Gilbert Helmberg

Getting Acquainted with Fractals

de Gruyter

Gilbert Helmberg

Getting Acquainted with Fractals

W DE G
Walter de Gruyter
Berlin · New York

Gilbert Helmberg
Kalkofenweg 5
6020 Innsbruck
Austria

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Helmberg, Gilbert.
Getting acquainted with fractals / by Gilbert Helmberg.
p. cm.
Includes bibliographical references.
ISBN 978-3-11-019092-2 (hardcover : alk. paper)
1. Fractals. I. Title.
QA614.86.H45 2007
514'.742–dc22

2006102211

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-11-019092-2

Printed in Germany.
Coverdesign: +malsy, kommunikation und gestaltung, Willich.
Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen.

de Gruyter Textbook

Helmberg · Getting Acquainted with Fractals

Preface

To someone, having heard about fractals but not yet acquainted with them, they might seem to be regarded with suspicion: How could "real" objects – accessible by sight and not only by thought – be replicas of arbitrarily small parts of themselves? How could a continuous path which runs almost everywhere parallel to sea level climb up to any height? How could a continuous curve pass through every point of a square?

Getting acquainted with fractals opens a glimpse into a world of wonders, but these wonders are strongly supported by a frame of serious mathematics in which various of its branches play together: geometry, analysis, linear algebra, topology, measure theory, functions of complex variables, algebra, ...

I have tried to do justice to both aspects: the fascination of geometric objects as well as the serious mathematical background – as far as an advanced undergraduate level. At some points, where the technicalities would transgress this level, I have at least indicated where an interested reader could find the whole story. I hope the presentation adds something worthwhile to the many remarkable books on this topic which also lead much farther into the world of fractals.

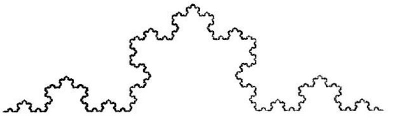
These books also contain something which a reader might miss in the present one: I have chosen to avoid the possibility of frustrating the reader by expecting him to do exercises; he will find them in abundance in the mentioned books (e.g. [Barnsley, 1988], [Falconer, 1990]) if he wants to. However, it is at least my intention to make accessible – via the internet address `http://techmath.uibk.ac.at/helmberg` – the programs producing the illustrations, thus enabling the reader to create and play with fractals according to his own taste.

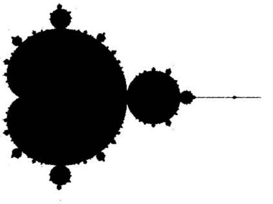
My thanks are due to the de Gruyter Publishing Company, in particular to Dr. Plato, for their interest in and support of this book project. My first book has been dedicated to my parents, my wife, and my two eldest children, but there are more people who mean very much to me. Therefore this book is dedicated

to Chri, Moni, and Mui.

Innsbruck, Cavalese, August 2006 Gilbert Helmberg

Contents





1 Fractals and dimension

1.1 The game of deleting and replacing

The word "fractal" comes from the Latin word "frangere" (with past participle "fractus") which means "to break", "to destroy". Let us begin with exploring how such a destruction process may still generate some new mathematical object displaying interesting features.

1.1.1 The CANTOR set

Let us define an operation f (such an operation is commonly called an *operator*) working on any closed segment $[a, b] \subset \mathbb{R}$ (= the real line) by deleting the open middle third $]a + \frac{b-a}{3}, b - \frac{b-a}{3}[$, and let us denote the interval $[0, 1] \subset \mathbb{R}$ by $A_{(0)}$. Application of f to $A_{(0)}$ deletes the interval $]\frac{1}{3}, \frac{2}{3}[$ and produces a closed set

$$A_{(1)} = [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1],$$

the union of the two disjoint closed intervals $A_0 = [0, \frac{1}{3}]$ and $A_1 = [\frac{2}{3}, 1]$, each of which has length $\frac{1}{3}$. If we apply f now to $A_{(1)}$ we get a closed set

$$A_{(2)} = f(A_{(1)}) = f(f(A_{(0)})) \subset A_{(1)}$$

consisting of four disjoint intervals $A_{0,0}$, $A_{0,1}$, $A_{1,0}$, $A_{1,1}$ of length $\frac{1}{9} = \frac{1}{3^2}$ each. Since we want to continue the application of f, in order to avoid the clumsy notation $f(f(\ldots))$ let us use the notation

$$\begin{aligned} f^{(0)}(A) &:= A, \\ f^{(1)}(A) &:= f(A), \\ f^{(k+1)}(A) &:= f(f^{(k)}(A)). \end{aligned}$$

