The Sausage of Local Convex Hulls of a Curve and the Douglas-Peucker Algorithm

FRANÇOIS NORMANT AND AXEL VAN DE WALLE

Département de Mathématiques et de Génie industriel
École Polytechnique de Montréal

Abstract We investigate the problem of the simplification of a curve from a geometrical and fractal viewpoint. We consider that the best methods are those that keep an estimated value of the fractal dimension constant, before and after simplification, and that consider local details during processing.

We propose here both a simplification method and a simplification criterion, based on a new algorithm for computing the fractal dimension introduced by Tricot (1994). We cover the curve with overlapping convex sets of constant breadth (the ε-sausage of convex structuring elements) and retain the points of the original curve that lie on the boundary of this covering set. This algorithm is both accurate and universal, and does not imply the need to make any arbitrary hypotheses regarding the structure of the curve.

We will show experimentally that the well-known Douglas-Peucker method shares some very interesting geometrical properties with this algorithm and thus has the ability to preserve the fractal dimension when evaluated with a lower cutoff equal to the threshold distance.

Introduction

THE simplification of lines involves reducing the number of points used for representation (McMaster 1989). On one hand, keeping all the data creates fuzzy, thick outlines when the scale is reduced, while on the other hand, characteristic details must be kept in the process of simplification. It is consequently necessary to search for a compromise.

Extensive literature (McMaster 1987a; Muller 1987; Thapa 1988; Boutoura 1989) on this problem raises two important questions:

1 What is a criterion of good simplification?
2 What do we really consider to be characteristic details?

To answer these questions, let us consider the *fractality* of cartographical curves. Cartographical curves may be considered as curves that are more or less space-filling to a degree that is measured by the fractal dimension—a global index based on the evaluation of all small, local irregularities. But are cartographical curves *fractals*? One could argue that they are *not* because:

- they involve *cutoff* scales (the inner cutoff may be defined as the size of the smallest details and the outer cutoff as the diameter of the whole curve) (Mandelbrot 1982, pp. 38–39). Strictly speaking, this is accurate: no *real* curve is a fractal curve but one may admit that any real curve is locally an approximation of a *mathematical* one and thus one never computes the fractal dimension but an *effective dimension* within the cutoff scales.
- they do not exhibit *self-similar* patterns because the various processes (geomorphological, geological, ecological, etc.) operating at different scales do not have the same influence, and consequently do not give self-similar details (Carpenter 1981; Shelberg et al. 1982; Goodchild and Mark 1987; Buttenfield 1989; Gatrell 1991). Even if we accept that cartographical lines are not fractal curves, it is necessary to stress that the property 'fractal' does not imply self-similarity. A self-similar curve is only a particular case of fractal curves. If similarity is used in a *strict* mathematical sense, no natural curve may be described as self-similar, as are mathematical models like Von Koch and Gosper curves. If similarity is used in a *statistical* sense (repetition of the same scaled irregularities all along the curve), then a self-similar curve may be described as *homogeneous*. One can generally say that a curve is fractal (Tricot 1994), if it is
  - *a)* non-rectifiable (every subarc not restricted to a single point has an infinite length), and
  - *b)* homogeneous (the dimension of any subarc is the same as the dimension of the whole curve).

As far as cartographical lines are concerned, they may be considered as non-rectifiable since their length increases steadily with the scale (Richardson 1961), but they are not necessarily homogeneous because they exhibit both straight parts (almost rectifiable) and chaotic parts (in which the local dimension is not the same everywhere). According to the above definition, we can conclude that

François Normant is a Research Associate at the Département de Mathématiques et de Génie industriel, École Polytechnique, C.P. 6079, Montréal QC H3C 3A7, e-mail fn@graf.polymtl.ca Axel van de Walle is a Ph.D. student in the Department of Materials Science and Engineering, MIT, Room 13-4069, 77 Massachusetts Ave., Cambridge MA USA 02139, e-mail avdw@mit.edu

a cartographical curve may not be a real fractal curve. Theoretically, the fractal dimension of a non-homogeneous curve is equal, to the limit, to its highest local dimension—but, in practice, no estimation algorithm can give that result. Instead, it yields a dimension[1] that can be viewed as an average of those of all the subarcs. It is thus a *global* space-filling characteristic of the curve, and then a significant cartographical parameter. From this point of view, a well-simplified curve should have the same fractal dimension as the original curve (Muller 1987; Muller 1991; Normant and Tricot 1993). However, a major problem occurs when one tries to compute the fractal dimension of a non-homogeneous curve: the estimation of the effective dimension will vary with the choice of the lower and upper cutoffs. Let us recall how one usually computes the fractal dimension:

- given a curve Γ in the plane and a point $x \in \Gamma$, we denote the open ball of radius ε centred in x by $\mathcal{B}_\varepsilon(x) = \{y : \text{dist}(x, y) < \varepsilon\}$.
- We call ε-*sausage of Minkowski* (Minkowski 1900) of Γ, the set of points $\Gamma(\varepsilon) = \{y : y \in \mathcal{B}_\varepsilon(x), x \in \Gamma\}$.

The fractal dimension, or more precisely the *Minkowski-Bouligand dimension* $\Delta(\Gamma)$ (Bouligand 1928; Bouligand 1929) of a planar curve Γ is defined by

$$\Delta(\Gamma) = \lim_{\varepsilon \to 0} \left\{ 2 - \frac{\log A(\Gamma(\varepsilon))}{\log \varepsilon} \right\}$$

where $A(\Gamma(\varepsilon))$ denotes the area of the Minkowski sausage. Since this limit cannot be computed, one usually sets a series of ε_i taken between a lower and an upper cutoff, draws the log-log plot of $A(\Gamma(\varepsilon))$ versus ε, estimates the value of the slope s of the graph with a linear regression, and obtains a dimension $\Delta = 2 - s$ (all other methods are theoretically equivalent to this method, called the Minkowski-sausage method).

Let us illustrate how the estimated dimension of a coastline can vary with the choice of the lower and upper cutoffs while the estimated dimension of a mathematical curve stays the same. In order to measure the fractal dimension, assuming that coastlines are not self-similar, we have used a general method, i.e., one not dedicated to a specific geometry. The Minkowski sausage method or the box-counting method could be used, but their convergence rate is too slow and in practice yields poor results (Tricot et al. 1988; Dubuc et al. 1989). We prefer to use a method co-developed by one of us (Normant and Tricot

1 Although one computes the effective dimension, never the fractal dimension, we will use the expression "fractal dimension," however incorrectly, throughout this article.

