FRACTAL STRUCTURES: IMAGE ENCODING AND COMPRESSION TECHNIQUES

V. Drakopoulos

Department of Computer Science and Biomedical Informatics

UNIVERSITY OF THESSALY

Applications

- Biology
- Botany
- Chemistry
- Computer Science (Graphics, Vision, Image Processing and Synthesis)
- Geology
- Mathematics
- Medicine
- Physics

Bibliography

- Barnsley M. F., *Fractals everywhere*, 2nd ed., Academic Press Professional, San Diego, CA, 1993.
- Barnsley M. F., *SuperFractals*, Cambridge University Press, New York, 2006.
- Barnsley M. F. and Anson, L. F., *The fractal transform*, Jones and Bartlett Publishers, Inc, 1993.
- Barnsley M. F. and Hurd L. P., *Fractal image compression*, AK Peters, Wellesley, 1992.
- Barnsley M. F., Saupe D. and Vrscay E. R. (eds.), *Fractals in multimedia*, Springer-Verlag, New York, 2002.
- Falconer K. J., *Fractal geometry: Mathematical foundations and applications*, Wiley, Chichester, 1990.
- Fisher Y., *Fractal image compression* (ed.), Springer-Verlag, New York, 1995.
- Hoggar S. G., *Mathematics for computer graphics*, Cambridge University Press, Cambridge, 1992.
- Lu N., *Fractal imaging*, Academic Press, San Diego, CA, 1997.
- Mandelbrot B. B., *Fractals: Form, chance and dimension*, W. H. Freeman, San Francisco, 1977.
- Mandelbrot B. B., *The fractal geometry of nature*, W. H. Freeman, New York, 1982.
- Massopust P. R., *Fractal functions, fractal surfaces and wavelets*, Academic Press, San Diego, CA, 1994.
- Nikiel S., *Iterated function systems for real-time image synthesis*, Springer-Verlag, London, 2007.
- Peitgen H.--O. and Saupe D. (eds.), *The science of fractal images*, Springer-Verlag, New York, 1988.

Outline

1. Introduction
2. On the dimension
3. Iterated function systems
4. Fractal interpolation
5. Fractal-based image encoding and compression

1. INTRODUCTION

- A brief history
- Classic fractals
- Space-filling curves

Introduction

- Many natural and artificial phenomena
  - have the very fundamental characteristic of invariance under different scales,
  - have infinite details at every point,
  - are self-similar across different scales and
  - can be described by a procedure that specifies a repeated operation for producing the details.
- *Fractal* comes from the Latin adjective *fractus*, which has the same root as *fraction* and *fragment* and means “irregular or fragmented;” it is related to *frangere*, which means “to break.”

The beginning

- Draw a line on a sheet of paper.
- Euclidean geometry tells us that this is a figure of one dimension, namely the length.
- Now extend the line.
- Make it wind around and around, back and forth, without crossing itself, until it fills the entire sheet of paper.
- Euclidean geometry says that this is still a line, a figure of one dimension.
- But our intuition strongly tells us that if the line completely fills the entire plane, it must be two-dimensional.

Genesis

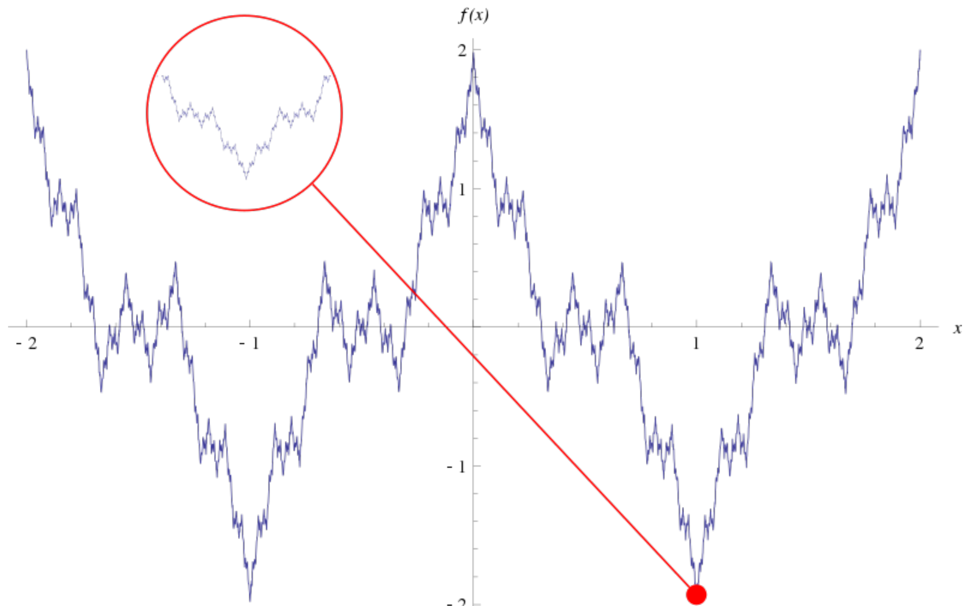
- Such thinking started a revolution in mathematics about a hundred years ago.
- Mathematicians such as Georg Cantor, Giuseppe Peano, David Hilbert, Felix Hausdorff, Helge von Koch and Wacław Sierpiński drew curves that were called “monsters”, “psychotic” and “pathological” by traditional mathematicians.
- A new type of dimensioning was proposed, in which a curve could have a fractional dimension, not just an integer one.

Classic fractals

- The Weierstrass function (1872).
- The (triadic, middle-third) Cantor set (discovered in 1874 by Henry John Stephen Smith and introduced by German mathematician Georg Cantor in 1883).
- Plane (space) filling curves
  - The Peano curve (1890)
  - The Hilbert curve (1891)
- The Koch curve (1904)
- The Sierpiński gasket (1915)
- The Sierpiński carpet (1916)
- The Sierpiński tetrahedron
- The Menger sponge (1926)
- The Mandelbrot set
- Spleenwort fern
- Natural phenomena
  - Terrain, coastline, clouds, water, trees, feathers, fur

Weierstrass function (1872)

- The Weierstrass function is an example of a pathological real-valued function on the real line.
- The function has the property of being continuous everywhere but differentiable nowhere.
- It is named after its discoverer Karl Weierstrass.



Plot of Weierstrass function over the interval [−2, 2]. Like some fractals, the function exhibits self-similarity: every zoom (red circle) is similar to the global plot.

