Computational Geometry: Theory and Applications 2 (1992) 31–50
Elsevier

Detecting and decomposing self-overlapping curves

Peter W. Shor and Christopher J. Van Wyk

AT&T Bell Laboratories, Murray Hill, NJ 07974, USA
Present address of C.J. Van Wyk: Department of Mathematics and Computer Science, Drew University, Madison, NJ 07940, USA

Communicated by János Pach
Submitted 7 February 1991
Accepted 12 September 1991

Abstract

Shor, P.W. and C.J. Van Wyk, Detecting and decomposing self-overlapping curves, Computational Geometry: Theory and Applications 2 (1992) 31–50.

A curve is self-overlapping if it can be divided by nontrivial line segments into simple curves. We show how to test whether a given curve is self-overlapping, and also how to construct sets of line segments that demonstrate that the curve is self-overlapping. We also describe several interesting topological properties of self-overlapping curves.

Keywords. Computational topology; constrained Delaunay triangulation; dynamic programming; immersions.

1. Introduction

All self-intersecting plane curves are nonsimple, but some are more nonsimple than others (cf. [11]). We say that a curve is *self-overlapping* if it can be divided by nontrivial line segments into simple curves. There are several other ways to define self-overlapping curves; all these definitions agree on curves with only a finite number of intersection points. Before we define the notion more formally, we offer some examples and motivation for the study of self-overlapping curves.

The curve in Fig. 1(a) is self-overlapping: a single slice at the bottom divides it into two simple curves. A simple curve is also, trivially, self-overlapping. Neither of the curves in Figs. 1(b) or 1(c), however, is self-overlapping: to divide either into simple pieces we would have to slice it at a crossing point.

Even when a self-overlapping curve is nonsimple, it has a natural 'interior.' Fig. 1(a), for example, naturally defines an annular region of the plane as its 'interior': we can choose an orientation on the curve so that if we trace the curve following the orientation, the interior always lies to our left. In contrast, neither of the

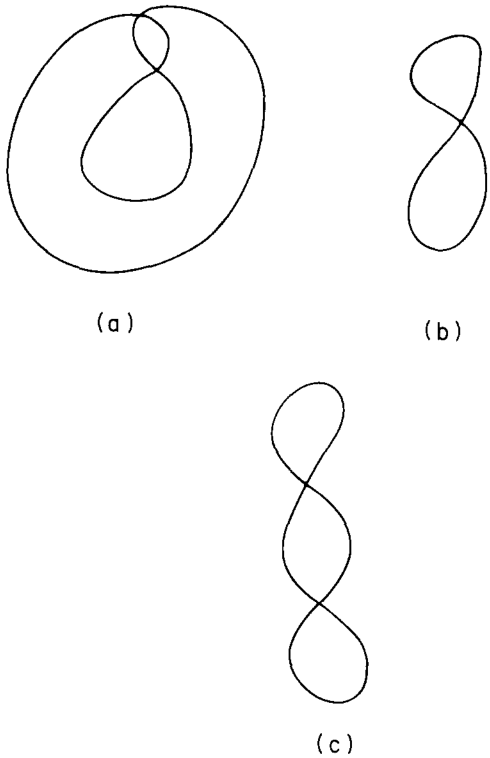


Fig. 1. The curve in (a) is a self-overlapping curve. Those in (b) and (c) are not.

curves in Figs. 1(b) and 1(c) has a natural interior: no matter what orientation we choose, at some point when we look to our left we shall be able to see infinity.

Micha Sharir posed the problem of identifying self-overlapping curves at the Eighth Geometry Day at the Courant Institute; he had in mind an application in robotics. Chi-Yuan Lo of AT&T Bell Laboratories' Integrated Circuit Design Aids Department suggests another application in the analysis of integrated-circuit layouts. If a layout contains a nonsimple but self-overlapping polygon, then it is probably correct; but if the layout includes a non-self-overlapping polygon, then the user should be advised of a likely error. His suggestion is motivated by the idea that self-overlapping curves have a natural interior, and is more conservative than some other recommendations for the treatment of nonsimple polygons in circuit layouts [10].

The identification of self-overlapping curves is not a simple matter of appealing to the celebrated Whitney–Graustein theorem [16], which characterizes curves that can be regularly deformed into a circle in terms of the net number of rotations a tangent vector makes as one traces the curve. It *is* necessary that a self-overlapping curve have unit tangent winding number, but Fig. 1(c) shows that this is not sufficient.

Another illustration of the difficulty of identifying self-overlapping curves is provided by *Milnor's paisley* curve, which is shown in Fig. 4. This curve can be realized as a self-overlapping region in two fundamentally different ways; that is, it can be divided into simple curves in two different ways, so that there is no way to map the pieces in the first decomposition onto the pieces in the second decomposition so that all paths in the first decomposition are mapped to paths in the second decomposition.

In this paper we give an algorithm for recognizing if a curve is self-overlapping. Our algorithm is based on 'triangulating' a self-overlapping polygon, which we do by using dynamic programming. We then give an algorithm for counting the number of fundamentally different ways of realizing a self-overlapping curve as a decomposition of polygons. This algorithm is based on an extension to self-overlapping polygons of the concept of constrained Delaunay triangulations [8]. We show that each fundamentally different way of realizing a self-overlapping curve corresponds to a unique constrained Delaunay triangulation, and give a dynamic programming algorithm for counting these constrained Delaunay triangulations.

Definition. A smooth curve $C = g(S^1)$ is *normal* if there is a finite number of self-intersections, each self-intersection x is a simple crossing point, and $g^{-1}(x)$ consists of two points. [16]

When C is normal, there is no ambiguity about where to proceed along C from each point of self-intersection: we can construct a function $g: S^1 \to C$ from C alone. Figs. 1, 3, and 4 depict normal curves, so we did not need to show g explicitly; Fig. 9, on the other hand, shows a curve that is not normal.

After we presented an earlier version of this work [14], we learned of a large body of closely related work; we mention here only the work of Titus [15], Blank [2], and Marx [9]. If we consider only the simple case where the curves are normal, then there are several equivalent definitions of self-overlapping curves. If $g: S^1 \to \mathbb{R}^2$ defines a normal self-overlapping curve, then it can be extended to an immersion $F: D^2 \to \mathbb{R}^2$, although some immersions of D^2 take the boundary to curves that are not normal. There is a converse to this statement; implicit in the work of Blank [2] and Marx [9] is the following: Let $F: D^2 \to \mathbb{R}^2$ be an immersion; if $C = F(S^1)$ is normal, then C is self-overlapping. Also, if $F: \mathbb{C} \to \mathbb{C}$ is analytic on an open region $D \subset \mathbb{C}$, and $\gamma: S^1 \to D$ is a homeomorphism, then $F \circ \gamma$ defines a self-overlapping curve; there are, however, self-overlapping curves that cannot be obtained this way. The case where the curves have an infinite number of crossing points, and thus are not normal, will be discussed later.

