer approximation, or one could also choose to approximate the circular arc from inside, to get a guaranteed upper distance bound of ε. Our time bounds assume that the size of R is proportional to n.

Definition 1 (proper gate stabbing) *A curve stabs gates $g_i, \dots, g_j$ properly, if and only if:*

- *the curve passes through gate $g_m \in \{g_i, \dots, g_j\}$ from the side of $\overline{p_{m-1}p_m}$ to the side of $\overline{p_mp_{m+1}}$*
- *the curve passes through every gate only once (the stabbing curve may pass through an endpoint of the gate in addition to passing through it once)*

*Department of Computer Science, Dartmouth College, scot@cs.dartmouth.edu

†Department of Computer Science, Free University Berlin, rote@inf.fu-berlin.de.

‡Department of Computer Science, Free University Berlin, sturm@inf.fu-berlin.de

§Partially supported by the IST Programme of the EU as a Shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS – Algorithms for Complex Shapes)

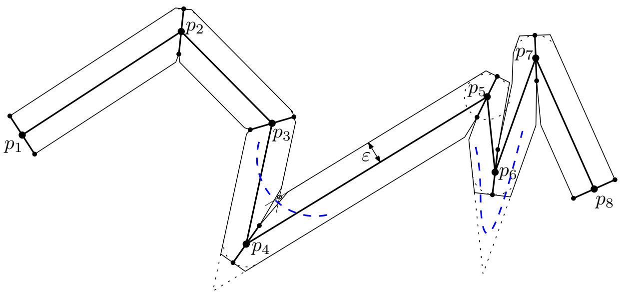

Figure 1: polygonal tolerance region

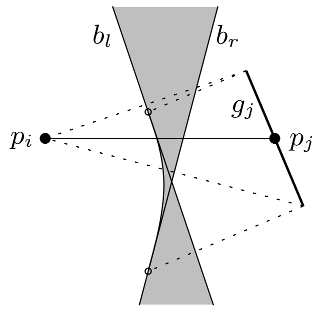

Figure 2: region of all centers of circles passing through p_i and gate g_j

Definition 2 (valid circular arc) *A circular arc a_{ij} with starting point p_i and endpoint p_j is a valid arc if:*

- *the arc stabs the gates $g_{i+1}, \ldots, g_{j-1}$ properly,*
- *the arc does not cross any piece of the boundary of the tolerance region R.*
- *the arc reaches p_i from the correct side of g_i and reaches p_j from the correct side of g_j.*

Note that because R passes through the gate endpoints, any arc that goes through a series of gates without crossing the tolerance boundary must go through them in the correct order, so we do not need to test for that separately.

We can split the problem of determining if a valid circular arc connects p_i with p_j into three parts. First, we compute all arcs between p_i and p_j that stab all intermediate gates properly. Second, we compute all arcs that start at p_i and end at p_j, reaching both from the correct side. Third, we compute all arcs between p_i and p_j that do not intersect with the tolerance boundary. A valid circular arc has to be member of all three result sets.

2 Stabbing through the gates

Definition 3 (point/gate bisectors) *Given a point p and a gate g, let b_l be the bisector of p and g's left endpoint and b_r be the bisector of p and g's right endpoint.*

Lemma 1 *The centers of all circles passing through a vertex p and intersecting a gate g exactly once lie in a double cone whose boundary is b_l and b_r. The sections we want are the ones where one half plane includes p and the other excludes it. (Figure 2 illustrates this.) In the degenerate case where b_l is parallel to b_r one "cone" is the strip between the bisectors and the other "cone" is empty.*

Proof. A case analysis of circles centered on the bisectors and in each of the regions confirms the claim. □

Lemma 2 *The region of the centers of all circles passing through a vertex p which are tangent to the gate g or intersect it twice forms a parabolic region (in Figure 2 the parabolic region is the filled region to the left of the double cone). The boundary of the parabolic region is given by a parabolic piece, defined by the centers of the circles which are tangent to the gate, and by two pieces of the boundary double cone. In the degenerate case when the bisectors are parallel the parabolic region is empty.*

Proof. Geometric analysis proves the claim. □

The centers of all circles which pass through a point p and intersect or are tangent to the gate are in the union of the double cone and the parabolic region described in Lemma 1 and Lemma 2. By Definition 1 an arc stabs the gates properly only if every gate is intersected only once. Circles centered in the parabolic region, intersect the gate twice or (on the parabola) never pass through the gate. Therefore we include only the two straight boundary segments of the parabolic region, which represent extreme cases of passing through a boundary point and an additional point on the gate. We exclude the rest of the parabolic region as centers for intermediate gates, even though in some circumstances when points on P are relatively close together an arc might end at an endpoint on P before it again intersects the gate. We will allow points in the parabolic region at endpoints of the arc.