Cantor set (1883)

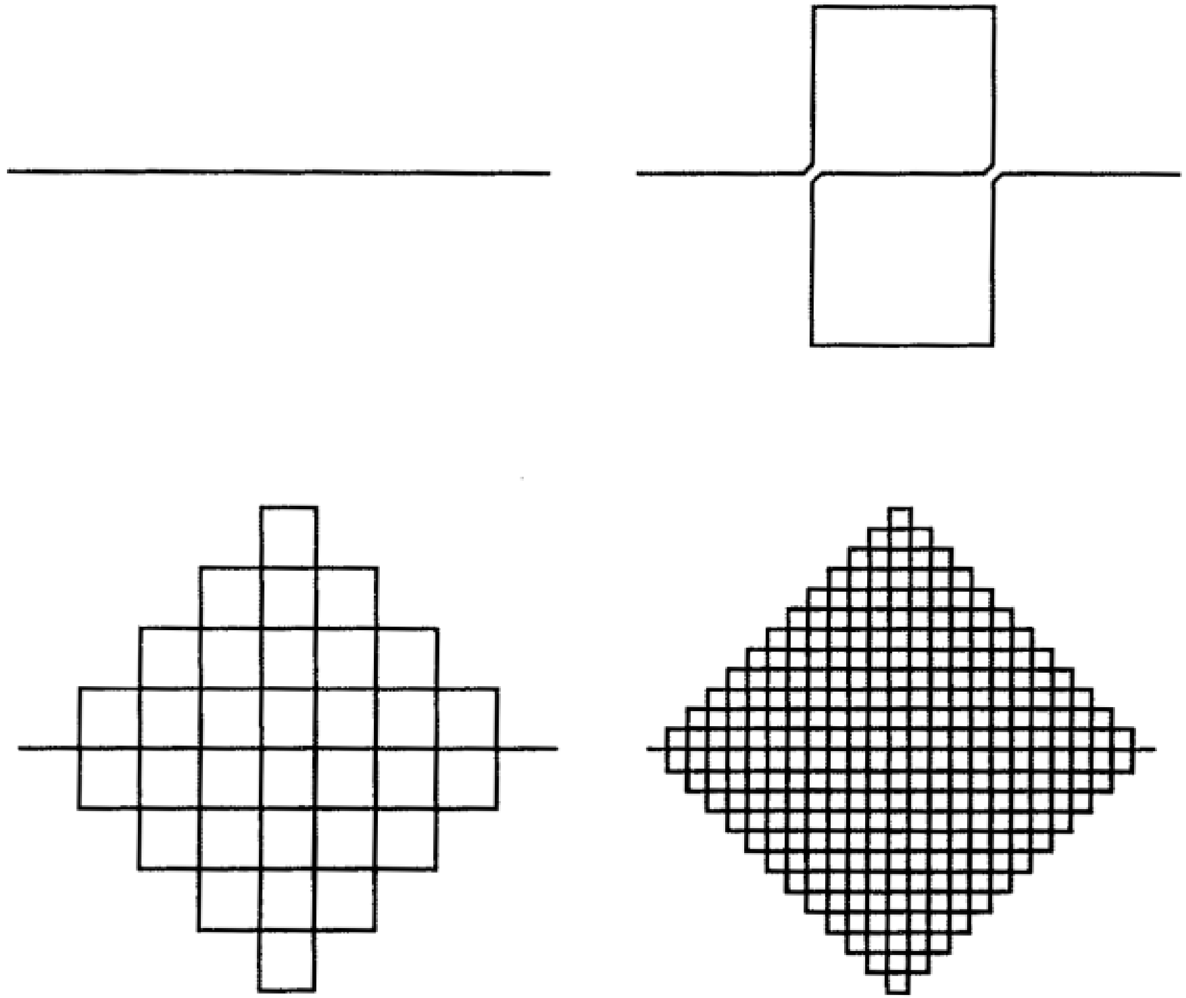
- Initially, we consider the closed set $c_0 = [0, 1]$.
- Remove from c_0 its middle third. What remains is the set $c_1 = [0, 1/3] \cup [2/3, 1]$.
- Remove the middle third of [0, 1/3] and [2/3, 1].
- Continuing this ad infinitum, we get the Cantor set

$$C = \bigcap_{n=0}^{\infty} c_n .$$

A brief history

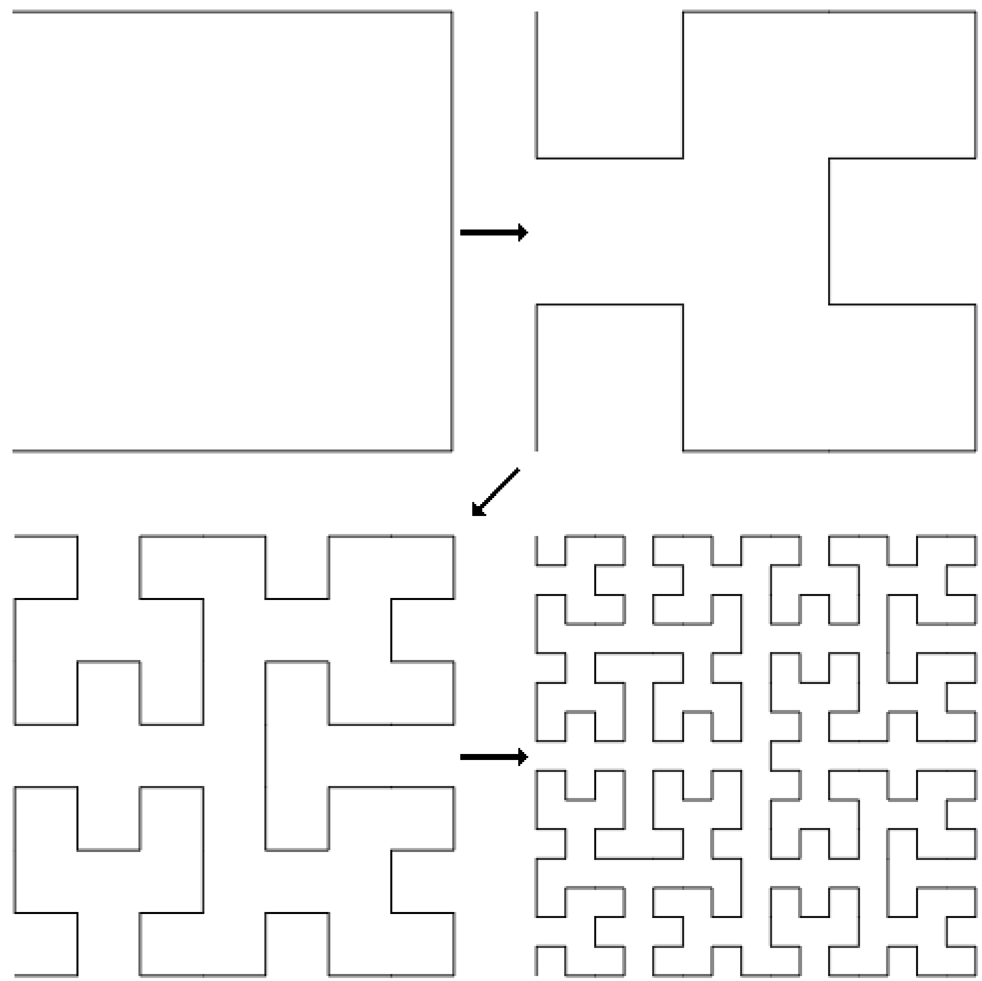
- In 1890, Giuseppe Peano discovered a densely self-intersecting curve that passes through every point of the unit square.
- His purpose was to construct a continuous mapping from the unit interval onto the unit square.
- He was motivated by Georg Cantor's earlier counterintuitive result that the infinite number of points in a unit interval is the same cardinality as the infinite number of points in any finite-dimensional manifold, such as the unit square.
- The problem Peano solved was whether such a mapping could be continuous; i.e., a curve that fills a space.