(We shall call the index k the *level* of the construction.) Applied to our interval $A_{(0)}$ this allows us to define a sequence of closed sets $A_{(k)}$ $(1 \leq k < \infty)$ by

$$A_{(k)} := f^{(k)}(A_{(0)})$$

satisfying

$$A_{(0)} \supset A_{(1)} \supset \cdots \supset A_{(k)} \supset A_{(k+1)} \supset \cdots . \tag{1.1}$$

The set $A_{(k)}$ is the union of 2^k closed intervals $A_{j_1,\ldots,j_k}$ $(j_i \in \{0, 1\},\ 1 \leq i \leq k)$ of length $\frac{1}{3^k}$ each. A sequence $\{A_{(k)}\}_{k=1}^{\infty}$ as well behaved as indicated by (1.1) raises the question whether there exists, in some sense, a limit set A. Indeed, by a well known topological theorem, the decreasing sequence of non-empty compact sets $\{A_{(k)}\}_{k=1}^{\infty}$ has the property that the set

$$A := \bigcap_{k=1}^{\infty} A_{(k)},$$

called the CANTOR *set* [Cantor, 1883], is compact and not empty. See Figure 1.1 for an illustration of the set $A_{(4)}$.

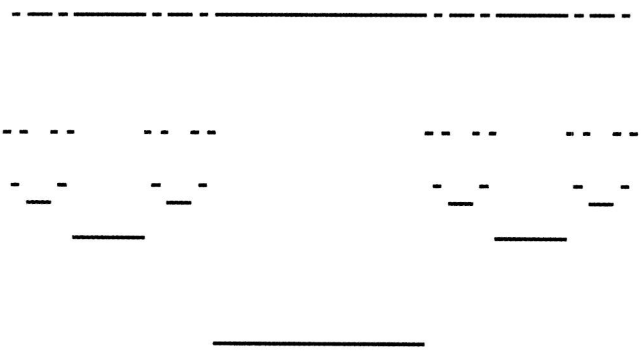


Figure 1.1. The set $A_{(4)}$, pictured in blue, is the union of sixteen closed component-intervals. The open set $[0,1] \setminus A_{(4)}$, pictured in red, is decomposed according to the intervals deleted at levels 1, 2, 3 and 4.

Still, as to the "size" of the set A, we notice that it is contained in all sets $A_{(k)}$ $(1 \leq k < \infty)$; as observed above, the total length of the 2^k component-intervals of $A_{(k)}$ is $\frac{2^k}{3^k} = (\frac{2}{3})^k$ which approaches zero as $k \to \infty$. If A is to have any "length" in some sense at all, it therefore must be zero. Indeed, using one-dimensional LEBESGUE measure $\mathcal{L}$ $(= \mathcal{L}^1)$, which on the sets $A_{(k)}$ coincides with their lengths, by a well-known theorem of measure theory we get

$$\mathcal{L}(A) = \mathcal{L}\Big(\bigcap_{k=1}^{\infty} A_{(k)}\Big) = \lim_{k\to\infty} \mathcal{L}(A_{(k)}) = \lim_{k\to\infty} \left(\frac{2}{3}\right)^k = 0.$$

It is not surprising that A is not empty: the countably many end points of all component intervals $A_{j_1,\ldots,j_k}$ $(0 \leq k < \infty,\ j_i \in \{0,1\},\ 1 \leq i \leq k)$ are never deleted by any application of f and therefore are all contained in A. But there are more points surviving all these applications:

Let us write every non-zero $x \in [0,1]$ as an infinite series $x = \sum_{k=1}^{\infty} \frac{x_k}{3^k}$ $(x_k \in \{0,1,2\})$, in short $x = 0.x_1x_2\ldots$ with the understanding that any finite sum of the form $x = 0.x_1x_2\ldots 1 = \sum_{k=1}^{n} \frac{x_k}{3^k}$ $(x_n = 1)$ shall be written as a non-ending periodic triadic fraction

$$x = 0.x_1x_2\ldots x_{n-1}022\ldots = \sum_{k=1}^{n-1} \frac{x_k}{3^k} + \sum_{k=n+1}^{\infty} \frac{2}{3^k}.$$

Application of f to $A_{(0)}$ eliminates all points x for which $x_1 = 1$. The set $A_{(1)}$ therefore contains none of these. Renewed application of f to $A_{(1)}$ now eliminates all points x for which $x_2 = 1$ (in both intervals A_0 and A_1 which are characterized by $x_1 = 0$ and $x_1 = 2$ respectively). Repeated application of f subsequently eliminates all points $x \in A_{(0)}$ for which $x_k = 1$ $(1 \leq k < \infty)$. What remains? Precisely the set of all points $x \in A_{(0)}$ whose "digits" x_k are either 0 or 2. It is well known that the points of the interval $[0, 1]$, apart from the countably many "dyadic rational" points, are in one-to-one correspondence with the points which in dyadic notation may be written as $y = 0.y_1 y_2 \ldots$ $(y_k \in \{0, 1\})$. The conclusion is that our set A is not countable but contains as many points as the interval $[0, 1]$, i.e. has the cardinality of the reals.