According to Lemma 1, one cone is the region of the centers of disks including the left boundary point of the gate and excluding the right boundary point. Circular arcs centered in these region pass the gate from the correct side, according to the stabbing condition, if they are in CCW orientation. In CW orientation the arc would walk around the left boundary point before intersecting the gate. The unbounded part of this cone lies to the left of P. Symmetrically the circular arcs in the other cone need CW orientation to pass the gate in the correct direction, and the unbounded part of this cone lies to the right of P.

So from now on we talk about the left cone and the right cone. A circular arc stabbing through the gates is not allowed to change its orientation.

Lemma 3 *A circular arc a starting at a point p stabs gates $g_i, \dots, g_j$ properly if and only if its center lies in the intersection of the left cones defined by p and the gates, or the intersection of the right cones.*

Proof. Straightforward. □

Lemma 4 *Incremetally computing the two regions of centers of all valid circular arcs passing through a point p and stabbing all $g_i, \dots, g_j$ gates properly, can be done in $O(n \log n)$ time.*

Proof. It is the incremental intersection of $O(n)$ half planes. □

3 Arc endpoints

A valid circular arc from p_i to p_j must reach each point from the correct side of its gate. All arcs that start at p_i and end at p_j have their centers on the bisector of the segment connecting p_i and p_j. However, some of these arcs will approach the gate from the wrong side. We want to eliminate all such arcs, but allow arcs that are tangent to a gate at its defining point or which would circle back and cut the gate a second time if the arc did not stop. This means that we want to consider not only centers in the double cone, but also in the parabolic region.

All points on the bisector of p_i and p_j can be centers of arcs from p_i to p_j that reach p_j from the correct side of g_j, but most can only go around the circle in one direction. The exception is the circle which is tangent to p_j at g_j. We call the center of this circle the splitting point. All points on the ray of the bisector that starts at the splitting point and goes right are centers of CW arcs that meet p_j from the correct side. All points on the ray of the bisector that starts at the splitting point and goes left are valid centers of CCW arcs.

To determine which arcs meet p_i from the correct side of g_i we do the symmetrical test, with the roles of p_i and p_j reversed.

Lemma 5 *Let b be the perpendicular bisector of the segment between p_i and p_j. Let s_i be the point of b which is the center of a circle tangent to g_i at p_i, and let s_j be defined symmetrically. The centers of all CW arcs that reach both p_i and p_j from the correct side are all points on b to the right of both s_i and s_j. CCW arcs are symmetrical.*

Proof. This ray is the intersection of the CW rays for both endpoints of the arc. □

4 Tolerance boundary

We break the tolerance boundary R into two polygonal chains, one on each side of the original polygonal chain P. When dealing with CCW circles we will exclude the right chain, which we call the CCW boundary. When dealing with CW circles we will exclude the left chain, which we call the CW boundary. The requirements that the arcs pass through gates and that the tolerance boundary not intersect the interiors of gates or cross the segments connecting boundary points of successive gates guarantees that there will be no conflict with the other boundary.

A circle passing though point p does not intersect or contain any edge on a polygonal chain C if its center lies closer to p than to any point on C. That is, if we compute the Voronoi diagram of $C \cup p$, the center of the circle must lie in point p's region, $V(p)$.

This is not quite the condition that we want, namely that a circular arc does not cross chain C. The Voronoi region guarantees that an entire circle does not cross C. However, in our case these are equivalent.

Lemma 6 *If an arc from g_i to g_j does not intersect a tolerance boundary between g_i and g_j then neither does the circle on which that arc lies.*

Proof. For each pair of consecutive gates g_k and g_{k+1} we are given that the section of the arc between them does not intersect the section of the tolerance boundary between g_k and g_{k+1}. But this section of the tolerance boundary is not allowed to cross either the segment connecting its start and end points or the lines extending g_k and g_{k+1}. Therefore this section of the boundary cannot intersect the rest of the circle, either. □

While we could compute the entire Voronoi diagram of $C \cup p$ to determine $V(p)$, this would be too expensive. Fortunately, we can interatively add n consecutive segments of C and update p's Voronoi region $V(p)$ in $O(n)$ total time.

Voronoi regions are “generalized star shaped”. This means that a shortest segment from a boundary point to a nearest point in the shape defining the region lies entirely within the region.