The Peano curve (1890)



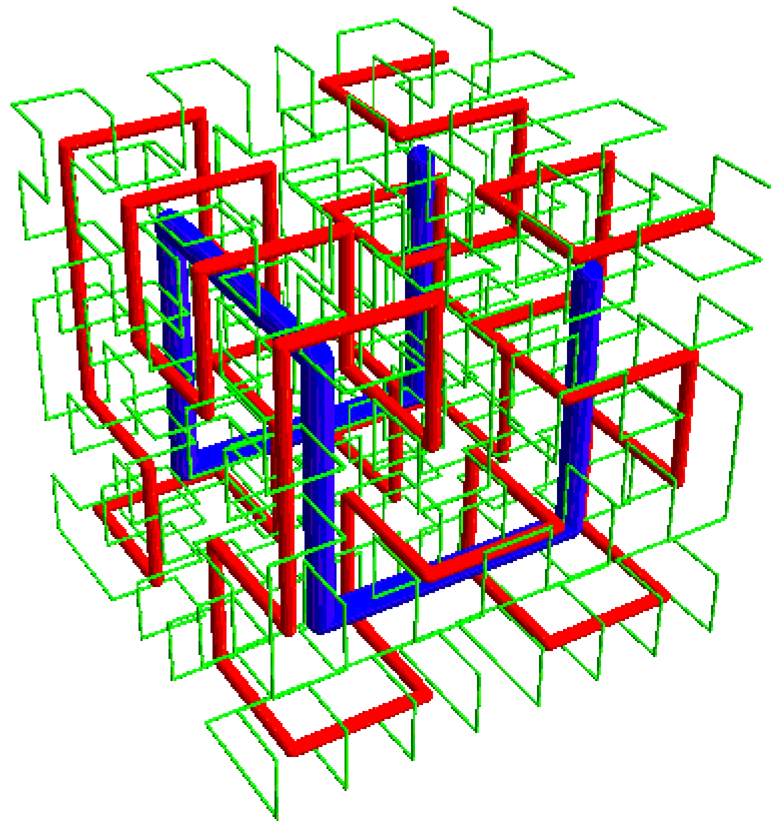
The Hilbert curve in 2D

- A continuous fractal space-filling curve first described by the German mathematician David Hilbert in 1891 as a variant of the space-filling curves discovered by Giuseppe Peano in 1890.
- The first four iterations are shown on the right.

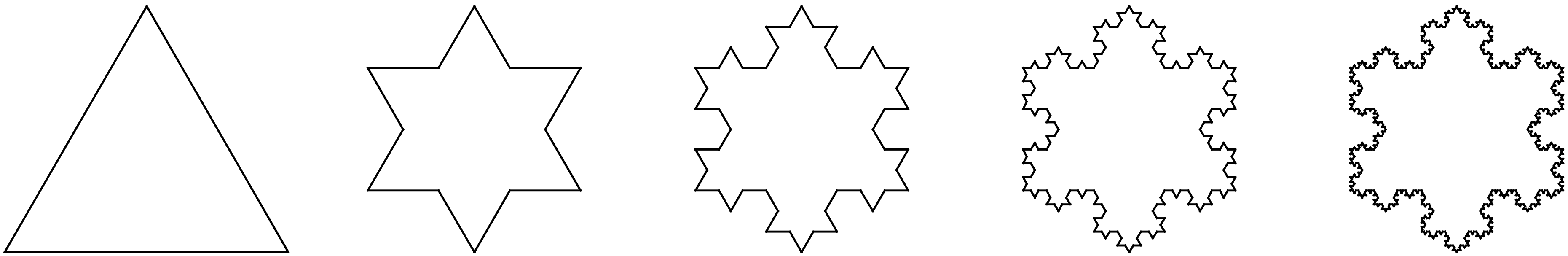


The Hilbert curve in 3D

- The Hilbert curve as well as the Moore curve are two famous plane-filling curves that can be extended to 3D space-filling curves.
- A three-dimensional analog of the Hilbert curve is shown on the right.



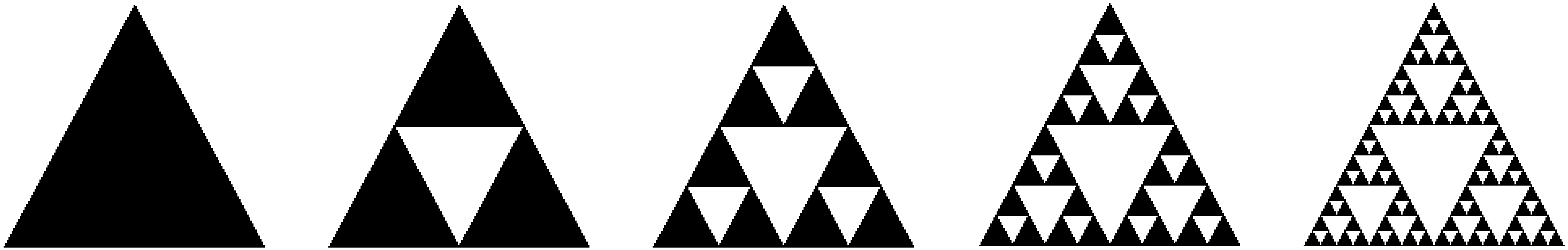
Koch snowflake (1904)



The first five iterations of the Koch snowflake

- The Koch curve can be constructed by starting with an equilateral triangle, then recursively altering each line segment as follows:
  - Divide the line segment into three segments of equal length.
  - Draw an equilateral triangle that has the middle segment from step 1 as its base and points outward.
  - Remove the line segment that is the base of the triangle from step 2.
- After one iteration of this process, the result is a shape similar to the Star of David.
- The Koch curve is the limit approached as the above steps are followed over and over again.

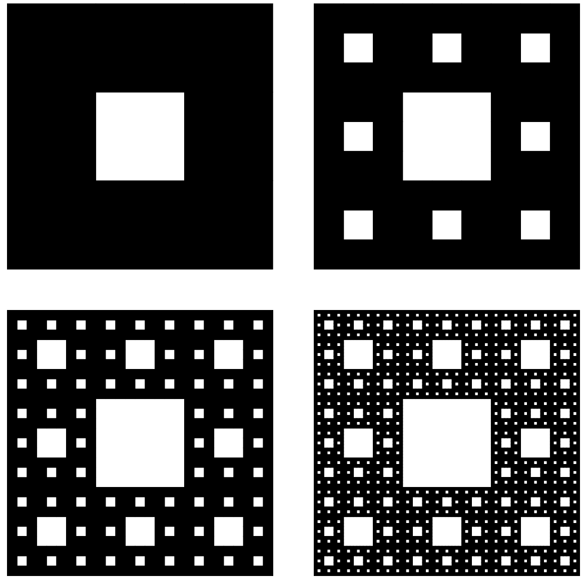
Sierpiński gasket (1915)



- Start with a solid (filled) equilateral triangle $S(0)$.
- Divide this into four smaller equilateral triangles using the midpoints of the three sides of the original triangle as the new vertices and remove the interior of the middle triangle to get $S(1)$.
- Repeat this procedure on each of the three remaining solid equilateral triangles to obtain $S(2)$ and continuing we get