Given a normal immersion $g: S^1 \to \mathbb{R}^2$, Blank and Marx show how to determine when a given g is self-overlapping (in their words, "when it has a light open extension to D^2") and how many incompatible decompositions ("pairwise inequivalent properly interior extensions") it has. First they construct a word W

from C, then they apply transformations to the word to answer these questions. They do not mention any estimate of the time complexity of their transformations, but a dynamic programming approach can perform the transformations in $O(|W|^3)$ time. On a normal curve with k self-intersections, $|W| = O(k^2)$, so Blank's and Marx's methods can run in $O(k^6)$ time; we believe that one can define the words differently so that $|W| = O(k)$, which would give an $O(k^3)$ time algorithm. We present an algorithm that runs in $O(n^3 \log n)$ time for a polygon with n vertices. By transforming a curve with k self-intersections to a polygon with $O(k)$ vertices, we derive a $O(k^3 \log k)$-time algorithm for curves with k intersections. To compare this with Blank's and Marx's methods, note that an n-gon can have $k = \Omega(n^2)$, so Blank's and Marx's methods would run in time $O(n^{12})$ (using their definition of word) or $O(n^6)$ (using our modification of their definition). Titus [15] also describes conditions to detect when a given normal curve is self-overlapping ("is an interior boundary"); the running time of an algorithm based on his conditions is exponential.

2. Definitions

For any two points a and b in the plane, $\overline{ab}$ is the line segment between them. Let S^1 and D^2 denote the unit circle and the unit disk, respectively. For a set $S \subset \mathbb{R}^2$ that is homeomorphic to the unit circle S^1, $\tilde{S}$ denotes the union of S and its interior. For a set $S \subset \mathbb{R}^2$ that is homeomorphic to the unit disc D^2, ∂S is the bundary of S and int S is the interior of S. The domain of a mapping f is denoted dom f.

Let $g: S^1 \to \mathbb{R}^2$ be a continuous map of the unit circle into the plane. Then $C = g(S^1)$ is a closed curve in the plane. To simplify the notation, we shall write θ instead of $e^{i\theta}$. We use g to keep track of our place when we trace the curve C, since C may contain self-intersections.

Definition. A point $\sigma \in S^1$ is a *self-intersection point* if $g^{-1}(g(\sigma)) \neq \sigma$. A *self-intersection* is a maximal connected region of self-intersection points.

(For the moment we allow g to trace a piece of curve C more than once, which means that some self-intersections may be of positive measure. This allows our definitions to apply to curves that are not normal, a case that arises frequently in practice.)

We define what it means for curve C to be *self-overlapping* in terms of a construction that decomposes it into simple closed curves. A decomposition of C corresponds to a dissection of the unit circle by chords; indeed, we begin the construction with such a dissection. Then we make more choices and define more notation as the construction proceeds. At some points in the construction, we can tell that certain choices that are found to fail. It is possible, however, to finish the construction before we can tell that it has not worked. Thus, even though our definition involves a construction, ultimately it is purely existential: a curve is self-overlapping if and only if there exists a construction that demonstrates that it is; the definition does not offer many clues about how to find such a construction.

Let $((\phi_i, \psi_i) \mid 1 \leq i \leq m)$ be a sequence of non-empty open counterclockwise ranges of angles, and let $\delta_i = \overline{\phi_i \psi_i}$ be the chord of the unit circle determined by ϕ_i and ψ_i. The construction can succeed only if no two chords in the set $\{\delta_i \mid 1 \leq i \leq m\}$ intersect at a point in $\operatorname{int} D^2$. The sequence of angle ranges gives rise to two sequences of subregions of D^2. Let $\partial\Gamma_0 = S^1$; for $1 \leq i \leq m$, let

$$\partial\Delta_{i-1} = \partial\Gamma_{i-1} - (\psi_i, \phi_i)_{\Gamma_{i-1}} \cup \delta_i$$

and

$$\partial\Gamma_i = \partial\Gamma_{i-1} - (\phi_i, \psi_i)_{\Gamma_{i-1}} \cup \delta_i;$$

finally, let $\partial\Delta_m = \partial\Gamma_m$. (In this definition, the notation $(\phi, \psi)_\Gamma$ means the open interval along $\partial\Gamma$ going counterclockwise from ϕ and ψ.) These subregions are defined only if $\delta_i \subset \Gamma_{i-1}$ for $1 \leq i \leq m$. Thus, δ_i splits Γ_{i-1} into Δ_{i-1} and Γ_i. The non-intersection condition on the chords implies that Γ_i and Δ_i are homeomorphic to D^2 for all i. Fig. 2 shows a dissection of D^2 by a brief sequence of angle ranges.

The dissection of D^2 by the sequence of subregions $(\Delta_i \mid 0 \leq i \leq m)$ gives rise to a decomposition of C, which we describe by a sequence of mappings. Let $g_0 = g$; for $1 \leq i \leq m$, let $x_i = g(\phi_i)$ and $y_i = g(\psi_i)$, and let g_i map $\partial\Delta_{i-1}$ onto $g_{i-1}((\phi_i, \psi_i)_{\Gamma_{i-1}}) \cup \overline{x_i y_i}$ and $\partial\Gamma_i$ onto $g_{i-1}((\psi_i, \phi_i)_{\Gamma_{i-1}}) \cup \overline{x_i y_i}$ such that if $x \in \operatorname{dom} g_{i-1}$, then $g_i(x) = g_{i-1}(x)$. Thus, g_i always maps pieces of S^1 to pieces of C, and for $1 \leq j \leq i$ it maps the chord δ_j to the jth *diagonal* $\overline{x_j y_j}$. Fig. 2 illustrates the development of a short sequence of mappings. For $0 \leq i < m$, let $D_i = g_{i+1}(\partial\Delta_i)$; let $D_m = g_m(\partial\Delta_m)$. The ith diagonal chops the closed curve D_{i-1} off of C during the construction.

Definition. The above construction is *valid* if D_i is simple for $0 \leq i \leq m$.

When a construction is valid, the restriction of g_m to $\partial\Delta_i$ is a homeomorphism for all i. Only a valid construction can demonstrate that a curve is self-overlapping, but validity is not a sufficient condition. For example, if any of the diagonals $\overline{x_i y_i}$ is trivial $(x_i = y_i)$, then the construction may be valid but still not succeed.

Let $F: D^2 \to \mathbb{R}^2$ be a continuous extenstion of g_m: for $x \in \operatorname{dom} g_m$, $F(x) = g_m(x)$, and for $0 \leq i \leq m$, F restricted to Δ_i is a homeomorphism between Δ_i and $\bar{D}_i$.