A mathematician may be tempted to exploit the relation between the set A and a subset of $[0, 1]$ even further. Just now we have associated with the point

$$y = \sum_{k=1}^{\infty} \frac{y_k}{2^k} \qquad \Big(y_k \in \{0, 1\}, \sum_{y_k=0} 1 = \sum_{y_k=1} 1 = \infty\Big)$$

the point

$$a(y) = \sum_{k=1}^{\infty} \frac{2y_k}{3^k} \in A.$$

Denoting by $\mathbb{N}$ the set of natural numbers, we may extend this mapping a to the compact topological product $\{0, 1\}^{\mathbb{N}}$ of all $\{0, 1\}$-sequences by defining

$$\tilde{a}(\tilde{y}) := \sum_{k=1}^{\infty} \frac{2y_k}{3^k} \qquad \text{for } \tilde{y} = \{y_k\}_{k=1}^{\infty} \quad (y_k \in \{0, 1\}),$$

e.g. if $\tilde{y} = \{0, 1, 1, \ldots\}$, then $\tilde{a}(\tilde{y}) = \sum_{k=1}^{\infty} \frac{2}{3^k} = \frac{1}{3}$, while for $\tilde{y} = \{1, 0, 0, \ldots\}$ we get $\tilde{a}(\tilde{y}) = \frac{2}{3}$. It is not hard to see that the mapping $\tilde{a} : \{0, 1\}^{\mathbb{N}} \to A$ is bijective (every point of A is the image of exactly one sequence in $\{0, 1\}^{\mathbb{N}}$) and continuous. A well-known topological theorem (cf. [Kelley, 1955, p. 141]) then asserts that A is homeomorphic to $\{0, 1\}^{\mathbb{N}}$. In particular, A is completely disconnected (in topology such a space is also called *zero-dimensional*) and *perfect* (i.e. closed without isolated points), but nowhere dense. Remembering that $[0, 1]$ may be considered as a subset of $\{0, 1\}^{\mathbb{N}}$ and roughly speaking, the mapping $\tilde{a}$ furnishes an extended parametrization of the set A (i.e. to every dyadic rational point of $[0, 1]$ there correspond two "neighbouring" points of A).

At this point we may notice one more property of the set A which is important for us since it will turn up in adapted form repeatedly in sets which we legitimately may call "fractals": suppose we omit the component A_1 of the set $A_{(1)}$ and restrict repeated application of f to the interval A_0. What would we have got? Evidently part of A, to wit a copy of the set A, only reduced by a factor $\frac{1}{3}$ in size. In fact, every component set $A_{j_1,\ldots,j_k}$ of $A_{(k)}$, treated by itself with successive applications of f, produces a set which is part of A and, at the same time, a copy similar to A but reduced by a factor $\frac{1}{3^k}$. In other words, one may say that the set A is "self-similar" in the sense that it consists of smaller parts which are still similar to A.

A slight adaption of our construction of the CANTOR set furnishes a set with strikingly different features. Since we are now going to move from $\mathbb{R}$ to $\mathbb{R}^2$, and more generally to $\mathbb{R}^n$, for any vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we shall use the EUCLIDEAN norm $|x| = \sqrt{\sum_{k=1}^{n} x_k^2}$. The (EUCLIDEAN) distance of two points $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$ is then given by $|a - b|$.

1.1.2 The KOCH curve

We shall now modify the operator f considered in Section 1.1.1 by allowing it to work on any closed segment $[a, b]$ in the plane $\mathbb{R}^2$, and in the following way: it not only deletes but replaces the open middle third $]a + \frac{b-a}{3}, b - \frac{b-a}{3}[$ by two sides of an equilateral triangle, side length $|\frac{b-a}{3}|$, located to the left of $[a, b]$ if this segment is directed from a to b. We shall apply f also to piecewise linear curves in $\mathbb{R}^2$. Such a curve E is the graph of a piecewise linear, not necessarily continuous, function $g : [0, 1] \to \mathbb{R}^2$. It consists of finitely many segments $[a_j, b_j]$ $(1 \leq j \leq n)$, at most pairwise joined at their endpoints. The result $f(E)$ of applying f to E is obtained by applying the operator f to each of the component segments of E.