Figure 1 A part of the eastern shoreline of Baffin Island (1:200,000, 1360 points

Lower cutoff	Dimension
0.4%	1.40
0.8%	1.45
1.6%	1.50

Table 1 Estimation of the fractal dimension of the Baffin Island shoreline when the upper cutoff is set to 12%

Lower cutoff	Dimension
0.4%	1.24
0.8%	1.25
1.6%	1.23

Table 2: Estimation of the fractal dimension of the Von Koch curve when the lower cutoff is set to 12%

1991), called the Constant Deviation Variable-Step Method, which is completely adapted to the qualitative and quantitative study of non-homogeneous curves. It consists in dividing the curve Γ into subarcs Γ_i of constant breadth[1] ε to obtain a polygonal approximation of Γ to the precision ε. The length L_ε of this polygonal approximation of Γ is given by the sum of the diameters of the subarcs Γ_i. This operation is done for a series of ε chosen appropriately, that is, on a scale of values such that the curve is not seen as either rectifiable or as a point. A log-log plot of L_ε Versus ε gives the fractal dimension Δ using the formula

$$\Delta = \lim_{\varepsilon \to 0}\left(1 + \frac{\log L_\varepsilon}{|\log \varepsilon|}\right)$$

As an example of a natural curve, we used part of the eastern shoreline of Baffin Island, situated north of Quebec in the Northwest Territories of Canada. The curve was obtained from a map of the area using a digitizing tablet (Figure 1). As an example of a mathematical curve, we used the well-known Von Koch curve, whose theoretical dimension is log 4/log 3 ≈ 1.26 (Figure 2). The lower and upper cutoffs are given as percentages of the side of the bounding square of the curve.

The estimate of the dimension, made with a lower cutoff of 0.2% and an upper cutoff of 12%, gives 1.35 for the Baffin Island shoreline and 1.25 for the Von Koch curve. In Tables 1 and 2 we have set the upper cutoff to 12% and have used several values of the lower cutoff, while in Tables 3 and 4 we have set the lower cutoff to 0.2% and have used several values of the upper cutoff.

Increasing the lower cutoff corresponds to estimating the dimension of the curve without taking into account the smallest details, while decreasing the upper cutoff corresponds to estimating the dimension of the curve without taking into account the largest details. We may observe that the effective dimension of the mathematical curve (Von Koch) is independent of the lower and upper cutoffs used (1.25 ± 0.01) while the effective dimension of the natural curve (Baffin Island) changes dramatically with the lower and upper cutoffs (1.35 ± 0.15).

Thus, preserving the fractal dimension of a non-homogeneous curve is only meaningful if the lower cutoff used to evaluate the dimension before and after simplification is not smaller than the size of the removed details.

Finally, the preservation of the fractal dimension is not sufficient because curves of various geometry may have the same dimension; thus the characteristic properties of *local irregularities* (small indentations, meanders, and promontories) of the curves should also be taken into account during the simplification process.

In conclusion, an accurate, simplified curve will be a curve of the same fractal dimension (for a given tolerance factor) and of the same local properties as the original one.

1 All these notions are explained in the next section

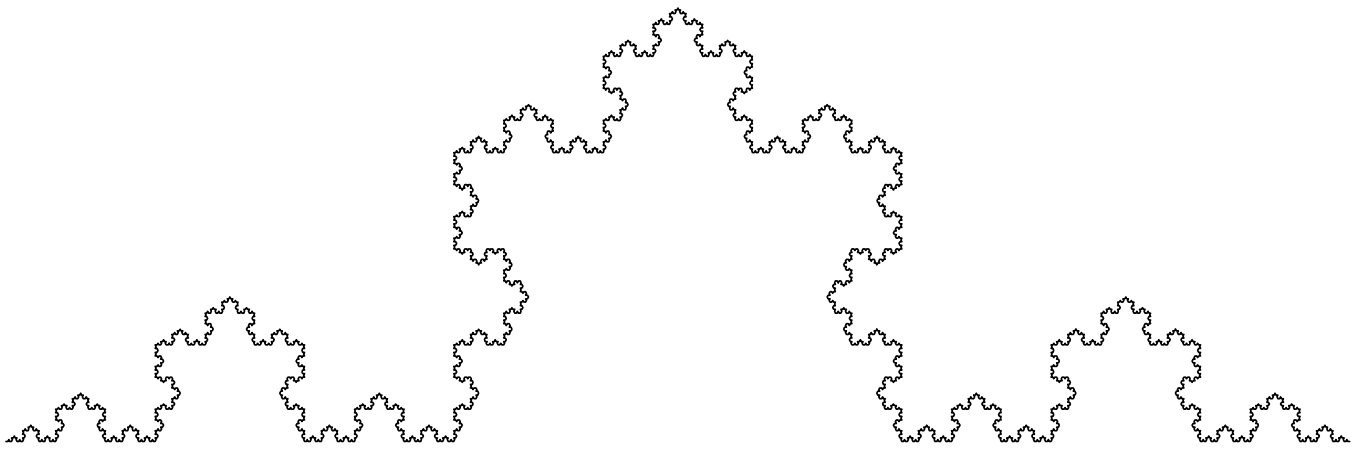

Figure 2 Van Koch curve (6 iterations: 4097 points)

Upper cutoff	Dimension
8%	1.32
4%	1.27
2%	1.23

Table 3 Estimation of the fractal dimension of the Baffin Island shoreline when the upper cutoff is set to 0.2%

Upper cutoff	Dimension
8%	1.26
4%	1.26
2%	1.26

Table 4 Estimation of the fractal dimension of the Von Koch curve when the upper cutoff is set to 0.2%

We base our simplification method on a new algorithm for computing the fractal dimension introduced by Tricot (1994). In order to simplify the curve with a precision ε, we cover it with overlapping convex sets of constant breadth ε (the ε-sausage of convex structuring elements) and retain the points of the original curve that lie on the boundary of this covering set. Notice that this new method is different from the one introduced in Normant and Tricot (1993), although both methods are based on local convex hulls.

The next section gives the mathematical definitions of local properties and describes the algorithm of the so-called ε-*Sausage of Convex Structuring Elements* method, and the following section presents some experiments on self-similar curves and coastlines for comparison with the Douglas-Peucker method.