Let us pause for a moment to review where in this construction we have made choices. The map $g: S^1 \to \mathbb{R}^2$ was given. A decomposition of C is defined by a sequence of angle ranges $((\phi_i, \psi_i) \mid 1 \leq i \leq m)$, which defines a sequence of

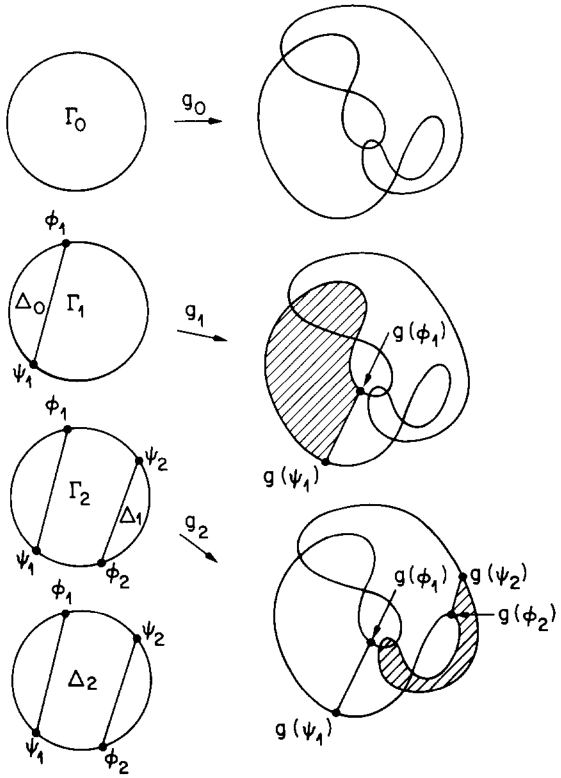


Fig. 2. The left column shows the dissection of D^2 by the sequence $((\phi_1, \psi_1), (\phi_2, \psi_2))$. The right column shows the image of D^2 together with appropriate chords as the sequence of mappings g_0, g_1, and g_2 is produced. The shaded regions indicate int D_0 and int D_1.

chords, (δ_i), and two sequences of subregions, (Γ_i) and (Δ_i), of D^2. We constructed a sequence of mappings that culminates in g_m, which defined a sequence of diagonals $(\overline{x_i y_i})$ and a sequence of closed curves (D_i). We chose F as a continuous extension of g_m to D^2. By construction, the restriction of F to S^1 agrees with g, and the restriction of F to $S^1 \cup \{\delta_i \mid 1 \leq i \leq k\}$ agrees with g_k; thus, we can recover the complete status of the construction from the pair (Θ, F), where $\Theta = ((\phi_i, \psi_i) \mid 1 \leq i \leq m)$.

Definition. Let $g: S^1 \to \mathbb{R}^2$ be given. If there exist Θ and F such that (Θ, F) is a valid construction and for each point $x \in D^2$ there is a neighborhood of x in which F is injective, then g defines a *self-overlapping* curve, and the construction (Θ, F) *demonstrates* this fact.

Once we have the dissection Θ and the function F, we can refer to the interior of C as the image of int D^2 under F. Function F is, in fact, *nearly* an immersion of D^2 [7].

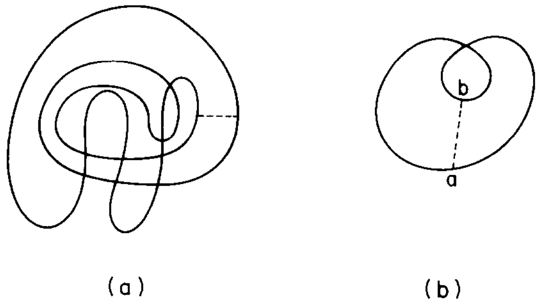


Fig. 3. The diagonal in (a) shows that the curve is self-overlapping. The diagonal in (b), however, does not yield a proper decomposition of the non-self-overlapping curve shown.

Fig. 3(a) shows a self-overlapping curve and a diagonal that suggests how it can be decomposed. Fig. 3(b) shows a curve that is not self-overlapping, and illustrates why we require that F be locally injective. Although diagonal $\overline{ab}$ separates C into two simple closed curves, so that we can define a valid F, such an F will not be injective in any neighborhood of b: the interiors of the two closed curves overlap in any neighborhood of b. (Here is another observation that suggests that Fig. 3(b) does not contain a self-overlapping curve: in an attempt to define the natural interior, we could orient and trace the curve counterclockwise; when we look to our left at a, however, we shall see the *exterior* of C at b.)

3. Compatible decompositions

For the self-overlapping curves in Figs. 1(a), 2, and 3(a), all decompositions are essentially the same; informally, there is essentially only one way to chop the curves into simple pieces. Fig. 4 shows Milnor's paisley curve [12], which illustrates that in general there can be different ways to chop a self-overlapping curve into simple pieces; the diagonals in decomposition I are $\overline{af}$ and $\overline{be}$, while in decomposition II they are $\overline{ch}$ and $\overline{dg}$. The possibility of essentially different decompositions motivates our definition of what it means for two decompositions to be *compatible*.

Definition. Let (Θ, F) and (Θ', F') define two decompositions. The two decompositions are *compatible* if whenever $(\phi', \psi') \in \Theta'$, $F^{-1}(\overline{g(\phi')g(\psi')})$ includes a path that connects ϕ' to ψ' and otherwise lies in int D^2, in which case we say that ϕ' and ψ' are *mutually visible* under (Θ, F).

We note without proof that compatibility between decompositions of g is an equivalence relation that depends only on Θ and Θ', not on the choice of F and F'; thus we refer below to mutual visibility under Θ alone. We also note that the visibility relation is symmetric and depends only on Θ. Fig. 5 shows the inverse

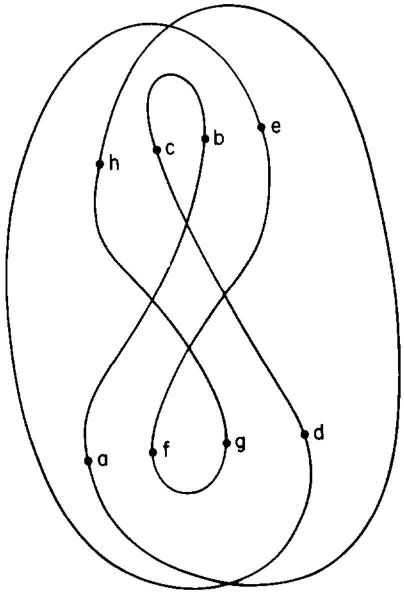


Fig. 4. This figure has two incompatible decompositions.

image of $\overline{dg}$ from decomposition II of Fig. 4 under the F of decomposition I; the dashed path at g ends at a preimage of d and vice versa; since neither path connects two points in S^1, decompositions I and II are incompatible.