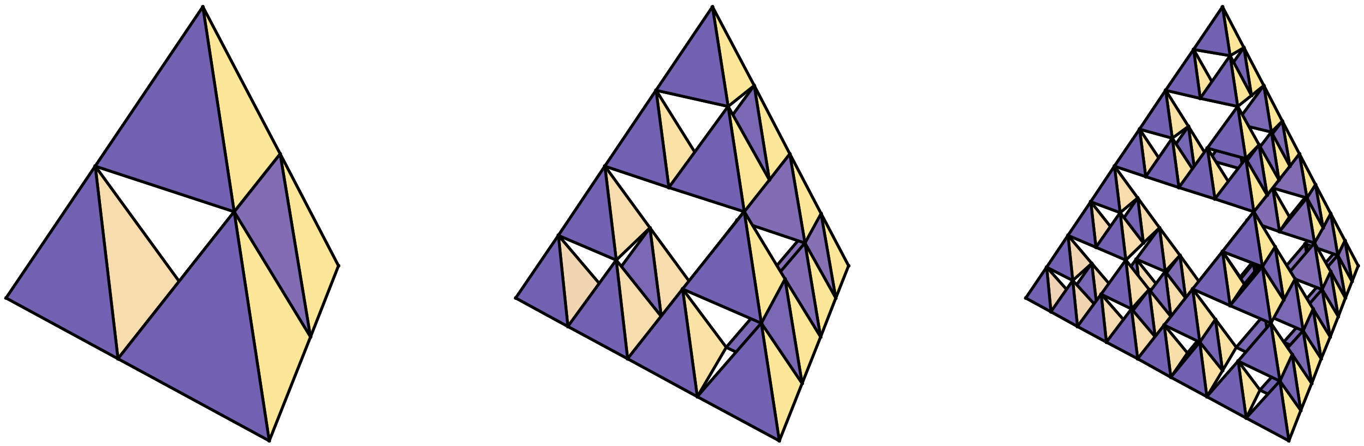
$$S = \bigcap_{i=1}^{\infty} S(i).$$

Sierpiński carpet (1916)



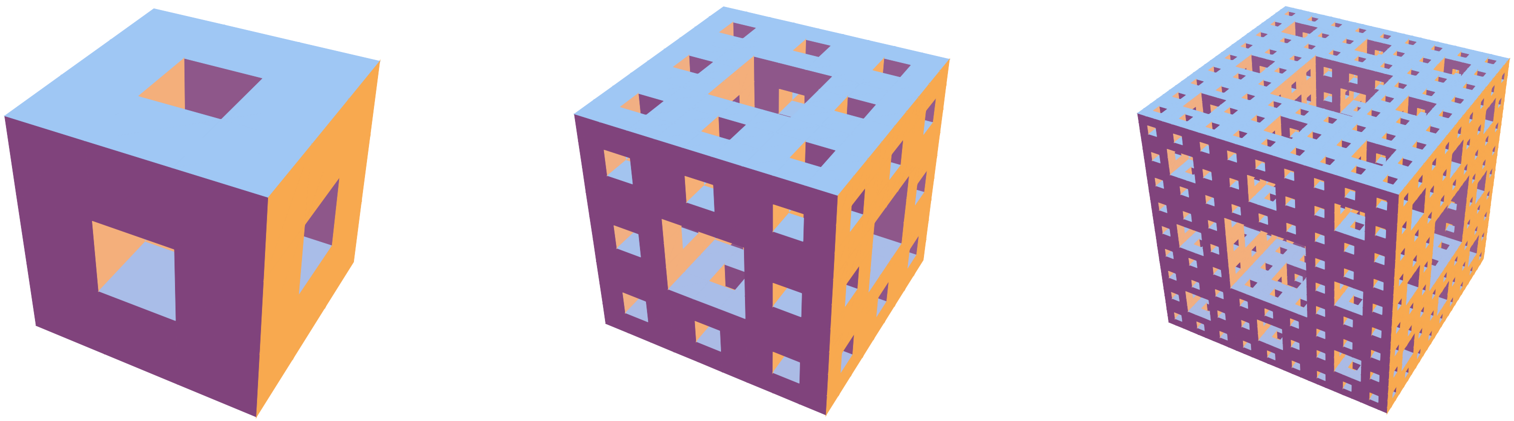
- Begin with a square.
- The square is cut into 9 congruent subsquares in a 3-by-3 grid, and the central subsquare is removed.
- The same procedure is then applied recursively to the remaining 8 subsquares, *ad infinitum*.

Sierpiński tetrahedron



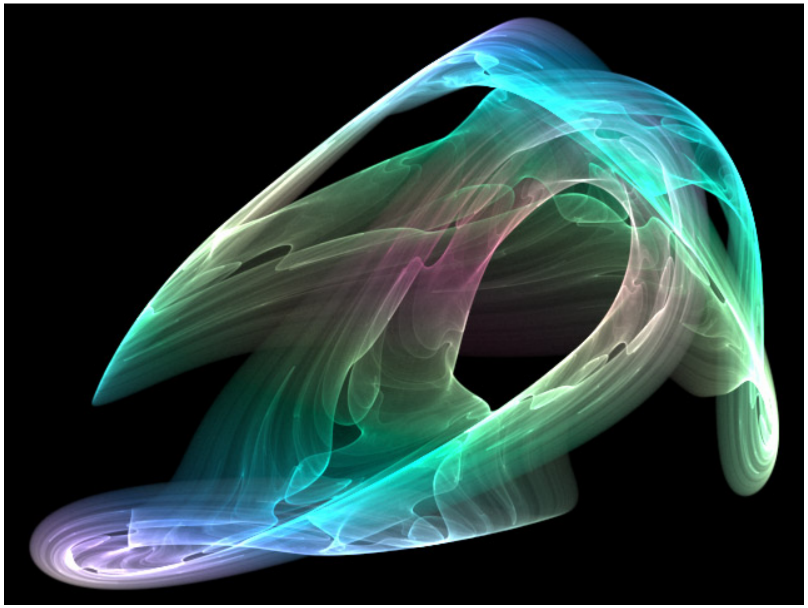
- The tetrix is the three-dimensional analogue of the Sierpiński triangle, formed by repeatedly shrinking a regular tetrahedron to one half its original height, putting together four copies of this tetrahedron with corners touching, and then repeating the process.
- This can also be done with a square pyramid and five copies instead.

Menger sponge (1926)

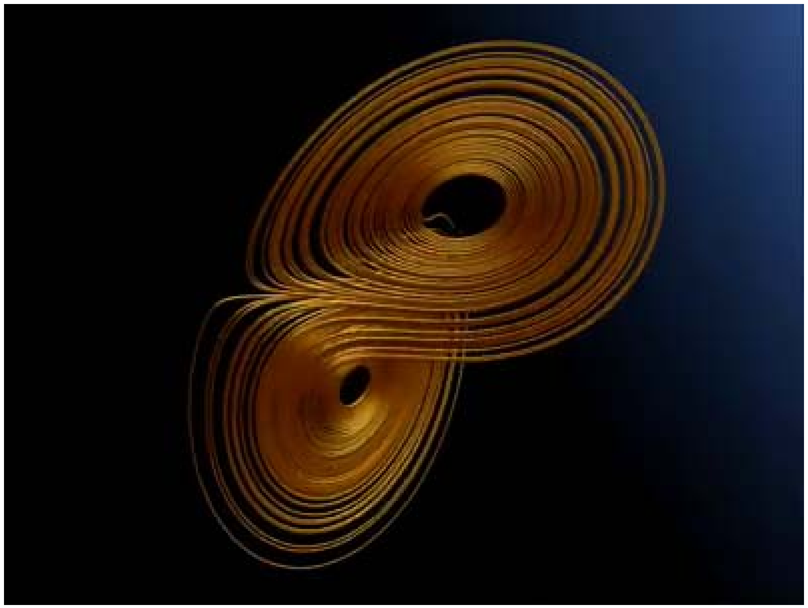


- Begin with a cube. (*first image*)
- Divide every face of the cube into 9 squares, like a Rubik's Cube. This will subdivide the cube into 27 smaller cubes.
- Remove the cube at the middle of every face and remove the cube in the center, leaving 20 cubes, resembling a Void Cube. (*second image*). This is a level-1 Menger sponge.
- Repeat steps 1–3 for each of the remaining smaller cubes.