Description of the Method

LOCAL CONVEX HULLS

Throughout this discussion, we shall be working solely with simple plane curves—i.e., plane curves without multiple points. It is convenient to consider such a curve Γ as a *trajectory*, where positions $\gamma(t)$ depend on a *time t*. The time is more generally called a *parameter*, and thus Γ is a *parametrized curve*. We denote by $[a, b]$ the time interval necessary to walk all along Γ. Every subarc of Γ can be written as $\gamma([t_1, t_2])$, where $[t_1, t_2]$ is a subinterval of $[a, b]$ (Figure 3).

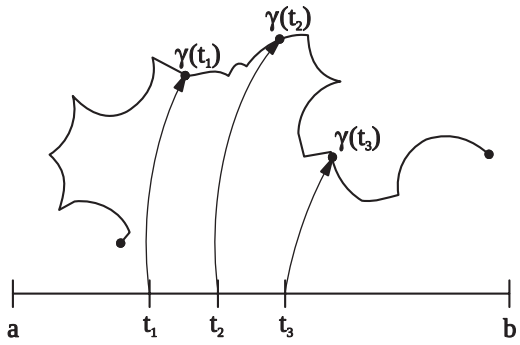

Figure 3 A parametrized curve with subarcs $\gamma([t_1, t_2])$ and $\gamma([t_2, t_3])$

We need to describe the local geometry of Γ, and for this purpose we use the notion of a *convex hull* of its subarcs. Before introducing the definition of a convex hull, let us give several definitions (Santaló 1976):

- A set of points K is called a *convex set* if for each pair of points $A \in K$ and $B \in K$, the segment $\overline{AB} \subset K$. For convenience, we shall assume throughout that the convex sets are bounded and closed, and we shall denote the boundary of K by ∂K (Figure 4).
- A line L is said to be a *line of support* of a convex set K at a point $P \in \partial K$, if $P \in L$ and K is contained in the closure of one of the two half planes into which L cuts the plane. Every point on the boundary of a convex set lies on a line of support and there are exactly two lines of support parallel to a given direction (Figures 4 and 5).
- The *diameter* of a convex set K, denoted diam(K), is the greatest distance between any two points of K. It is also the greatest distance between parallel lines of support of K (Figure 5).

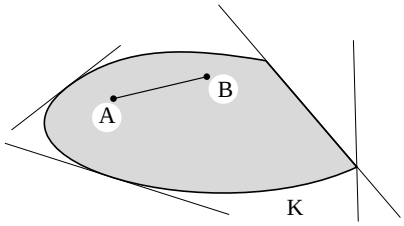

Figure 4 A convex set K, *a segment* $\overline{AB} \in K$, *and several lines of support of* K

- The *breadth* of a convex set K, denoted breadth (K), is the least distance between parallel lines of support of K (Figure 5).

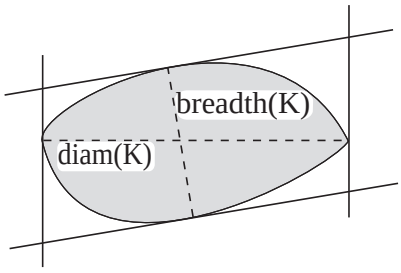

Figure 5 A convex set K, *its diameter* (K) *and its breadth* (K), *with parallel lines of support*

Intuitively, the diameter is an index that describes how *long* a convex set is, and the breadth describes how *thick*

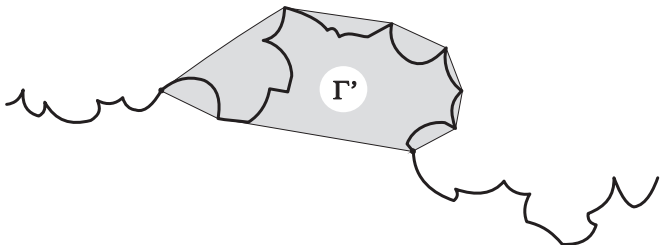

Figure 6 A subarc Γ' and its convex hull (shaded set)

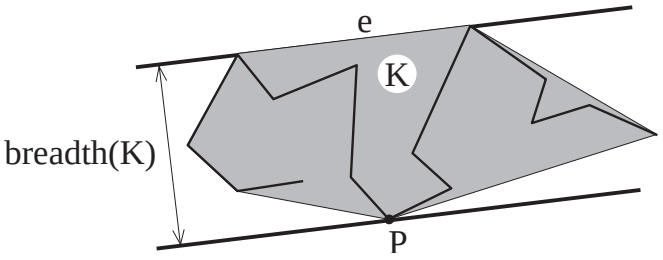

Figure 7 The convex hull K of a digitized curve; its breadth breadth (K); *the antipodal point P of the edge e with parallel lines of support*

it is.

For every subarc Γ' of Γ, the smallest convex set enclosing Γ' is called the *convex hull* of Γ', and denoted $K(\Gamma')$ (Figure 6). It is also the union of all the segments $\overline{AB} \subset K$ with $A \in \Gamma'$ and $B \in \Gamma'$.

Although wrongly, we shall use the expressions 'diameter' and 'breadth' of a subarc, meaning, in fact, diameter and breadth of its *convex hull*. We may note that a subarc is a line segment if and only if its breadth is zero. The more the breadth of the subarc approaches its diameter, the more that subarc may be considered chaotic. All these notions are introduced in Tricot (1990) and Tricot (1994), together with the corresponding mathematical analysis.

In the case of a digitized curve, every subarc Γ' of Γ is represented by a discrete number $n = j - i + 1$ of points $\{X_i, X_{i+1}, ..., X_j\}$, where point X_k corresponds to the value t_k of the parameter. The convex hull of Γ' is the convex hull of the set of these n points and thus is a *convex polygon* having at most n sides. The computation of the breadth of the convex hull of Γ' can be facilitated with the use of special points, called *antipodal points*. For an edge e of the convex polygon $K(\Gamma')$, we call the vertex of the convex hull that is the farthest away from the supporting line containing e the antipodal point of the edge, and denote by $d(e)$ the distance of the antipodal point e from the supporting line containing e (Figure 7), (Imai and Iri 1986). The breadth of $K(\Gamma')$ is the least distance $d(e_i)$, for all the edges e_i of the polygon.