4. An algorithm for polygons

In this section we present an algorithm to discover whether a polygon is self-overlapping. This is a key step in the algorithm to solve the problem on general curves.

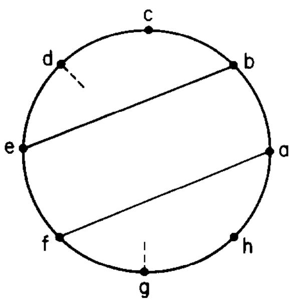


Fig. 5. The points in S^1 that are preimages under g of the labelled points in Fig. 4 are labelled with the corresponding letters.

Definition. A curve $C = g(S^1)$ is a *polygon* if C consists of n line segments or *sides*.

When C is a polygon, there is an increasing sequence of angles $(\theta_i \mid 0 \leq i < n)$ such that $g(\theta_i)$ is an endpoint of one of the sides of C; the image of these angles under g is the set of vertices of C. When C is a polygon on n vertices and $n < \infty$, each point of C can have only a finite number of preimages under g. Moreover, there can be only a finite number of self-intersections, and the image of each is either a point or a line segment.

Definition. A decomposition construction (Θ, F) is a *triangulation* of polygon C if it demonstrates that C is self-overlapping, for $(\phi, \psi) \in \Theta$, $\phi = \theta_j$ and $\psi = \theta_k$ for some j and k, and for (ϕ_i, ψ_i), $(\phi_j, \psi_j) \in \Theta$, $i \neq j$, the set $\{\phi_i, \psi_i, \phi_j, \psi_j\}$ contains at least three distinct values.

If there is a triangulation of the polygon then it is certainly self-overlapping. Conversely, Theorem 1 below shows that if (Θ, F) demonstrates that polygon C is self-overlapping then there is a triangulation compatible with (Θ, F). This means that an algorithm to determine whether a polygon is self-overlapping can work by seeking a triangulation of the polygon.

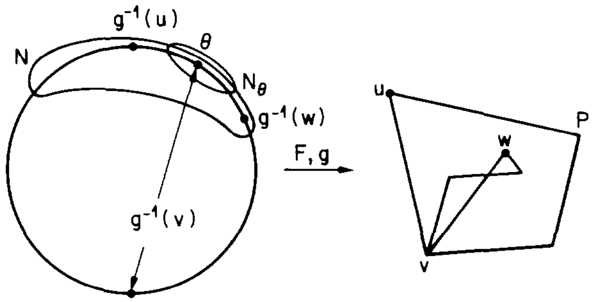
Lemma 1. *Let (Θ, F) demonstrate that g defines a self-overlapping polygon P. Let θ be such that $v = g(\theta)$ is a convex vertex of P. There exists an open neighborhood $N_\theta \subset S^1$ of θ such that whenever ϕ and ψ lie on opposite sides of θ in N_θ, ϕ and ψ are mutually visible under Θ.*

Proof. Let $N \subset D^2$ be an open neighborhood of θ on which F is injective. The neighborhood $F(N)$ must include a non-empty triangle T two of whose sides coincide with sides of P that are incident to v. Let $N_\theta \subset S^1$ be an open neighborhood of θ such that $g(N_\theta) \subset T$. (See Fig. 6.)

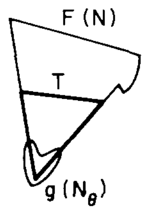
Let $\phi, \psi \in N_\theta$ be two points that lie on opposite sides of θ in N_θ. By construction, $\overline{g(\phi)g(\psi)}$ lies in T. Since F is injective on N and $T \subset N$, $F^{-1}(\overline{g(\phi)g(\psi)})$ includes a path between ϕ and ψ that otherwise lies in the interior of N. Thus, ϕ and ψ are mutually visible under Θ. □

Lemma 2. *Let (Θ, F) demonstrate that g defines a self-overlapping polygon P. There exist two points $\theta_1, \theta_2 \in S^1$ such that $g(\theta_1)$ and $g(\theta_2)$ are vertices of P and θ_1 and θ_2 are mutually visible under Θ.*

Proof. Let θ be such that $v = g(\theta)$ is one of the vertices of P with minimum y-coordinate. Since P is self-overlapping, the region $F(D^2)$ can be understood as the 'inside' of P. Since v has minimum y-coordinate, all of the inside of P lies above v, so v is a convex vertex. Let ϕ and ψ be the preimages under g of the



(a)



(b)

Fig. 6. Fig. (a) depicts most of the notation in the proof of Lemma 1. Fig. (b) shows the images of the neighborhoods N and N_θ under F and g, respectively, as well as triangle T.

vertices that precede and follow v on P. If ϕ and ψ are mutually visible under Θ, then we are done.

Otherwise, let $N_\theta = (\phi', \psi')$ be a maximal open neighborhood of θ such that ϕ, ϕ', θ, ψ', and ψ appear in that cyclic order on S^1, and for any ζ and η on opposite sides of θ in N_θ, ζ and η are mutually visible under Θ, as constructed in

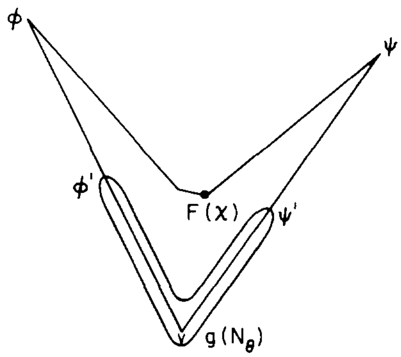


Fig. 7. This illustration for the proof of Theorem 1 uses a simple polygon. Thus, F and g are homeomorphisms, and we need only draw the situation in the plane that contains P. The figure shows the image of one possible choice of maximal open neighborhood N_θ.

the proof of Lemma 1 and illustrated in Fig. 7. Let $l=\overline{g(\phi')g(\psi')}$. By construction, $F^{-1}(l)$ includes no path that joins ϕ' to ψ' and otherwise lies in int D^2. By continuity, however, $F^{-1}(l)$ includes a path Π between ϕ' and ψ' that lies in the inverse image under F of the triangle $\Delta g(\theta)g(\phi')g(\psi')$. Therefore, the path Π intersects S^1 in more than two points. Let χ be a point on the boundary of the intersection of the interior of Π with S^1. Then $F(\chi)$, like $g(\theta)$, is a vertex of P, and θ and χ are mutually visible under Θ. □

Theorem 1. *Given a decomposition* (Θ, F) *that demonstrates that polygon* P *is self-overlapping, there exists a compatible triangulation of* P.