We start out again with the segment $A_{(0)} = [0, 1]$ on the x-axis. Application of f to $A_{(0)}$ produces a continuous piecewise linear curve $A_{(1)} = f(A_{(0)})$ consisting of four segments denoted consecutively by A_j $(0 \leq j \leq 3)$, each of which has length $\frac{1}{3}$ (Figure 1.2).

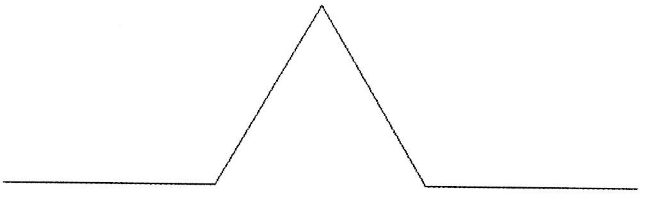


Figure 1.2. The generator $A_{(1)}$ of the KOCH curve.

Why not apply f again, this time to $A_{(1)}$, i.e. to each of these four segments? The result is a continuous piecewise linear curve $A_{(2)} = f^{(2)}(A_{(0)})$ consisting of 4^2 segments A_{j_1,j_2} $(0 \leq j_i \leq 3)$ of length $\frac{1}{3^2}$ each. Repetition of this procedure furnishes a sequence of continuous piecewise curves $A_{(k)}$ consisting of 4^k segments $A_{j_1,\ldots,j_k}$ $(0 \leq j_i \leq 3,\ 1 \leq i \leq k)$ of length $\frac{1}{3^k}$ each (Figures 1.3–1.5). Unfortunately, however, these curves, considered as subsets of $\mathbb{R}^2$, do not anymore satisfy (1.1). Do they still converge in some sense to some limit? The eye emphatically approves, but does mathematics support this impression?

In order to investigate the situation, we turn our attention to the sequence of segments $A_{j_1,\ldots,j_k}$ in $\mathbb{R}^2$ constituting the curve $A_{(k)}$ successively, starting at $(0, 0)$ and ending at $(1, 0)$. Notice that the endpoints of these segments are preserved when f

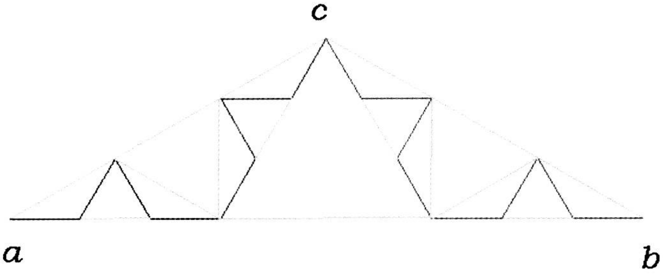


Figure 1.3. The approximating set $A_{(2)}$ for the KOCH curve. The grey lines illustrate the open set condition (Definition 1.1.3.2) needed for the computation of $\dim_S(A)$.

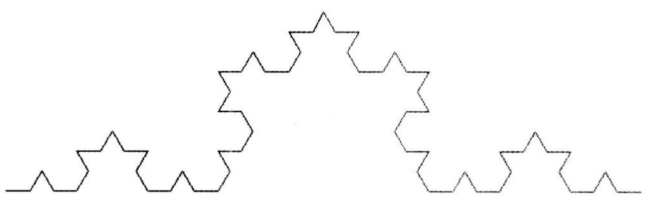


Figure 1.4. The approximating set $A_{(3)}$ for the KOCH curve.

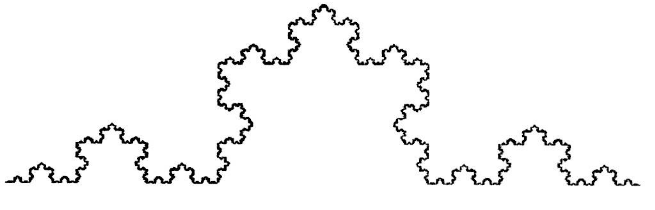


Figure 1.5. A closer approximation ($A_{(7)}$) of the KOCH curve A. The four colours indicate subsets of A which are similar to the whole of A.

is applied to $A_{(k)}$ since they only become endpoints of smaller subsegments. Let us define a map $\phi_k : [0,1] \to A_{(k)}$ in the following way: write every $x \in [0,1]$ in its "4-adic expansion"

$$x = \sum_{i=1}^{\infty} \frac{x_i}{4^i} = \sum_{i=1}^{k} \frac{x_i}{4^i} + r_k(x) \qquad (x_i \in \{0,1,2,3\}, 1 \le i < \infty). \tag{1.2}$$

If we agree to use the finite sum expansion when possible, then $0 \le r_k(x) < \frac{1}{4^k}$. Now define $\phi_k(x)$ to be the point of the segment $A_{x_1,x_2,\dots,x_k}$ lying at distance $r_k(x)$ from the starting point of this segment (in the positive direction). Evidently ϕ_k is a continuous piecewise linear map of $[0,1]$ onto $A_{(k)}$.