To simplify the notations, since any subarc $\Gamma' = \gamma([t_1, t_2])$ corresponds to some parameter interval $([t_1, t_2])$, the diameter and the breadth of $K(\Gamma')$ may be considered as variables of two functions, denoted respectively by diam (t_1, t_2) and breadth (t_1, t_2). Thus

$$\text{diam}(t_1, t_2) = \text{diam}\ (K(\gamma([t_1, t_2])))$$

$$\text{breadth}(t_1, t_2) = \text{breadth}\ (K(\gamma([t_1, t_2])))$$

Principle

Let Γ be a parametrized curve, $[a, b]$ be the time interval and ε be the resolution. The principle of the method may be described in three steps as follows:

1 Starting from each point $\gamma(t_i)$ of Γ, find the first point $\gamma(t_j)$, with $t_i < t_j$, such that the convex hull $K(\gamma([t_i, t_j]))$ has a breadth greater or equal to ε, i.e., breadth $(t_i, t_j))) \geq \varepsilon$. If this point does not exist, then let t_j equal b. Convex hulls are built starting with three non-colinear points (forming a triangle) and then adding one point at a time until the proper breadth has been reached. This is a well-known problem in computational geometry, referred to as the *on-line problem* (Preparata 1979; Avis et al. 1985). The breadth is computed using the *rotating caliper method* (Shamos 1978; Toussaint 1983).

2 Merging all the convex hulls together, one obtains the ε-*sausage of the convex structuring elements* (Figure 8):

$$K_\varepsilon(\Gamma) = \bigcup_{ij} K(\gamma([t_i, t_j])) \text{ with } i, j \text{ s.t. breadth}(t_i, t_j)) \leq \varepsilon$$

$K_\varepsilon(\Gamma)$ may also be defined as the union of all the chords $\overline{\gamma(t_i), \gamma(t_j)}$, such that the arc $\gamma([t_i, t_j])$ is of breadth $\leq \varepsilon$. The algorithm for merging the local convex hulls together is described in detail in Normant and van de Walle (1994).

3 Following the parameterization, i.e., from time a to time b, define a sequence $\{t_k\}$ of $[a, b]$ such that points $\gamma(t_k)$ belong to the boundary of the ε-sausage of convex structuring elements $K_\varepsilon(\Gamma)$. These are, in some sense, extreme points of Γ where the "direction" changes at the precision ε. The simplified curve consists of all points $\gamma(t_k)$, which we consider the significant points.

Note that this procedure is only useful if one is interested in both the sausage and the simplified curve. A simpler procedure, which gives only the simplified curve, may be described in two steps as follows:

1 Starting from each point $\gamma(t_i)$ of Γ, find the first point $\gamma(t_j)$, with $t_i < t_j$, such that the convex hull $K(\gamma([t_i\ t_j]))$ has a breadth greater or equal to ε, i.e., breadth $(t_i, t_j) \geq \varepsilon$. If this point does not exist, then let t_j equal b. For each local convex hull, store in an array all the indices of the points that lie **inside** the hull.

2 The simplified curve consists of all points of the original curve whose indices are **not** in the array assigned at the previous step.

Note that both these algorithms are "naive" in the

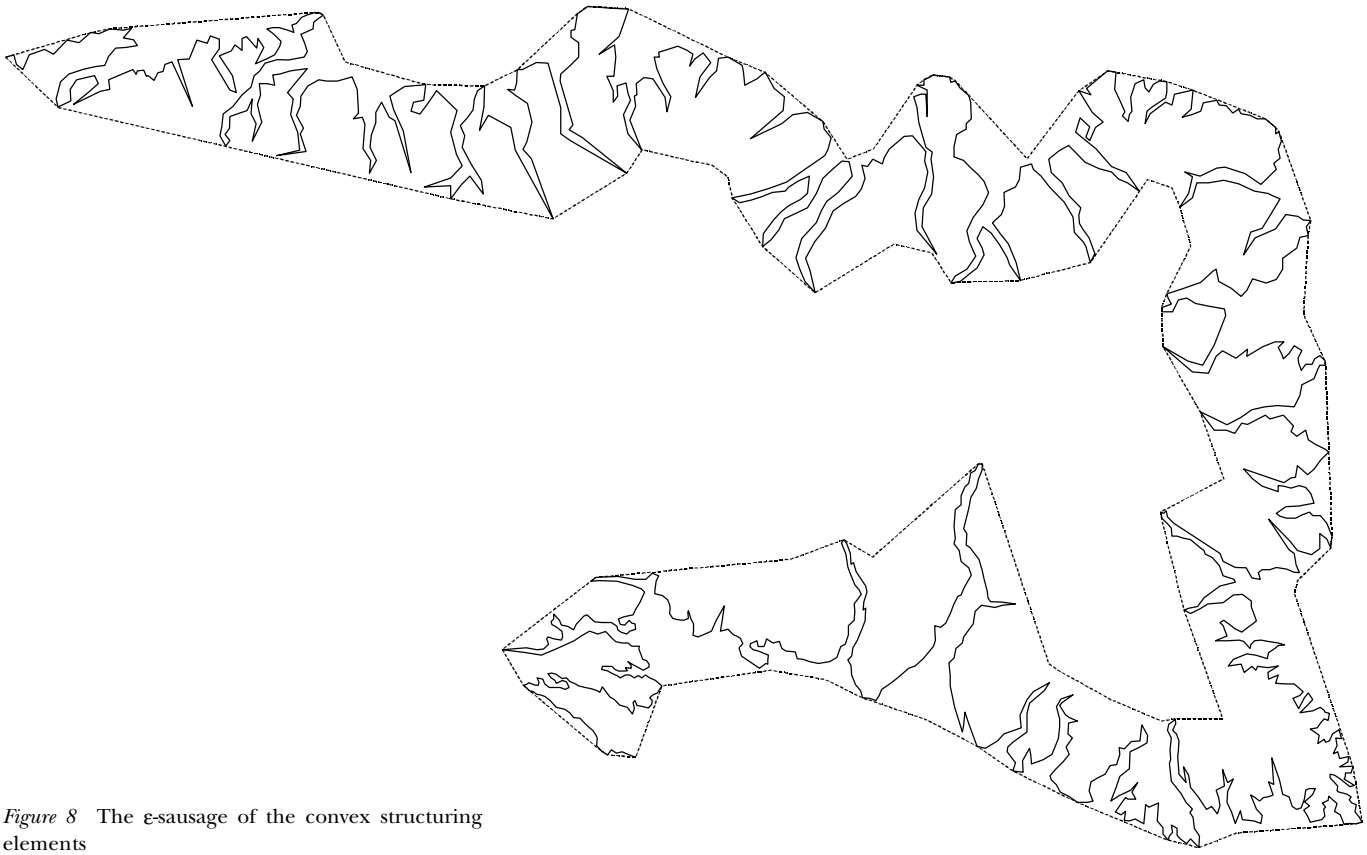

Figure 8 The ε-sausage of the convex structuring elements

sense that we have to compute all the local convex hulls. We have developed improved algorithms that allow us to avoid the construction of local convex hulls that would be totally included in others, and thus would not provide any information. Those algorithms are very technical, and beyond the scope of this paper. The reader may however refer to Normant and van de Walle (1994).