Proof. Use Lemma 2 to find points θ_1 and θ_2 that are mutually visible under Θ such that $g(\theta_1)$ and $g(\theta_2)$ are vertices of P. Then $F^{-1}(\overline{g(\theta_1)g(\theta_2)})$ includes a path Π that joins θ_1 to θ_2 and divides D^2 into two simply connected regions Ω_1 and Ω_2. We can modify Θ so it demonstrates that $F(\partial\Omega_1)$ and $F(\partial\Omega_2)$ are both self-overlapping polygons, as follows. We build the two sequences by considering in order the chords defined by Θ. For each, if $(\phi, \psi) \in \Omega$ has both endpoints in $\partial\Omega_1$ or $\partial\Omega_2$, assign it to the modified sequence for the appropriate region. Otherwise, the chord $\overline{\phi\psi}$ intersects Π in a single point (since $F(\overline{\phi\psi})$ and $F(\Pi)$ are both line segments); assign $\overline{\phi\psi} \cap \Omega_1$ to the modified sequence for Ω_1 and $\overline{\phi\psi} \cap \Omega_2$ to the modified sequence for Ω_2.

This proves that a self-overlapping polygon on n vertices can always be decomposed into two self-overlapping polygons, each of which has no more than $n-1$ vertices. A simple induction suffices to complete the proof of the theorem. □

Next we use Theorem 1 to state an algorithm that tests whether a given polygon is self-overlapping.

Algorithm 1. Let $v_0, \ldots, v_{n-1}$ be the vertex sequence of polygon P. For convenience, assume that no three consecutive vertices form a straight angle.

The algorithm uses dynamic programming to find a triangulation of P; it constructs a table $Q_{n\times n}$ where Q_{ij} is one if it is possible to triangulate the $(j-i+1)$-gon whose vertices are $v_i, \ldots, v_j$, and zero otherwise. (All arithmetic on subscripts is carried out modulo n, so if $j<i$, the vertex sequence wraps around from v_{n-1} to v_0, and we treat $j-i+1$ as $j-i+n+1$.) For convenience we set $Q_{i,i+1}=1$.

The first step of the dynamic program is to fill in the values $Q_{i,i+2}$; if v_{i+1} is a convex vertex, then $Q_{i,i+2}=1$, but if v_{i+1} is a reflex vertex then $Q_{i,i+2}=0$. The numbering of the vertices defines an orientation on the polygon, and all triangles for which $Q_{i,i+2}=1$ will be oriented the same way; without loss of generality we assume that this orientation is counterclockwise as one travels from v_i through v_{i+2} to v_{i+1}.

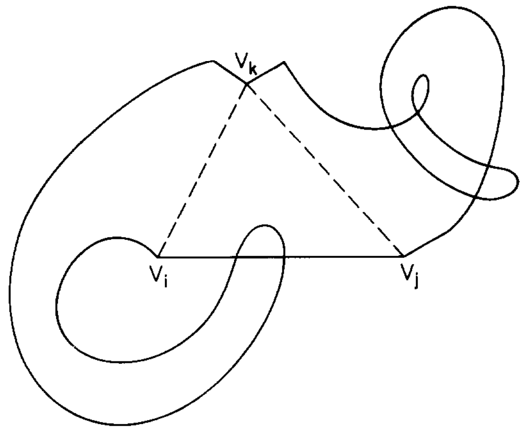


Fig. 8. General step of the dynamic program.

In general, the value of Q_{ij} is one if and only if there exists an index k such that $Q_{ik} = Q_{kj} = 1$, $\Delta v_i v_j v_k$ is oriented counterclockwise, v_i, v_j, v_{k+1} and v_{k-1} appear in that order counterclockwise around v_k, and the following four segments do not intersect the interior of $\Delta v_i v_j v_k$: $v_i v_{i+1}$, $v_{k-1} v_k$, $v_k v_{k+1}$, and $v_{j-1} v_j$ (see Fig. 8). A simple induction shows that when these conditions hold, v_i, v_j, and v_k form a triangle along whose sides we can glue triangulations of v_i through v_k and v_k through v_j so that we can construct an F that is locally injective around v_i, v_j, and v_k.

Since we can compute each element Q_{ij} in $O(n)$ time, Algorithm 1 runs in $O(n^3)$ time. The polygon is self-overlapping if and only if there is an index i such that $Q_{i,i-1} = 1$. Therefore we can test in time cubic in the number of vertices whether a polygon is self-overlapping.

To make it possible to reconstruct a triangulation of P from the dynamic program, Algorithm 1 can record at each location Q_{ij} that is set to one a value of k that permitted us to set $Q_{ij} = 1$. From these values it is possible to reconstruct a sequence Θ that demonstrates that P is self-overlapping.

5. Generalization to curves

In this section we show how to use Algorithm 1 to test whether a curve $C = g(S^1)$ is self-overlapping. We shall assume that the set of self-intersections is finite.

Algorithm 2. First we construct a graph G whose embedding in the plane is the same as that of curve C. Let $\{P_i\}$ be the set of pieces of C that can be expressed

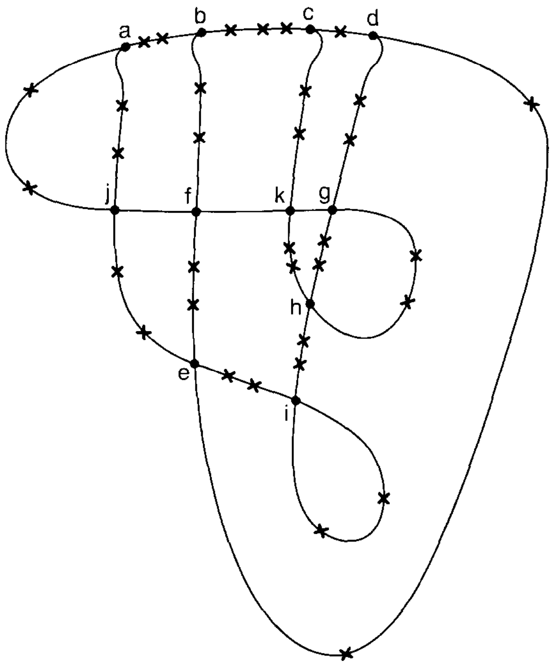


Fig. 9. The transformation of a curve to a planar graph. The 'Gauss code' is *abcdefbcdghiiejabckhgkfj*. Bullets mark vertices from W, while crosses mark vertices added to keep multiple edges and self-loops from collapsing.

as $g(\Sigma_1) \cap g(\Sigma_2)$ for some self-intersections Σ_1 and Σ_2, and let W be the set of points that lie on the boundary of some P_i. If P_i is a point, it will give rise to a single vertex in W; if P_i is of positive measure, it will give rise to two vertices in W, one at each endpoint. The set $\Phi = g^{-1}(W)$ is a set of points on S^1. Construct Φ' by adding to Φ two distinct points that lie between each pair of neighboring points in Φ. Then $V = g(\Phi')$ is the vertex set of G.