What happens to ϕ_k if k increases to ∞? In order to find out about this let us observe what happens to the point $\phi_k(x)$ if x is given as in (1.2) and if we define

$$x_{(m)} := \sum_{i=1}^{m} \frac{x_i}{4^i}.$$

As pointed out above, for the "4-adic rational" part $x_{(m)}$ of x and for all $k \ge m$ we get

$$\phi_k(x_{(m)}) = \phi_m(x_{(m)}) \in A_{x_1,\dots,x_m,0,\dots,0}; \subset A_{(k)} \tag{1.3}$$

(in fact, $\phi_k(x_{(m)})$ is the starting point of this subsegment of $A_{(k)}$). Observing the effect of consecutive applications of f to $A_{x_1,\dots,x_m}$, we find that every such application moves the point $\phi_k(x) \in A_{x_1,\dots,x_m,\dots,x_k}$ to its new position $\phi_{k+1}(x) \in A_{x_1,\dots,x_m,\dots,x_k,x_{k+1}}$ about a distance of at most $\frac{4}{3^{k+1}}$ (four times the length of $A_{x_1,\dots,x_{k+1}}$; a rough estimate since at most $\frac{\sqrt{3}}{2 \cdot 3^{k+1}}$ would do). Adding this up for $k > m$ we get the estimate

$$|\phi_k(x) - \phi_m(x)| \le \sum_{i=m}^{k-1} \frac{4}{3^{i+1}} < \sum_{i=m}^{\infty} \frac{4}{3^{i+1}} = \frac{2}{3^m}. \tag{1.4}$$

As a consequence, we see that the sequence $\{\phi_k(x)\}_{k=1}^{\infty}$ is a CAUCHY sequence (= fundamental sequence) in the plane $\mathbb{R}^2$ and has to converge to a limit point $\phi(x)$. By (1.4) we even see that the functions ϕ_k converge uniformly on $[0,1]$ and that therefore the limiting map ϕ furnishes a continuous curve in $\mathbb{R}^2$. This curve is called the KOCH *curve* [von Koch, 1904].

The fact that endpoints of subsegments $A_{x_1,\dots,x_m}$ of $A_{(m)}$ do not change position under further applications of f, as expressed by (1.3), helps to realize that the KOCH curve is nowhere differentiable. Non-differentiability is readily seen at such an endpoint itself: For any $x \in [0,1[$ consider the point $p_0 = \phi(x_{(m)})$ as defined above (for $x = 1$ the reasoning has to be slightly adapted). For $k > m$ let

$$p_1 := \phi\Big(x_{(m)} + \frac{1}{4^k}\Big), \qquad p_2 := \phi\Big(x_{(m)} + \frac{2}{4^k}\Big).$$

p_1 is the endpoint of the subsegment of $A_{(k)}$ beginning at $\phi(x_m)$, p_2 is the endpoint of the following subsegment of $A_{(k)}$. If $k > m$ is sufficiently large, then the points p_1 and

p_2 are arbitrarily close to p_0, while the secants p_0p_1 and p_0p_2 always include the same positive angle.

It is somewhat more tedious to deal with a point $\phi(x)$ if x is not a 4-adic rational number. Roughly speaking, if ϕ were differentiable in x, then two different points close to $\phi(x)$ (as which will be taken endpoints of subsegments) would have to define a secant close to the tangent in $\phi(x)$, and this will be shown to be impossible. Recall that for complex-valued functions g and h of the argument y one writes $h = o(g)$ as $y \to a$ if $\lim_{y\to a} \frac{h(y)}{g(y)} = 0$, while $h = O(g)$ as $y \to a$ means $\limsup_{y\to a} \frac{|h(y)|}{|g(y)|} < \infty$. Correspondingly $o(1)$ (as $y \to x$) will denote a function which vanishes as $y \to x$, and $O(1)$ will denote a function which remains bounded as $y \to x$. Although not strictly necessary the notation $o_1(1)$, $o_2(1), \ldots$, $O_1(1)$, $O_2(1), \ldots$ will be used to indicate different such functions.