Strange attractors



Visual representation of a strange attractor



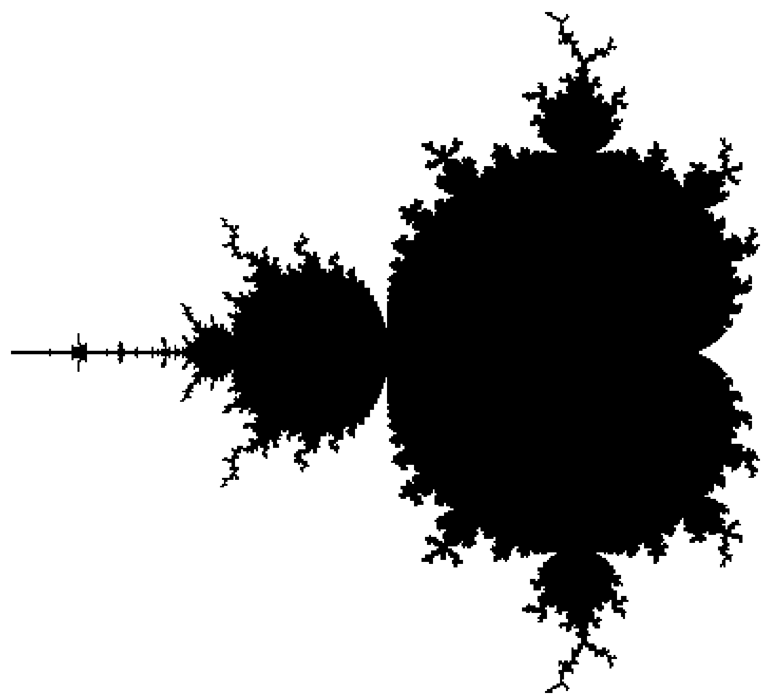
Lorenz attractor

Barnsley's fern

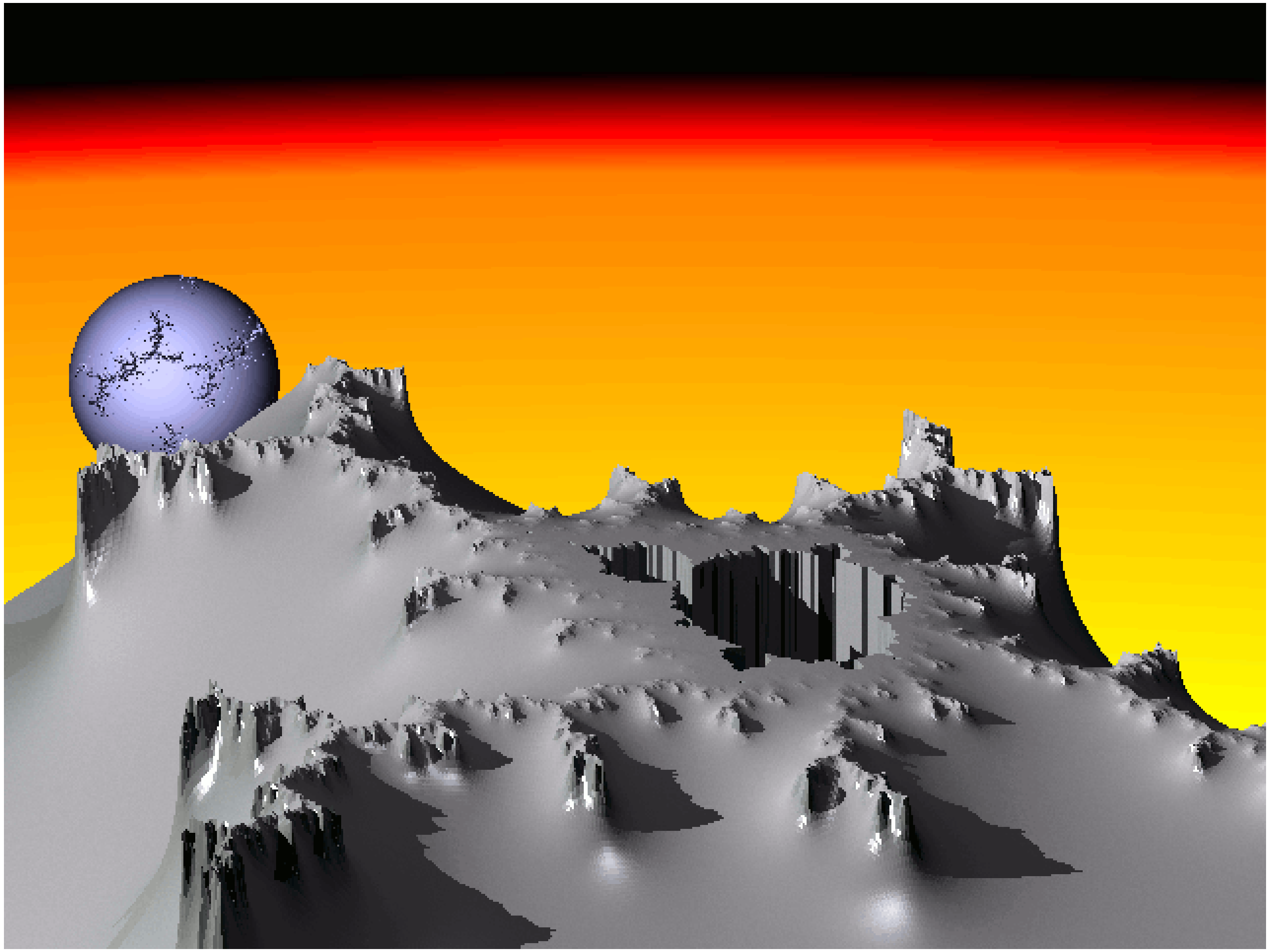
- It is a fractal named after the British mathematician Michael Barnsley who first described it in his book *Fractals Everywhere*.
- He made it to resemble the Black Spleenwort, *Asplenium adiantum-nigrum*.