Lemma 7 *Each segment added will either cause no change to $V(p)$ or will replace a section of $V(p)$ by at most three new segments (two straight lines and a parabola). (If $V(p)$ is unbounded we think of an edge “at infinity” connecting the two infinite rays, so that these three “segments” are considered consecutive.)*

Proof. Follows from the connectedness of C and the generalized star-shaped property. □

There are two parts to updating p's Voronoi region when adding a segment s to the diagram. First, we find a place on the boundary of $V(p)$ that is equidistant from p and S, if such a place exists. If so, we next walk around the boundary of $V(p)$, eliminating boundary sections until we reach the other place on the boundary where p is equidistant from S. (Note that either of these places could be "at infinity".)

The second part is easy - walk around the boundary of $V(p)$ from the starting point, eliminating obsolete bisector segments until you get to the finish point.

Because C is a polygonal chain, the first part is also easy. $V(p)$ is bounded by bisector pieces between p and a subset of the segments in C. Of the segments in this subset, there is a first segment F and a last segment L, according to the order along the chain.

Lemma 8 *If $V(p)$ changes, then its boundary with either $V(F)$ or $V(L)$ must change.*

Proof. The proof formalizes the idea if you can't go through the chain C, then the only way to get to $V(p)$ is through $V(F)$ or $V(L)$. □

Lemma 9 *We can compute the centers of all circular arcs that pass between g_i and g_j without crossing the tolerance boundary in $O(n)$ time.*

Proof. Incrementally add segments from C and amortize the update time. □

5 Computing the shortest path

To determine the shortest path from the start point to the end point of the polygonal curve we can build a directed graph of all possible valid arcs and do a BFS to find the shortest path from p_1 to p_n.

Theorem 10 *A point c is the center of a valid CW circular arc from p_i to p_j if and only if it is in the intersection of:*

- *the intersection of the right cones between p_i and each of the gates g_{i+1} through g_{j-1}.*
- *The region of $V(p_i)$ in the Voronoi diagram of p_i and all of the segments on the CW boundary between g_i and g_j.*
- *all points on b to the right of both s_i and s_j, where b, s_i, and s_j are as defined in Lemma 5.*

The conditions for valid CCW arcs are symmetrical.

Proof. Direct consequence of earlier lemmas. □

We find the possible arcs from a point p_i to all points further along P incrementally. We maintain the intersection of the right cones, the intersection of the left cones, the Voronoi region of p_i with the CW boundary, and the Voronoi region of p_i with the CCW boundary. At each step we update each of the four items. We intersect each bisector ray with an intersection of cones and with a Voronoi region, and then test if the two intersections overlap.

Note that we can quit if both of the cone intersection regions become empty. In fact, we could quit when the intersection of the right cones with the CW boundary Voronoi region and the intersection of the left cones with the CCW boundary Voronoi region are both empty, if we could test this quickly.

Theorem 11 *Given an open polygonal curve $P = (p_1, \ldots, p_n)$, a polygonal tolerance boundary of size $O(n)$, and a gate for each p_i, we can approximate P by a minimum number of valid circular arcs in $O(n^2 \log n)$ time.*

Proof. Each starting point takes $O(n \log n)$ time. □

6 Future Work

Because we compute all possible circular arcs from p_i to p_j, we expect to be able to use this information to match tangents of successive arcs or to compute bi-arcs. We have partial results along these lines.

References

[1] J. Eibel. Approximation of Planar Curves Within an Asymmetric Tolerance Band. *MSc.thesis, Universität Salzburg, Computerwissenschaften*, 2002.

[2] M. Held and J. Eibel. Biarc Approximation of Polygons Within Asymmetric Tolerance Bands. *Computer-Aided Design*, 37:357–371, 2005.

[3] DS. Meek and DJ. Walton. Approximation of discrete data by G^1 arc splines. *Computer-Aided Design*, 24:301–306, 1992.

[4] DS. Meek and DJ. Walton. Approximating quadratic NURBS curves by arc splines. *Computer-Aided Design*, 25:371–376, 1993.

[5] DS. Meek and DJ. Walton. Approximating smooth planar curves by arc splines. *Journal of Computational and Applied Mathematics*, 59:221–231, 1995.

[6] L. Piegl. Curve Fitting for Rough Cutting. *Computer-Aided Design*, 18:79–82, 1986.

[7] J. Schönherr. Smooth Biarc Curves. *Computer-Aided Design*, 25:365–370, 1993.

[8] M. Yeung and DJ. Walton. Curve Fitting With Arc Splines for NC Toolpath Generation. *Computer-Aided Design*, 26:845–849, 1994.