Let us now present an heuristic proof of the preservation of the fractal dimension. An equivalent definition of the Minkowski-Bouligand dimension is given by Tricot (1990; 1994) as follows:

$$\Delta(\Gamma) = \lim_{\varepsilon \to 0} \left\{ 2 - \frac{\log A(K_\varepsilon(\Gamma))}{\log \varepsilon} \right\}$$

where $A(K_\varepsilon(\Gamma))$ denotes the area of the ε-sausage of convex structuring elements defined earlier.

It is evident that if we use the present algorithm with a tolerance factor τ, for all $\varepsilon > \tau$, the area ε-sausage of convex structuring elements will not be affected significantly by the simplification. It appears that most of the points removed by the simplification process are inside the local convex hulls of breadth ε, and consequently would not change the shape of those hulls if they were present. Thus, by fixing the lower cutoff to τ (with the upper cutoff remaining the same), the effective dimension will be preserved.

Note that the notion of convexity has already been used in line generalization. Perkal (1956) introduced the notion of ε-convex sets and used it in Perkal (1956; 1966), to simplify the geometrical shape of an arc. This method consists of rolling a circle of radius ε along both sides of the line. The envelope of all positions of the circle has two branches that usually do not coincide. In fact, this method is not symmetrical, and the result depends on the side on which the ball is rolled. If the water side is chosen, the promontories are unaffected and the mouths smaller than ε are suppressed. The converse is true on the land side. For an extended discussion on the applications of Peucker's epsilon generalization, see Nystuen (1966), Rechlin (1972) and Chrisman (1983). It is important to stress the fact that the ε-sausage of convex structuring elements is not an ε-convex set according to Perkal's definition.

Applications

A REVIEW OF THE DOUGLAS-PEUCKER ALGORITHM

Douglas and Peucker (1973) developed an holistic, tolerance-band algorithm that considers the line in its entirety during processing. They gave two versions, one sequential and the other recursive, which were not exactly equivalent. We will use the recursive one, which is more efficient and uses less CPU time (Monmonier 1982), as follows:

The two end points of the line are selected for display. Distances are then computed between the intervening points and a

straight line connecting these end points. If the greatest perpendicular distance from the straight line to any of these points in less than a threshold distance, all intervening points can be eliminated from the generalized line. If the threshold distance is exceeded, the point most distant from the straight line is selected for display and is used as an end point for the two resulting subdivisions of the original line. Each of these portions is then examined separately, with its own hypothetical straight line and recomputed perpendicular distances to intervening points. The process continues until no further points are available for elimination.

This is a straightforward method of implementation and is, according to the literature, one of the most efficient (Muller 1987; McMaster 1987a; McMaster 1987b).

It is very interesting to notice that the Douglas-Peucker algorithm, with a threshold distance ε, selects points that are almost always on the boundary of the ε-sausage of convex structuring elements. We have the following theorem:

Theorem 1 The first point C selected by the Douglas-Peucker method (i.e., the point the most distant from the straight line connecting the two end-points A and B of the curve Γ) is always on the boundary of the ε-sausage of convex structuring elements of Γ;

$$C \in \partial(K_\varepsilon(\Gamma))$$

To prove this theorem, we first need two lemmas:

Lemma 1 *Let $K_1 \subset K_2 \subset K_3$ be three closed sets; then*

$$(K_1 \cap \partial K_3) \subset \partial K_2.$$

Proof If we denote respectively by $\complement K$ and $\mathring{K}$ the complement and the interior of a closed set K, we have $\partial K = K \cap \complement \mathring{K}$ and for all K, K' we have $\mathring{K} \subset \mathring{K}' \Rightarrow \complement \mathring{K}' \subset \complement \mathring{K}$. Then,

$$\begin{aligned} K_1 \cap \partial K_3 &= K_1 \cap (K_3 \cap \complement \mathring{K}_3) \\ &= K_1 \cap \complement \mathring{K}_3 \\ &\subset K_1 \cap \complement \mathring{K}_2 \\ &\subset K_2 \cap \complement \mathring{K}_2 \\ &= \partial K_2 \end{aligned}$$

Lemma 2 Let Γ be a curve of endpoints A and B, and let C be the point of Γ the most distant from the straight line $\mathcal{L}$ going through A and B. Then

$$C \in \partial(K(\Gamma)).$$

Proof Let d = dist (C, $\mathcal{L}$) = sup {dist P, $\mathcal{L}$) | P ∈ Γ} be the distance from C to $\mathcal{L}$ (Figure 9). For all Q ∈ $K(\Gamma)$:

- either Q ∈ Γ and then dist (Q, $\mathcal{L}$) ≤ *d*; or
- Q belongs to a chord $\overline{MN}$ of Γ: then

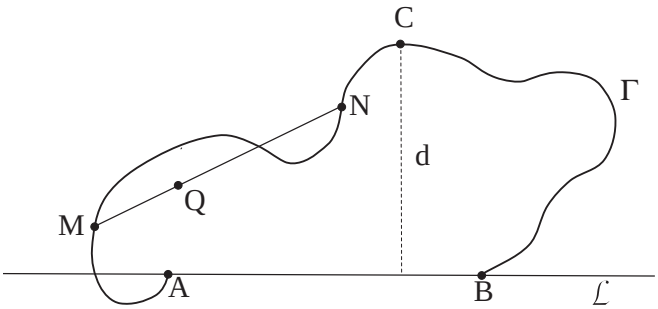

Figure 9 A curve Γ *with end points* A *and* B, *a line* $\mathcal{L}$ *going through* $\overline{AB}$ *and a point* Q *on the chord* $\overline{MN}$

$$\text{dist}(Q, \mathcal{L}) \le \max\{\text{dist}(M, \mathcal{L}), \text{dist}(N, \mathcal{L})\} \le d.$$

In both cases the result is the same: dist (Q, $\mathcal{L}$) ≤ *d*. Let us suppose that C ∈ $\mathring{K}(\Gamma)$: there exists ε > 0 such that the ball $B_\varepsilon(C) \subset K(\Gamma)$, thus there exists a point Q ∈ $K(\Gamma)$ such that dist (Q, $\mathcal{L}$) = dist (C, $\mathcal{L}$) + ε. This leads to dist (C, $\mathcal{L}$) ≤ *d* – ε, which is impossible. As a consequence C ∈ $\partial(K(\Gamma))$.