1.1.2.1 Lemma. *Suppose ϕ is differentiable in x, i.e. there exists a vector $q \in \mathbb{R}^2$ such that*

$$\phi(x) - \phi(y) = (q + o(1)) \cdot (x - y) \quad \text{as } y \to x,$$

and let $y_1 \to x$ and $y_2 \to x$ in such a way that

$$\begin{aligned} x - y_1 &= O_1(y_1 - y_2), \\ x - y_2 &= O_2(y_1 - y_2). \end{aligned} \tag{1.5}$$

Then

$$\phi(y_1) - \phi(y_2) = (q + o(1)) \cdot (y_1 - y_2). \tag{1.6}$$

Proof of the lemma.

$$\begin{aligned} \phi(y_1) - \phi(y_2) &= \phi(x) - \phi(y_2) - (\phi(x) - \phi(y_1)) \\ &= \big(q + o_2(1)\big) \cdot (x - y_2) - \big(q + o_1(1)\big) \cdot (x - y_1) \\ &= q \cdot (y_1 - y_2) + o_2(1) \cdot (x - y_2) + o_1(1) \cdot (x - y_1) \\ &= \big[q + o_2(1) \cdot O_2(1) + o_1(1) \cdot O_1(1)\big] \cdot (y_1 - y_2). \end{aligned}$$

□

In order to show that ϕ cannot be differentiable in x let, for arbitrarily large m,

$$y_1 := x_{(m)}, \qquad y_2 := x_{(m)} + \frac{1}{4^{m+1}}, \qquad \overline{y_2} := x_{(m)} + \frac{2}{4^{m+1}}.$$

Then

$$\begin{aligned} |y_1 - \overline{y_2}| &= 2 \cdot |y_1 - y_2| = \frac{2}{4^{m+1}}, \\ |x - y_1| &\le \frac{4}{4^{m+1}} = 4 \cdot |y_1 - y_2|, \\ |x - y_2| &\le \frac{3}{4^{m+1}} = 3 \cdot |y_1 - y_2|, \\ |x - \overline{y_2}| &\le \frac{2}{4^{m+1}} = |y_1 - \overline{y_2}|. \end{aligned}$$

So the requirements (1.5) are satisfied, but not (1.6) since we have already seen that the secants $\phi(y_1)\phi(y_2)$ and $\phi(y_1)\phi(\overline{y_2})$ always include the same small but non-zero angle. Consequently, the function ϕ cannot be differentiable in x.

How do we measure the length of a continuous curve? Take any finite sequence $S = \{p_j\}_{j=0}^n$ of points corresponding to increasing parameter values and compute $l_S := \sum_{j=1}^n |p_j - p_{j-1}|$. The length of the curve is then by definition the supremum over all values l_S obtained in this way. For the KOCH curve it seems convenient to choose $S_k := \{\phi(\frac{j}{4^k})\}_{j=0}^{4^k}$, the endpoints of the 4^k subsegments in $A_{(k)}$. Each of these has length $\frac{1}{3^k}$, therefore we get $l_{S_k} = (\frac{4}{3})^k$. As $k \to \infty$ this also tends to ∞. We conclude that the KOCH curve has infinite length, rather a contrast to the CANTOR set.

One last question (for the time being): what would have happened if we had restricted the action of f to one subsegment A_j ($j \in \{0,1,2,3\}$) of $A_{(0)}$ or, more generally, to a subsegment $A_{j_1,\ldots,j_k}$ of $A_{(k)}$? Obviously we would have got a curve similar to A but reduced to $\frac{1}{3}$, resp. $\frac{1}{3^k}$, in size. In other words, again the KOCH curve is self-similar, it consists of parts which are smaller copies of itself.

1.1.3 Heuristics of dimension

We have not yet pinned down any property of a set in $\mathbb{R}^2$ or $\mathbb{R}^3$ or, more generally, $\mathbb{R}^n$ which might be strange and characteristic enough to make it reasonable to call the set a "fractal". Self-similarity as encountered in the CANTOR set or the KOCH curve seems a possible candidate but there are perfectly harmless sets which also are self-similar, for instance a square in the plane. Shrinking its sides to half the original length again produces a square and the original square consists of four copies thereof – if we allow the sides of the small squares to coincide. In fact, this is intimately connected with the assertion that a full square is a set of dimension 2: reducing the sides to $\frac{1}{n}$ of their original length produces a set, n^2 copies of which (allowing sides to coincide) constitute the original square. Similarly n^3 cubes of side length $\frac{1}{n}$ make up the unit cube – corresponding to its three-dimensionality – and n^1 intervals of length $\frac{1}{n}$ joined together give the one-dimensional unit interval.