The Mandelbrot set

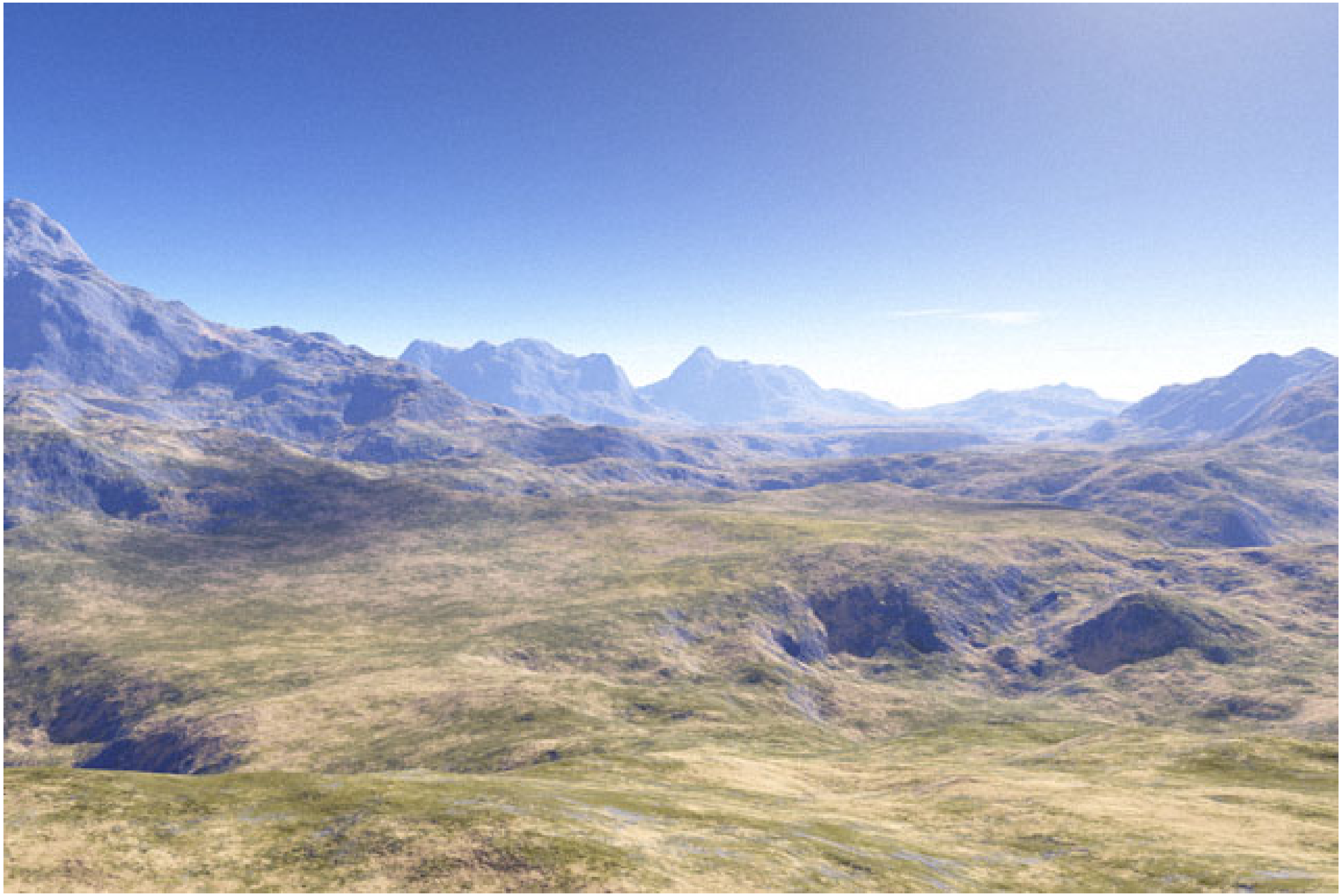
- The set of complex values c that do not diverge under the squaring transform $p(z) = z^2 + c$ beginning with $z = 0$.



Rendering of electrostatic potential



Landscapes



2. ON THE DIMENSION

- General concept
- Metric space
- Self-similarity

The concept

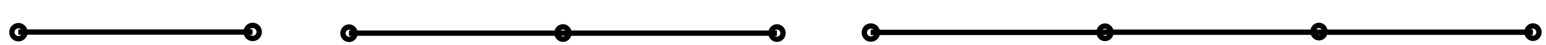
- The dimension of a space or object is informally defined as the minimum number of coordinates needed to specify each point within it.
- A line has a dimension of one because only one coordinate is needed to specify a point on it.
- A surface such as a plane or the surface of a cylinder or sphere has a dimension of two because two coordinates are needed to specify a point on it.
- The inside of a cube, a cylinder or a sphere is three-dimensional because three coordinates are needed to locate a point within these spaces.
- As one would expect, the (topological) dimension is always a natural number.

Inductive dimension

- Consider a discrete set of points (such as a finite collection of points) to be 0-dimensional.
- By dragging a 0-dimensional object in some direction, one obtains a 1-dimensional object.
- By dragging a 1-dimensional object in a new direction, one obtains a 2-dimensional object.
- In general one obtains an (n + 1)-dimensional object by dragging an n-dimensional object in a *new direction*.
- The inductive dimension of a topological space may refer to the small inductive dimension or the large inductive dimension, and is based on the analogy that (n + 1)-dimensional balls have n-dimensional boundaries, permitting an inductive definition based on the dimension of the boundaries of open sets.

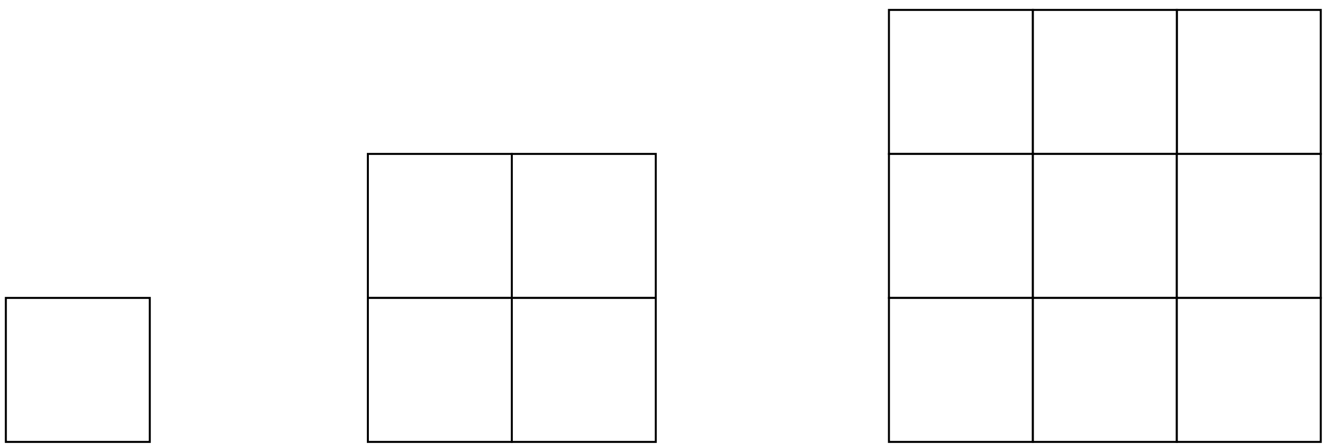
Dimension 1

- Consider a line segment.
- Blow up the segment by a factor of two. The segment is now twice (2^1) as long as before.
- Blowing up the segment by a factor of three, the segment is now three (3^1) times as long as before.



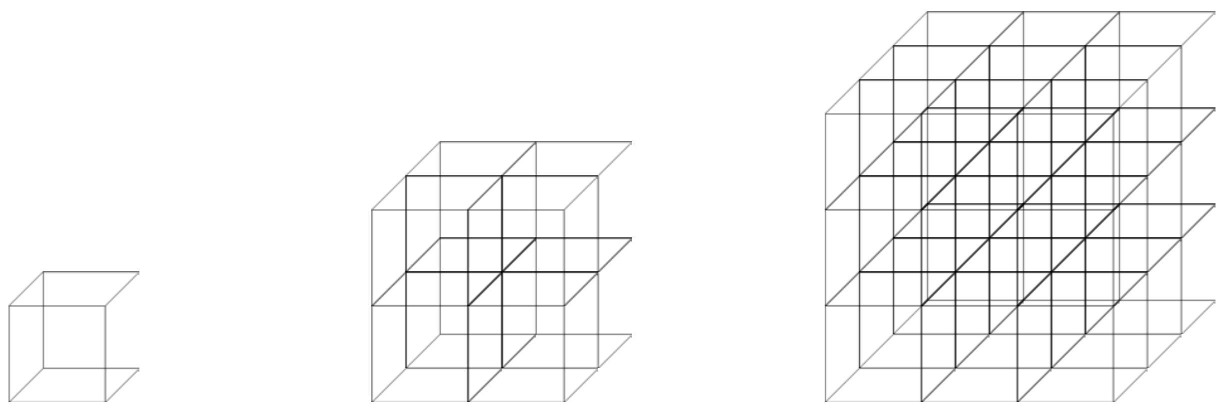
Dimension 2

- Consider a square.
- Blow up the square by a factor of two. The square is now $4 = 2^2$ times as large as before (i.e. 4 original squares can be placed on the original square).
- Blowing up the square by a factor of three, the square is now $9 = 3^2$ times as large as before.