We now prove Theorem 1:

Proof: We have $\Gamma \subset K_\varepsilon(\Gamma) \subset K(\Gamma)$. Thus, Lemma 2 implies that C ∈ $(\Gamma \cap \partial(K(\Gamma))$, and Lemma 1 implies that C ∈ $\partial(K_\varepsilon(\Gamma))$, which completes the proof.

It would be helpful if the next points selected by the Douglas-Peucker algorithm were also on the boundary. Unfortunately, that is not always true. The task remains to determine sufficient and necessary conditions that the curve must meet to achieve that result. A sufficient condition, although certainly not a necessary one, is that any two disjoint arcs of the curve have disjoint convex hulls. Geometrically, this means that the curve does not come back close to itself too often—a property related to the notion of *expansive curve*, which is studied in Tricot (1994). We can observe, in Figure 10, the Baffin Island shoreline from Figure 1 simplified with a threshold of 8%, drawn within the ε-sausage of convex structuring elements. All but one of the points selected by the Douglas-Peucker algorithm are on $\partial(K_\varepsilon(\Gamma))$.

In the next section, we will compare the Douglas-Peucker method (D.P. method) and the Epsilon Sausage method (E.S. method) on their ability to preserve the fractal dimension of a self-similar curve and a non-homogeneous curve for a given tolerance factor ε.

Experimental Results and Discussion

As an example of natural curves, we use parts of the eastern shoreline of Baffin Island situated north of Quebec in the Northwest Territories of Canada. The curves were ob-

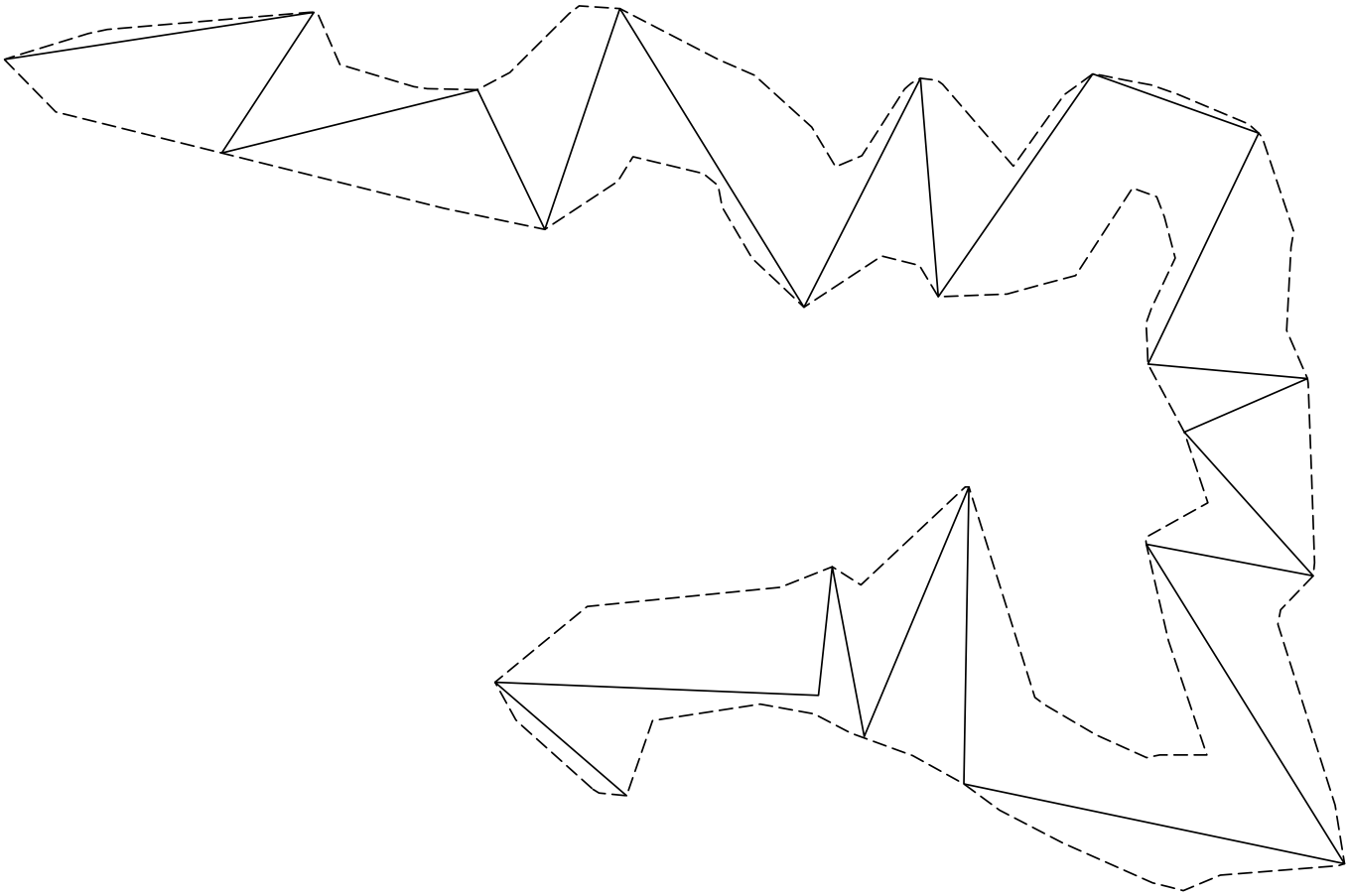

Figure 10 A curve simplified with the Douglas-Peucker algorithm (threshold distance ε = 8%*) and the* ε*-sausage of convex structuring elements*

tained from a map of the area using a digitizing tablet (Figures 1 and 11). As an example of a self-similar curve, we use the well-known Von Koch curve whose theoretical dimension is log 4/log 3 ≈ 1.26 (Figure 2). The fractal dimension is viewed as a characteristic parameter of the space-filling behaviour of the line and should be independent of the simplification when estimated with a lower cutoff, equal to the tolerance factor used. In order to estimate the fractal dimension, we use a Constant Deviation Variable-Step Method described in the first section 1.