The edge set of G is defined by C. If vertices v_1 and v_2 belong to V, there is an edge between them if and only if they are connected by a piece of C that contains no other vertices in V. Thus, G is a planar graph (see Fig. 9).

Use any of several algorithms ([5, 6, 13]) to modify the embedding of G so that each edge is replaced by O(1) straight line segments. Let H be a cycle of G that corresponds to traversing g around S^1. If there are no self-intersections of positive measure, then H is a Hamiltonian cycle of G; otherwise, it will traverse some edges more than once. In either case, H is a polygon, on which we can use Algorithm 1 to test whether it is self-overlapping.

If curve C has k self-intersections, then graph G will have $O(k^2)$ vertices and edges, and Algorithm 1 will run in $O(k^6)$ time, since the time to construct the dynamic program dominates the time to construct the straight-line embedding of C. If all of the self-intersections of C are points, the bound on the running time can be improved: G will have $O(k)$ vertices and edges, so this algorithm will run in $O(k^3)$ time.

6. Counting incompatible decompositions

Let $P = g(S^1)$ be a polygon with n vertices, and let (Θ, F) demonstrate that P is self-overlapping. If the sequence of angle pairs Θ defines a triangulation of P, then for $0 \leq i \leq n-3$, the image $F(\partial\Delta_i)$ is a triangle T_i whose vertices are vertices of P, and $\partial\Delta_i$ contains three values α_i, β_i, γ_i which are preimages of the vertices of T_i.

Definition The triangulations defined by two sequences of angle ranges, Θ and Θ', are *combinatorially equivalent* if they define the same set of chords in D^2.

Combinatorially equivalent triangulations are produced by different orderings of the same set of diagonals. (Notice, however, that in general a sequence of angle ranges cannot be reordered arbitrarily, since each diagonal is required to cut off a simple curve.)

It is straightforward to modify Algorithm 1 to count the number of combinatorially equivalent ways there are to triangulate P. Instead of setting Q_{ij} to be zero or one, we store in Q_{ij} the number of combinatorially different trangulations of $v_i, \ldots, v_j$. Since a convex n-gon has exponentially many compatible but combinatorially different triangulations, however, this count does not tell how many incompatible decompositions there are. Now we shall define a class of triangulations that have special properties that allow us to count the incompatible decompositions of a polygon.

Definition. A triangulation (Θ, F) of P is a *constrained Delaunay triangulation* (CDT) with respect to a decomposition if for each $0 \leq i \leq n-3$, there is no value θ such that $g(\theta)$ is a vertex of P, $g(\theta)$ lies inside the circumcircle of T_i, and θ is visible under Θ to all of α_i, β_i, and γ_I.

This definition is essentially the same as for simple polygons [8], except the notion of visibility is defined with respect to a decomposition.

Definition. A triangulation (Θ, F) of P is *locally optimal* if the following is true for every two regions Δ_i and Δ_j that share a chord on their boundaries: Without loss of generality, label the preimages of the vertices so that $\alpha_i = \alpha_j$ and $\gamma_i = \gamma_j$; then β_i does not lie inside the circumcircle of T_j and β_j does not lie inside the circumcircle of T_I.

Following [8], we note that if Δ_i and Δ_j share a chord and do not have this property, then $\Delta\beta_i\beta_j\alpha_i$ and $\Delta\beta_i\beta_j\gamma_i$ do have this property. From the definiton and this observation, it is clear that a locally optimal triangulation of a polygon P always exists, and that a constrained Delaunay triangulation is locally optimal. The next theorem shows that a locally optimal triangulation is a constrained Delaunay triangulation, which proves that constrained Delaunay triangulations exist, and also that we can compute them relatively easily.

Theorem 2. *Suppose that $P = g(S^1)$ has no four cocircular vertices and Θ defines a locally optimal triangulation of P. Then Θ defines a CDT of P.*

Proof. The proof is a modification of the proof for simple polygons [8]. We proceed by contradiction. Let α, β, γ, and ϕ be such that $g(\phi)$ lies inside the circumcircle of $\Delta g(\alpha)g(\beta)g(\gamma)$, ϕ is visible to α, β, and γ, and the distance from $g(\phi)$ to $\overline{g(\alpha)g(\gamma)}$ is the smallest over all pairs of triangles and visible vertices that lie in their circumcircles in the triangulation defined by Θ. If the perpendicular projection of $g(\phi)$ onto the line containing $\overline{g(\alpha)g(\gamma)}$ does not hit $\overline{g(\alpha)g(\gamma)}$, there might be a choice of triangles; in this case, choose the values of α, β, and γ so that the perpendicular projection of ϕ is closest to the segment $\overline{g(\alpha)g(\gamma)}$. (See Fig. 10.)

Since ϕ is visible to β, $\overline{g(\alpha)g(\gamma)}$ must be a diagonal in the triangulation, not an edge of P. Thus it belongs to another triangle whose third vertex is $g(\delta)$. Since Θ defines a locally optimal triangulation, $g(\delta)$ lies outside the circumcircle of $\Delta g(\alpha)g(\beta)g(\gamma)$. Moreover, since $g(\phi)$ cannot lie inside $\Delta g(\alpha)g(\gamma)g(\delta)$, one of the edges of $\Delta g(\alpha)g(\gamma)g(\delta)$, say $\overline{g(\gamma)g(\delta)}$, lies between $g(\phi)$ and $\overline{g(\alpha)g(\gamma)}$.

Since ϕ is visible to β, $\overline{g(\gamma)g(\delta)}$ is also a diagonal in the triangulation. If ϕ and δ were not mutually visible under Θ, then there would be another triangle and visible point where the point was closer to the triangle than $g(\phi)$ is to $\Delta g(\alpha)g(\beta)g(\gamma)$; this would violate the choice of α, β, γ, and ϕ. Not only is ϕ visible to δ: $g(\phi)$ is also closer to $\overline{g(\gamma)g(\delta)}$ than to $\overline{g(\alpha)g(\gamma)}$. Finally, since $g(\delta)$ lies outside the circumcircle of $\Delta g(\alpha)g(\gamma)g(\phi)$, and $\overline{g(\alpha)g(\gamma)}$ is a chord of the circumcircle of $\Delta g(\alpha)g(\gamma)g(\delta)$, $g(\phi)$ must lie inside this circumcircle, which contradicts the initial choice of α, β, γ, and ϕ. This proves that no such α, β, γ, and ϕ exist, so Θ defines a constrained Delaunay triangulation. □

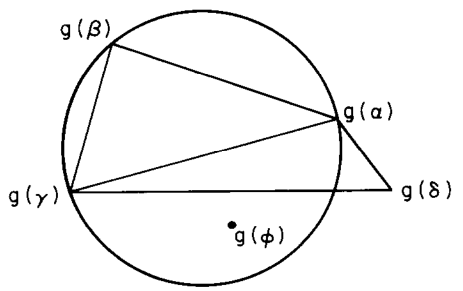


Fig. 10. Illustration for the proof of Theorem 2.