Dimension 3

- Consider a cube.
- Blow up the cube by a factor of two. The cube is now $8 = 2^3$ times as large as before (i.e. 8 original cubes can be placed on the original cube).
- Blowing up the cube by a factor of three, the cube is now $27 = 3^3$ times as large as before.



Metric space

A non-empty set V becomes a metric space when supplied with a mapping (metric) of the form $\rho: V \times V \to \mathbb{R}$ given by the formula $(x, y) \mapsto \rho(x, y)$ which for each $x, y, z \in V$ has the properties:

(M1) $\rho(x, y) \geq 0$, (non-negativity)

and

$\rho(x, y) = 0 \Leftrightarrow x = y$ (identity)

(M2) $\rho(x, y) = \rho(y, x)$ (symmetry)

(M3) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ (triangle inequality)

The members of V are frequently called ‘points’ and the non-negative, real number $\rho(x, y)$ the ‘distance’ from the ‘point’ x to the ‘point’ y.

Examples

- The set $\mathbb{R}$ of all real numbers with the usual metric $\rho(x, y) = |x - y|$ for all $x, y \in \mathbb{R}$ is a metric space, which is called a *real line*.
- The most important space for us is the familiar n-dimensional Euclidean space $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}, i = 1, 2, \ldots, n\}$ with the *Pythagorean* or *root mean square error metric* defined by

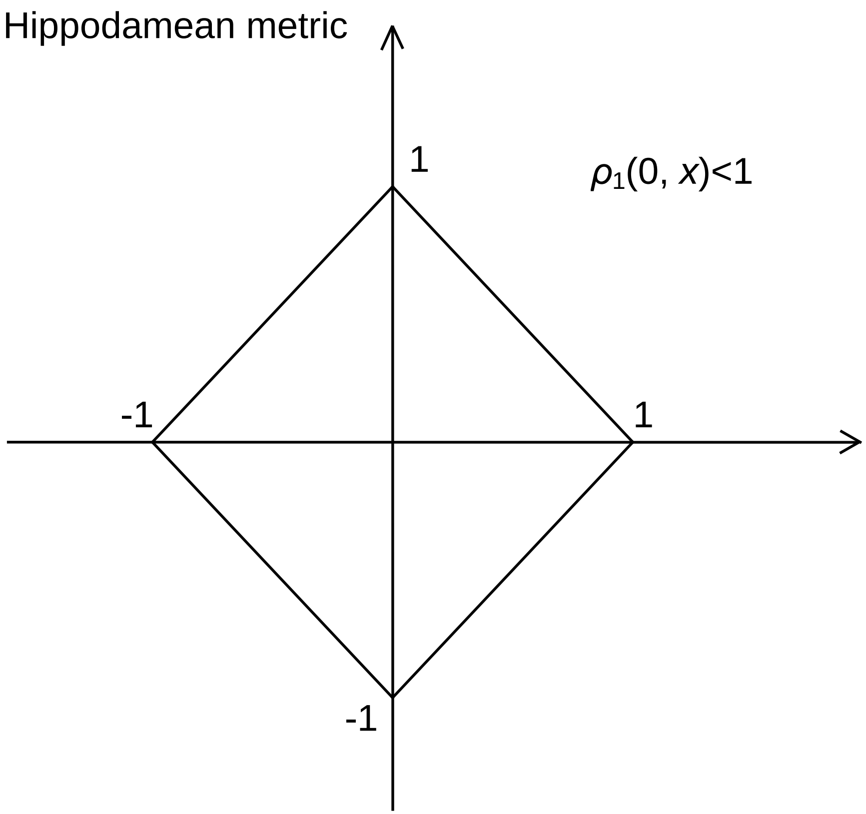
$$\rho_2(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

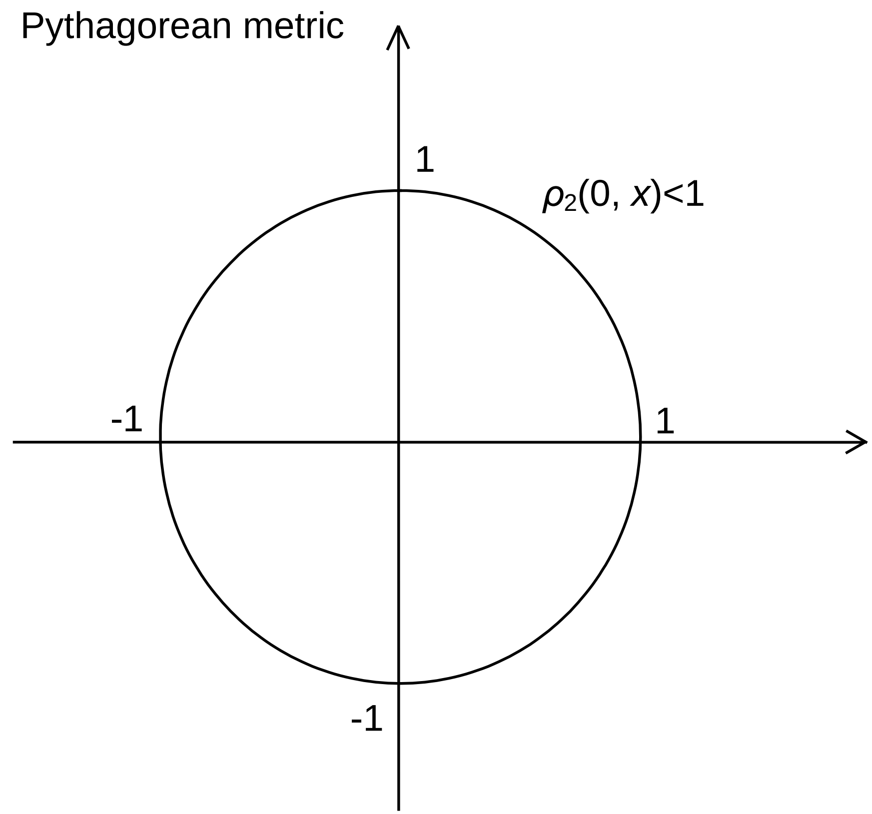
or with the *Hippodamean metric*

$$\rho_1(x, y) = \sum_i |x_i - y_i|$$

where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, $x_i, y_i \in \mathbb{R}$, sometimes called the *box* or *city-block* metric.