We use three different tolerance factors, expressed as percentages of the side of the bounding square of the curve, to obtain three levels of simplification. The results appear in Table 5 for the Von Koch curve and in Tables 6 and 7 for the Baffin shoreline. Δ is the estimated fractal dimension and *N* the number of points of the simplified curve.

It may be seen immediately that the Douglas-Peucker algorithm retains many fewer points than the ε-Sausage algorithm, and thus selects only some of the points of the curve lying on the boundary of the ε-sausage. That could be expected because—in the Douglas-Peucker algorithm—no point is at a distance smaller than ε from the next and the previous points, while—with the ε-sausage algorithm—several points can be very close to one another (that is the case when a part of the curve overlaps with the boundary of the ε-sausage).

As expected, the ε-sausage method preserves the dimension. However, surprisingly enough, the Douglas-Peucker method is also very accurate. The slight difference, 0.0l to 0.03, shown by the Douglas-Peucker is not meaningful, and thus one may conclude that this method satisfies our criterion of good simplification.

Conclusion

We try to describe the inner structure of a curve using the qualities of breadth and diameter. Even if they are sufficient to determine the fractal dimension (Tricot 1990), they are only some of the descriptors of the curve shape, and there is no doubt that other parameters will have to be found in order to describe a curve totally: the fractal dimension is not the complete answer, but only a criterion of global behaviour. The fractal dimension is mathematically defined for non-homogeneous curves, such as cartographic ones, but unfortunately while algorithms exist to estimate the effective dimension, there are none

Figure 11 A part of the eastern shoreline of Baffin Island (1:100,000, 1655 points)

ε in percent	Original Curve	D. P. Method		E. S. Method	
	D	N	D	N	D
0.25	1.25	513	1.22	1219	1.24
0.50	1.14	257	1.22	623	1.24
1.00	1.24	129	1.21	3670	1.24

Table 5 Number of points and estimated dimension of the Von Koch curve simplified with different tolerance factors

ε in percent	Original Curve	D. P. Method		E. S. Method	
	D	N	D	N	D
0.25	1.37	698	1.36	1208	1.37
0.50	1.41	410	1.40	913	1.41
1.00	1.46	259	1.44	559	1.46

Table 6 Number of points and estimated dimension of the Baffin shoreline (from Figure 1) simplified with different tolerance factors

ε in percent	Original Curve	D. P. Method		E. S. Method	
	D	N	D	N	D
0.25	1.33	526	1.32	1189	1.33
0.50	1.37	266	1.36	706	1.37
1.00	1.41	152	1.38	425	1.41

Table 7 Number of points and estimated dimension of the Baffin shoreline (from Figure 10) simplified with different tolerance factors

to estimate the fractal dimension. We show that the effective dimension of non-homogenous curves depends on the lower and upper cutoffs used, and thus should be estimated carefully with an accurate tool such as the one we have developed in Normant and Tricot (1991). Dedicated methods that assume that the curve is self-similar, such as the divider method, should be discarded.

The algorithm we propose here for simplification does not depend on arbitrary starting points as do almost all other algorithms (it is well known that if we divide a curve into two parts and apply the Douglas-Peucker algorithm to each, the two branches obtained will differ from the simplified curve throughout, and not just at their end points). One of the drawbacks of the new method is that it retains too many points from the original curve, especially on convex subarcs. However, that can be overcome with a simple post-processing filtering technique such as removal of points within a given threshold.

This new algorithm is also useful for checking the validity of other simplification methods against the criterion of preservation of the fractal dimension. There are good geometrical reasons to think that a simplification method that would retain points on the boundary of the ε-sausage of convex structuring elements would also preserve the effective dimension. In fact, the Douglas-Peucker method bears this out, though further studies are needed to determine in which particular cases a point of a simplified curve does not lie on the boundary of the sausage.

References

AVIS, D., H. EL GINDY, and R. SEIDEL. 1985. Simple on-line algorithms for convex polygons. In G.T. Toussaint, ed., *Computational Geometry*, volume 2 of *Machine Intelligence and Pattern Recognition*, pp. 23–36. North Holland, Amsterdam.

ISBN 0-444-87806-8.

BOULIGAND, G. 1928. Ensembles impropres et nombre dimensionnel. *Bull. Soc. Math. France* 52: 20–334 and 361–376.

———. 1929. Sur la notion d'ordre de mesure d'un ensemble plan. *Bull. Soc. Math. France* 53: 185–192.

BOUTOURA, C. 1989. Line generalization using spectral techniques. *Cartographica* 26/3-4: 33–48.

BUTTENFIELD, B.B. 1989. Scale-dependence and self-similarity in cartographic lines. *Cartographica* 26/1, Monograph 40: 79–100.

CARPENTER, L.C. 1981. Computer rendering of fractal curves and surfaces. In *Proceedings ACM SIGGRAPH Conference*, Seattle, Washington.

CHRISMAN, N.R. 1983. Epsilon filtering: A technique for automated scale changing. In *Proceedings of the Annual Conference of the American Congress for Surveying and Mapping*, pp. 322–331, Washington DC.

DOUGLAS, D.H. and T.K PEUCKER. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Canadian Cartographer* 10/2: 112–122.

DUBUC, B., J.-F. QUINIOU, C. ROQUES-CARMES, C. TRICOT, and S.W. ZUCKER. 1989. Evaluating the fractal dimension of profiles. *Physical Review A* 39/3: 1500–1512.

GATRELL, A.C. 1991. Concepts of space and geographical data. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, eds., *Geographical Information Systems*, volume 1. pp. 119–134. Longman Scientific and Technical.

GOODCHILD, M.F. and D.M. MARK. 1987. The fractal nature of geographic phenomena. *Annals of the Association of American Cartographers* 77/2: 265–278.

IMAI, H. and M. IRI. 1986. Computational-geometric methods for polygonal approximations of a curve. *Computer Vision, Graphics, and Image Processing* 36: 31–41.

MANDELBROT, B.B. 1982. *The Fractal Geometry of Nature*. pp. 38–39. W.H. Freeman and Company, San Francisco. ISBN 0-7167-1186-9.

McMASTER, R.B. 1987a. Automated line generalization. *Cartographica* 24/2: 74–111.

———. 1987b. The geometric properties of numerical generalization. *Geographical Analysis* 19/4: 330–346.