If no four of the vertices of a simple polygon are cocircular, then the constrained Delaunay triangulation is unique [8]. We shall prove a stronger result.

Theorem 3. *Suppose that $P = g(S^1)$ has no four cocircular vertices. Two decompositions Θ and Θ' of P have combinatorially equivalent constrained Delaunay triangulations if and only if they are compatible.*

Proof. The proof in one direction is easy. If the CDTs with respect to Θ and Θ' are combinatorially equivalent, then each diagonal in one decomposition is a diagonal in the other. Thus the two decompositions are compatible by definition.

Suppose that the CDTs defined by Θ and Θ' are combinatorially different. We must show that the decompositions are incompatible. Let ϕ be such that $\overline{g(\phi)g(\psi)}$ is the first diagonal (beginning counterclockwise from an edge of P incident to $g(\phi)$) that radiates from $g(\phi)$ and is not shared by the CDTs defined by Θ and Θ'. Assume without loss of generality that $\overline{g(\phi)g(\psi)}$ belongs to the CDT defined by Θ. Let $\overline{g(\phi)g(\theta_1)}$ be the diagonal that immediately precedes $\overline{g(\phi)g(\psi)}$; this diagonal belongs to both CDTs. (If $\overline{g(\phi)g(\psi)}$ is the first diagonal in Θ that follows an edge incident to $g(\phi)$, then $\overline{g(\phi)g(\theta_1)}$ is a side of P.) Let $\overline{g(\phi)g(\theta_2)}$ be the diagonal that immediately follows $\overline{g(\phi)g(\psi)}$ in the CDT defined by Θ. (See Fig. 11.)

If ϕ and ψ are not mutually visible under Θ', then there is nothing to prove: since they are mutually visible under Θ, the CDTs defined by Θ and Θ' are incompatible. So assume that ϕ and ψ are mutually visible under Θ'. We seek the preimage ψ' of the vertex of the triangle in the CDT defined by Θ' that has $\overline{g(\phi)g(\theta_1)}$ as an edge and that lies on the same side of that edge as $g(\psi)$. Note that by the choice of ϕ and ψ, $g(\psi')$ lies on the same side of $\overline{g(\phi)g(\psi)}$ as $g(\theta_2)$.

First we exclude two possible locations of $g(\psi')$. If $g(\psi')$ lay inside $\Delta g(\phi)g(\psi)g(\theta_2)$, then it would be visible under Θ from all three of ϕ, ψ, and θ_2 (since ϕ, θ_2, and ψ are pairwise mutually visible under Θ), which would

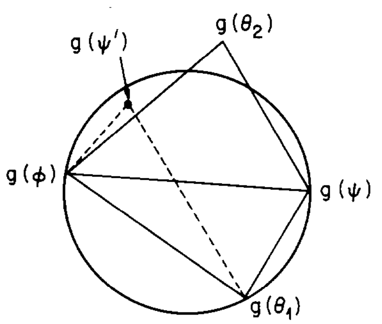


Fig. 11. The notation for the proof of Theorem 3. The solid triangles belong to the CDT defined by Θ. The dashed triangle shows where ψ' might lie.

contradict the fact that Θ defines a CDT. If $g(\psi')$ lay outside the circle through $g(\phi)$, $g(\psi)$, and $g(\theta_1)$, then the circumcircle of $\Delta g(\phi)g(\psi')g(\theta_1)$ would contain $g(\psi)$; since both ϕ and ψ' and ϕ and θ_1 are mutually visible under Θ', and ϕ and ψ are mutually visible under Θ' by assumption, ψ would be visible under Θ' to all three of ϕ, ψ' and θ_1, which would contradict the fact that Θ' defines a CDT.

Now suppose that $g(\psi')$ lay inside the circle segment bounded away from $g(\psi)$ by $\overline{g(\phi)g(\theta_2)}$. Then since Θ defines a CDT, ψ' must not be visible to all three of ϕ, ψ, and θ_1 under Θ. Since ϕ and ψ are mutually visible under Θ' by assumption, however, there must be an edge of P that intersects the circumcircle of $\Delta g(\phi)g(\psi)g(\theta_1)$ in a chord that lies between $\overline{g(\phi)g(\theta_2)}$ and $g(\psi')$. Therefore, ϕ and ψ' are not mutually visible under Θ, but are mutually visible under Θ', so Θ and Θ' define incompatible decompositions.

If $g(\psi')$ lay inside the circle segment bounded away from $g(\phi)$ by $\overline{g(\psi)g(\theta_2)}$, reasoning similar to that preceding shows that the decompositions defined by Θ and Θ' are incompatible.

This reasoning covers all possible locations for $g(\psi')$ except on the circumcircle of $\Delta g(\phi)g(\psi)g(\theta_1)$. Such a placement, however, would violate the assumption that no four vertices of P are cocircular. □

Theorem 3 implies that when P has no four cocircular vertices, we can count the number of incompatible decompositions it has by finding the number of combinatorially inequivalent constrained Delaunay triangulations. Algorithm 3 is a modification of Algorithm 1 that does this. It fills a table $Q_{n\times n\times n}$ by setting Q_{ijk} to be the number of combinatorially different locally optimal triangulations of the $(j-i+1)$-gon whose vertices are $v_i, \ldots, v_j$ that include $\Delta v_i v_j v_k$. Obviously, $Q_{ijk}=0$ when k does not follow i and precede j in cyclic order.

Algorithm 3. The first step of the dynamic program sets $Q_{i,i+2,i+1}=1$ if and only if v_{i+1} is a convex vertex. The general step of the dynamic program sets Q_{ijk} to $(\Sigma_a Q_{ika}) \times (\Sigma_b Q_{kjb})$ where the summation indices a and b are such that $Q_{ika}>0$, $Q_{kjb}>0$, $\Delta v_i v_j v_k$ is oriented counterclockwise and obeys the local optimality property with respect to both triangles $\Delta v_i v_a v_k$ and $\Delta v_k v_b v_j$, the vertices v_j, v_j, v_{k+1} and v_{k-1} appear in that order counterclockwise around v_k, and the following four segments do not intersect the interior of $\Delta v_i v_j v_k$: $v_i v_{i+1}$, $v_{k-1}v_k$, $v_k v_{k+1}$, and $v_{j-1}v_j$. Thus, v_i, v_j, and v_k form a triangle along whose sides we can glue locally optimal triangulations of v_i through v_k and v_k through v_j to derive a locally optimal triangulation of v_i through v_j.