If we had not already been familiar with the concept of dimension we could have "computed" the dimension of a square A, say, using the following reasoning: the dimension of A is the exponent d determined by the fact that the set A is a ("almost disjoint", whatever this may mean) union of n^d similar copies of A, reduced in size by the factor $\frac{1}{n}$ (such a similar copy $S(A)$ is congruent with the set $\frac{1}{n}A$, which originates by multiplying every vector in A by the factor $\delta(S) = \frac{1}{n}$). In other words and roughly speaking (to be made more precise in later sections), if $\delta(S) = \frac{1}{n}$ and if A happens to be decomposable into $N_S(A)$ sets of the form $\delta(S)A$, then the *self-similarity dimension* $\dim_S(A)$ may be considered as the solution of the equation

$$\left(\frac{1}{\delta(S)}\right)^{\dim_S(A)} = N_S(A),$$

i.e.

$$\dim_S(A) = \frac{\log N_S(A)}{-\log \delta(S)}. \tag{1.7}$$

Applying this reasoning to the CANTOR set A we recall that it is indeed the disjoint union of two similar copies of itself, reduced by the factor $\delta(S) = \frac{1}{3}$. Formula (1.7) now gives for its dimension

$$\dim_S(A) = \frac{\log 2}{\log 3} \approx 0.63.$$

There is one objection to be dealt with: $\dim_S(A)$ has not been defined in a unique way. If A is the (almost) disjoint union of $N_S(A)$ copies of $\delta(S)A$, then $\delta(S)A$ is the (almost) disjoint union of $N_S(A)$ (almost) disjoint copies of $(\delta(S))^2 A$ and A is the (almost) disjoint union of $N_{S^{(2)}}(A) = (N_S(A))^2$ copies of $\delta(S^{(2)})A = (\delta(S))^2 A$. Should we have been told, before applying formula (1.7), whether to work with S or with $S^{(2)}$, or even with the k-fold iteration $S^{(k)}$ of S? Fortunately this does not matter, since

$$\frac{\log N_{S^{(k)}}(A)}{\delta(S^{(k)})} = \frac{k \log N_S(A)}{k \log \delta(S)} = \frac{\log N_S(A)}{\log \delta(S)}.$$

If there is some doubt left, please be patient until dimension is discussed more thoroughly in Section 1.2 and Section 1.3.

The startling fact is that the dimension of the CANTOR set, with this understanding, is not 1 but less, to wit approximately 0.63 (also different from its topological dimension as a completely disconnected set, which is zero). Looking now at the KOCH curve, formula (1.7) tells us that (if the points in which the subsegments join do not do any damage) its self-similarity dimension is $\frac{\log 4}{\log 3} \approx 1.26$, while of course the topological dimension of each of its approximating sets $A_{(k)}$ is 1.

A theorem (Theorem 1.3.8) to be stated later tells a condition under which this reasoning is applicable, the so-called *open set condition.* Let us first state explicitly what is meant by a similarity.

1.1.3.1 Definition. A map $S : \mathbb{R}^n \to \mathbb{R}^n$ is called a *similarity* with *similarity factor* s if $|S(x) - S(y)| = s \cdot |x - y|$ for some positive number s and for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$.

1.1.3.2 Definition. The similarities S_i $(1 \le i \le k)$ satisfy the *open set condition* if there exists a bounded non-empty open set V with mutually disjoint image sets $S_i(V)$ $(1 \le i \le k)$ satisfying $\bigcup_{i=1}^k S_i(V) \subset V$.

In essence the mentioned theorem states that if the similarities S_i $(1 \le i \le k)$ satisfy the open set condition and if $A = \bigcup_{i=1}^k S_i(A)$, then (1.7) and even a more general formula for the computation of $\dim_S(A = \bigcup_{i=1}^k S_i(A))$ may be applied. The open set condition is obviously satisfied in the case of the CANTOR set: denoting by S_1 and S_2 the similarities mapping the unit interval into its first and last third, as the set V we may take e.g. the open unit interval. It is satisfied also in case of the KOCH curve: let V be the open isosceles triangle with vertices $a = (0,0)$, $b = (1,0)$, $c = (\frac{1}{2}, \frac{1}{2\sqrt{3}})$ (see Figure 1.3). V contains its four images under the similarities mapping the unit interval into the four line segments constituting the set $A_{(1)}$. As a consequence, we may also note that the whole set A is contained in the closure of the triangle V.