———. 1989. Numerical generalization in cartography. *Cartographica* 26/1, Monograph 40.

MINKOWSKI, H. 1900. Über die Begriffe Länge, Oberfläche und Volumen. *Jahresber. Deutsch. Math.-Verein.* 9/2: 115–121.

MONMONIER, M.S. 1982. *Computer Assisted Cartography, Principles and Prospects*. Prentice-Hall Inc., Englewood Cliffs, NJ. ISBN 0-13-165308-3.

MULLER, J.-C. 1987. Fractal and automated line generalization. *The Cartographic Journal* 24: 27–34.

———. 1991. Generalization of spatial databases. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, eds., *Geographical Information Systems*, volume 1. pp. 457–475. Longman Scientific and Technical.

NORMANT, F. and C. TRICOT. 1991. Method for evaluating the fractal dimension of curves using convex hulls. *Physical Review A* 43/12: 6518–6525.

———. 1993. Fractal simplification of lines using convex hulls. *Geographical Analysis* 25/2: 118–129.

NORMANT, F. and A. VAN DE WALLE. 1994. An algorithm for computing the ε-sausage of convex structuring elements of a planar curve. Unpublished technical report.

NYSTUEN, J.D. 1966. Effects of boundary shape and the concept of local convexity. In J.D. Nystuen, ed., *Michigan Inter-University Community of Mathematical Geographers*, number 10, pp. 1–22, University of Michigan, Ann Arbor, MI.

PERKAL, J. 1956. Sur les ensembles ε-convexes. *Colloquium Mathematicum* 4/1: 1–10.

———. 1958. O dlugosci krzywych empirycznych. *Zastosawania Matematyki* 3/3–4: 257–286.

———. 1966. On the length of empirical curves. In J.D. Nystuen, ed., *Michigan Inter-University Community of Mathematical Geographers*, number 10, pp. 17–20. University of Michigan, Ann Arbor, MI. Translated by W. Jackowski.

PREPARATA, F.P. 1979. An optimal real-time algorithm for planar convex hulls. *Communications of the ACM* 22/7: 402–405.

RECHLIN, A.T.M. 1972. The epsilon generalization. *Annals of the Association of American Cartographers* 62/4: 578–581.

RICHARDSON, L.F. 1961. The problem of contiguity: An appendix of statistics of deadly quarrels. *General System Yearbook* 6: 139–187.

SANTALÓ, L.A. 1976. Integral geometry and geometric probability. In G.-C. Rota, ed., *Encyclopedia of Mathematics and its Applications*, volume 1. Addison-Wesley Publishing Company. ISBN 0-201-13500-0.

SHAMOS, M.I. 1978. *Computational Geometry*. PhD thesis, Yale University.

SHELBERG, M.C., H. MOELLERING, and N. LAM. 1982. Measuring the fractal dimensions of empirical cartographic curves. In J. Foreman, ed., *Proceedings Auto-Carto 5*, pp. 481–490, Crystal City, VA. American Society of Photogrammetry and American Congress on Surveying and Mapping. ISBN 0-937294-44-6.

THAPA, K. 1988. Automated line generalization using zero crossing. *Photogrammetric Engineering and Remote Sensing* 54/4: 511–517.

TOUSSAINT, G.T. 1983. Solving geometric problems with the 'rotating calipers'. In *Proceedings IEEE Melecon*, Athens, Greece.

TRICOT, C. 1990. Local convex envelopes of a curve. *Real Anal. Exchange* 15: 675–695.

———. 1994. *Curves and Fractal Dimension*. Springer-Verlag, New York. ISBN 0-387-94095-2.

TRICOT, C., J.-F. QUINIOU, D., WEHBI, C. ROQUES-CARMES, and B. DUBUC. 1988. Évaluation de la dimension fractale d'un graphe. *Revue de Physique Appliquée*, 23: 111–124.

Résumé Nous étudions le problème de la simplification d'une courbe du point de vue de la géométrie et des fractales. Nous trouvons que les méthodes les meilleures sont celles qui gardent constante la valeur estimée de la dimension fractale, tant avant qu'après la simplification, et qui incluent les détails durant le traitement.

Nous proposons ici à la fois une méthode et un critère de

simplification basés sur un nouvel algorithme de calcul de la dimension fractale introduit par Tricot en 1994. Nous recouvrons la courbe par des ensembles convexes non-disjoint de largeur constante (pour obtenir la saucisse d'éléments structurants convexes) et retenons les points de la courbe originale qui touchent à la frontière de ce recouvrement. Cet algorithme est à la fois précis et universel : il n'est pas nécessaire de poser d'hypothèses arbitraires à propos de la structure de la courbe.

Nous allons démontrer de façon expérimentale que la méthode bien connue de Douglas-Peucker partage des propriétés géométriques très intéressantes avec cet algorithme. Elle a donc la capacité de maintenir la dimension fractale lorsqu'elle est évaluée avec une valeur approchée plus basse, égale à la distance de seuil.

Zusammenfassung Die Enveloppe von lokalkonvexen Hüllen einer Kurve und der Douglas-Peucker-Algorithmus Wir untersuchen das Problem der Vereinfachung einer Kurve aus der Sicht der Geometrie und der Fraktaltheorie. Nach unserer Meinung sind die besten Methoden diejenigen, die einen geschätzten Wert der fraktalen Dimension bei der Vereinfachung unverändert lassen und die während der Datenverarbeitung lokale Details berücksichtigen. Wir schlagen hier sowohl eine Methode als auch ein Kriterium der Vereinfachung vor. Beide basieren auf einem neuen Algorithmus zur Berechnung der fraktalen Dimension, den Tricot (1994) eingeführt hat. Wir bedecken die Kurve mit einander überschneidenden konvexen Punktmengen konstanter Breite (der Enveloppe von konvexen strukturierenden Elementen) und behalten diejenigen Punkte der Ausgangskurve bei, die auf der Grenze der Gesamtmenge liegen. Der behandelte Algorithmus ist sowohl exakt als auch allgemein anwendbar, und bei seiner Verwendung braucht man keine willkürlichen Hypothesen über die Struktur der Kurve zu aufzustellen. Wie wir experimentell zeigen werden, teilt die bekannte Douglas-Peucker-Methode einige sehr interessante geometrische Eigenschaften mit dem hier behandelten Algorithmus. Sie kann daher auch die fraktale Dimension bewahren, wenn man einen niedrigeren Trennwert vorgibt, der dem Entfernungs-Schwellenwert gleich ist.