Since the range of values of a and b that must be considered to compute Q_{ijk} do not overlap, it is easy to compute Q_{ijk} in $O(n)$ time, which leads to a running time of $O(n^4)$ for Algorithm 3. If instead of a simple three-dimensional table we maintain a matrix $Q_{n\times n}$ of sorted sequences, where Q_{ij} contains the partial sums of the values of Q_{ijk}, sorted by increasing angle at v_k in $\Delta v_i v_j v_k$, then we can reduce this running time to $O(n^3 \log n)$.

If a polygon contains four cocircular vertices that are mutually visible under some decomposition, then it has combinatorially different Delaunay triangulations that are compatible. Since Algorithm 3 counts combinatorially different Delaunay triangulations, it will not count correctly the number of incompatible decompositions of the polygon. To prevent this, modify the dynamic program so that if T_{ijk_1} and T_{ijk_2} are to be set to the same value because v_i, v_j, v_{k_1}, and v_{k_2} are cocircular and mutually visible, then only $T_{i,j,\min\{k_1,k_2\}}$ is set to this value, while $T_{i,j,\max\{k_1,k_2\}}$ is set to zero.

If we use Algorithm 3 instead of Algorithm 1 as a step in Algorithm 2, then we can count the number of incompatible decompositions of any curve.

7. Infinite self-intersection

Curves with an infinite number of self-intersections pose some interesting challenges. Some, like the one shown in Fig. 12, are self-overlapping according to the definition in Section 2, since they can be decomposed into a finite number of simple pieces. It is not clear, however, how to apply Algorithm 2 in any reasonable way.

The curve shown in Fig. 13 can be decomposed into a infinite number of simple pieces, but not into any finite number. To account for this example by some sort of explicit construction, we might take $m = \infty$ in our definition, and require that all Δ_i, including 'Δ_∞,' be simple. This change, however, causes us to lose the immersion property of Section 1.

8. Open problems

A variety of questions remain to be answered about self-overlapping curves.

Let C be a self-overlapping plane curve and let R be an open region in $\mathbb{R}^2 - C$. Under any mapping F defined by a decomposition of C, every point in R has the

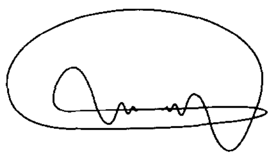


Fig. 12. One piece of this self-overlapping curve is a portion of the curve $y = x^2 \sin 1/x$, and another piece is a portion of the x-axis.

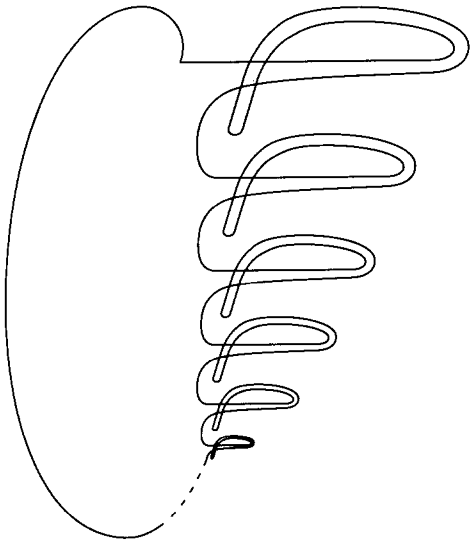


Fig. 13. Should this curve be self-overlapping?

same number of preimages under F (this follows from theorems about winding numbers [1]); thus we can speak of the numbers of *layers* that cover R, independent of the mapping F. We say that C is a k-*layer* curve if k is the maximum value such that no region in $\mathbb{R}^2 - C$ is covered by more than k layers; for example, the curve in Fig. 4 is a three-layer curve. We know of no two-layer curve that has two incompatible decompositions, and conjecture that none exists.

From a computational standpoint, an obvious question is whether one can test whether a curve is self-overlapping in sub-cubic time. Leo Guibas asked how one can find a decomposition that demonstrates that a curve is self-overlapping and uses a minimum number of cuts.

In our earlier paper [14], we made some incorrect remarks and conjectures. We suggested that one way to think of self-overlapping curves is as the projection of the boundaries of stretched but untwisted disks in three dimensions. A generalization of Milnor's paisley shows that this property is sufficient to guarantee that a curve is self-overlapping, but not necessary [3], and also disproves our conjecture about the indivisibility of the number of incompatible decompositions by any prime larger than the number of layers.

Acknowledgements

We are grateful for several discussions with Ken Clarkson, Subhash Suri, Gabor Tardos, Bob Tarjan, and Bill Thurston. Steve Fortune, David Johnson, David Levine, and Alex Schaffer offered helpful comments on the manuscript.

References

[1] L.V. Ahlfors, Complex Analysis, 2d Ed. (McGraw-Hill, New York, 1966) 114–118.
[2] S.J. Blank, Extending immersions of the circle, Ph.D. Dissertation, Brandeis University, 1967.
[3] D. Bennequin, Exemples d'immersions du disque dans le plan qui ne sont pas projections de plongements dans l'espace, C. R. Acad. Sci. Paris Ser. A 281 (1975) 81–84.
[4] G.K. Francis, The folded ribbon theorem. A contribution to the study of immersed circles, Trans. Amer. Math. Soc. 141 (1969) 271–303.
[5] H. de Fraysseix, J. Pach, and R. Pollack, Small sets supporting Fary embeddings of planar graphs, Proc. 20th Ann. Symp. Th. Comput. (1988) 426–433.
[6] D.H. Greene, Efficient coding and drawing of planar graphs, 1983, unpublished manuscript.
[7] J. Jewett, Differentiable approximations to light interior transformations, Duke Math. J. 23 (1956) 111–124.
[8] D.T. Lee and A.K. Lin, Generalized Delaunay triangulation for planar graphs, Discrete Comput. Geom. 1 (1986) 201–217.
[9] M.L. Marx, Extensions of normal immersions of S^1 into $\mathbb{R}^2$, Trans. Amer. Math. Soc. 187 (1974) 309–326.
[10] M.E. Newell and C.H. Sequin, The inside story on self-intersecting polygons, Lambda 1 (Second Quarter) (1980) 20–21.
[11] G. Orwell, Animal Farm (1945).
[12] V. Poenaru, Extensions des immersions en codimension 1 (d'après Blank), Séminaire Bourbaki: 1967/68, Exposé 342 (Benjamin, New York, 1969).
[13] P. Rosenstiehl and R.E. Tarjan, Rectilinear planar layouts and bipolar orientations of planar graphs, Discrete Comput. Geom. 1 (1986) 343–353.
[14] P. Shor and C.J. Van Wyk, Detecting and decomposing self-overlapping curves, Proc. 5th Ann. Symp. Comput. Geom (1989) 44–50.
[15] C.J. Titus, The combinatorial topology of analytic functions on the boundary of a disk, Acta Math. 106 (1961) 45–64.
[16] H. Whitney, On regular closed curves in the plane, Compositio Math. 4 (1937) 276–284.