Are we now in position to define what is meant by a "fractal"? Yes and no. Yes, since the original definition of MANDELBROT [Mandelbrot, 1982, Section 3] says: A subset of $\mathbb{R}^n$ is a *fractal* (also *fractal set*) if its topological dimension (which is always an integer, zero for the CANTOR set and one for the KOCH curve) is less than its "fractal" dimension (for the CANTOR set and the KOCH curve as computed above). According to this definition a set with a non-integral dimension (as discussed more generally later) is automatically a fractal. No, since it has turned out that there are sets (as the dragon Section 1.1.5.3 to be discussed later) that one would like to consider as fractals but are not included by the just mentioned definition. Up to now it has seemed difficult to find a satisfying definition including all sets which one would like to consider as fractals.

1.1.4 Initiators and generators

There are evidently two ways to produce more general fractals besides the CANTOR set and the KOCH curve: we can start from a set different from the unit interval $A = [0, 1]$ and we can (as we have done already) change the definition of the map f. The first is done by defining a set of segments $A_{(0)}$, then called the *initiator*, upon which the iterates $f^{(k)}$ of f should act. An example is provided as follows.

1.1.4.1 The KOCH island

Let $A_{(0)}$ be the equilateral triangle below the x-axis, one side of which is the unit interval $[0, 1]$. Applying the map f defining the KOCH curve (Section 1.1.2) to the three segments constituting the set $A_{(0)}$ produces a star with six vertices which we may also imagine as an equilateral hexagon, each side of which carries an equilateral triangle of side length $\frac{1}{3}$ (Figure 1.5). The next application of f adds twelve smaller equilateral triangles of side length $\frac{1}{9}$. Continuing this procedure eventually produces the contour of a set looking like a snow flake, consisting of three copies of the KOCH curve we know from Section 1.1.2. The idea of it being surrounded by water leads to calling it the KOCH *island* (Figures 1.6, 1.7).

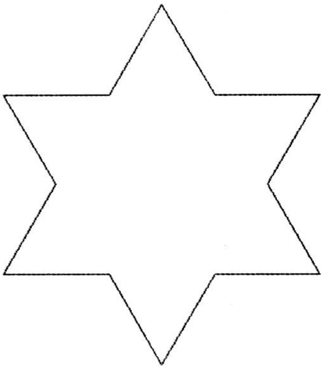


Figure 1.6. The first approximating set $A_{(1)}$ for the KOCH island.

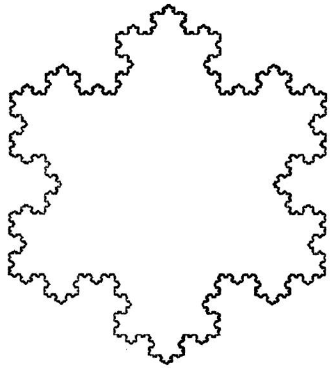


Figure 1.7. A closer approximation ($A_{(7)}$) of the KOCH island.

Still, there is more to this: somebody seeing it for the first time and being asked to estimate the length of the coast line may think: "Well, a little bit more than the perimeter of a circle roughly the same size; taking into account the coasts of the peninsulas and the bays, perhaps twice this perimeter." Asked to look a bit closer and perhaps to use a compass with a rather small opening he may to his surprise find that his measurement of the coast line becomes longer and longer as he decreases this opening, until we disclose to him that already one third of the coast line – our well-known KOCH curve – has infinite length.

It is in this line of thought that MANDELBROT [Mandelbrot, 1982, Section 5] points out that also e.g. the coast line of England, measured with increasing precision, turns out to have infinite length.

Keeping, for the time being, the unit interval as our initiator $A_{(0)}$, we may change the mapping f by requiring that it should act on every segment of any union B of segments by replacing this segment with a – suitably diminished – similar copy of a given union G of segments, called the *generator*. Let us look at several samples of the vast family of fractals obtained in this way.

1.1.4.2 A modified KOCH curve

Suppose the generator G consists of five segments of length $\frac{1}{3}$ each, obtained by replacing the two middle segments of the KOCH curve generator with three sides of a square (Figure 1.8).

We may think of f as employing five similarity maps S_i $(1 \leq i \leq 5)$ each with similarity factor $\frac{1}{3}$. Now the fractal $A = \lim_{k\to\infty} f^{(k)}(A_{(0)})$, defined in essentially the same way as in Section 1.1.2, consists of the union of five similar copies S_iA $(1 \leq i \leq 5)$ joined at the vertices of the original generator $G = A_{(1)}$, but the first two and the last two copies having a lot more points (in fact a whole diagonal segment) in common (Figure 1.9). Still, the open set condition (Definition 1.1.3.2) is satisfied: the open isosceles right-angled triangle D with $A_{(0)}$ as hypotenuse contains the union