The locus


Hippodamean metric
1
$\rho_1(0, x)<1$
-1
1
-1


Pythagorean metric
1
$\rho_2(0, x)<1$
-1
1
-1

Similarities

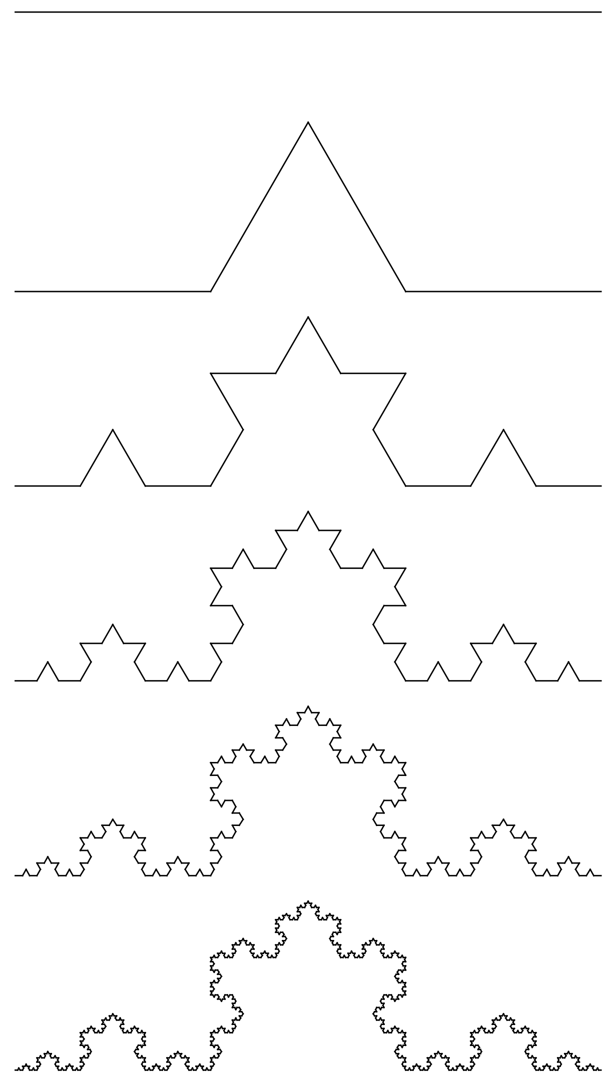
- Two geometrical objects are called similar if one can be obtained from the other by uniformly scaling (enlarging or reducing), possibly with additional translation, rotation and reflection.
- In geometry two triangles, $\triangle ABC$ and $\triangle A'B'C'$, are similar if and only if corresponding angles have the same measure: this implies that they are similar if and only if the lengths of corresponding sides are proportional.
- A mapping $f: X \to Y$, where (X, ρ) and (Y, σ) are metric spaces is a similarity or similitude of ratio or scale r, if $\sigma(f(x), f(y)) = r\,\rho(x, y)$ for every $x, y \in X$ and some fixed $r \in \mathbb{R}_+$.
- When $r = 1$ a similarity is called an isometry (rigid motion).

Self-similarity

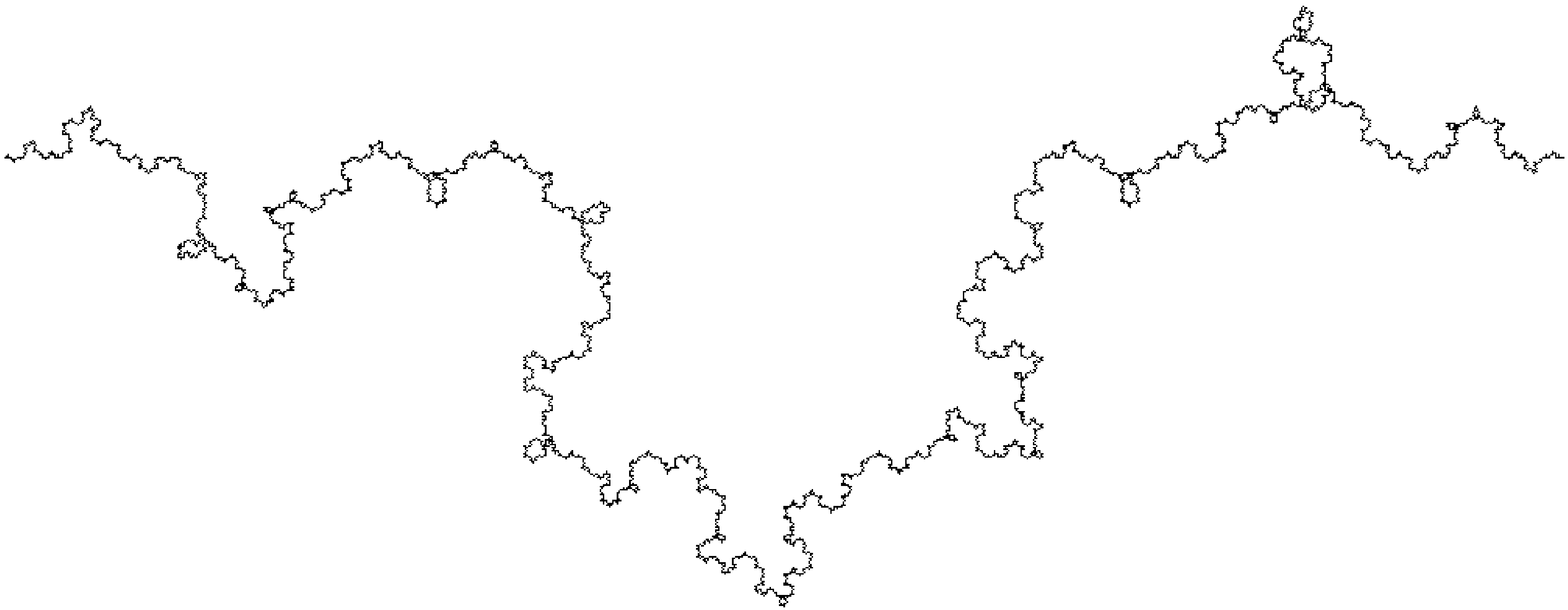
- A self-similar object is exactly or approximately similar to a part of itself (i.e. the whole has the same shape as one or more of the parts).
- Many objects in the real world, such as coastlines, are statistically self-similar: parts of them show the same statistical properties at many scales.
- Self-similarity is a typical property of fractals.
- Scale invariance is an exact form of self-similarity where at any magnification there is a smaller piece of the object that is similar to the whole.

Koch curve

- The single line segment in Step 0 is broken into four equal-length segments in Step 1.
- This same “rule” is applied an infinite number of times resulting in a figure with an infinite perimeter.
- The first five stages are shown on the right.



Randomly placed generator



Dimension 1.2619...

- Consider a Koch curve, where each of the 4 new lines is 1/3 the length of the old line.
- Blowing up the Koch curve by a factor of 3 results in a curve 4 times as large (one of the old curves can be placed on each of the 4 segments)
- Therefore, $4 = 3^d$ or

$$d = \frac{\ln 4}{\ln 3}.$$

Similarity dimension

- A set F is called self-similar, if

$$F = w_1(F) \cup w_2(F) \cup \cdots \cup w_i(F),$$

  where w_i are similitudes with common similarity ratio r and the sets $w_i(F)$ do not overlap.
- For a self-similar shape F made of N copies of itself, each scaled by a similarity with contraction factor r, the similarity dimension is

$$\dim_s F = \frac{\log(N)}{\log(1/r)}.$$

Examples

- Cantor set
  $N = 2$, $r = 1/3$, $\dim_s \mathcal{C} = \log 2/\log 3 = 0{,}630929...$
- Koch snowflake
  $N = 4$, $r = 1/3$, $\dim_s K = \log 4/\log 3 = 1{,}261859507\ldots$
- Sierpiński gasket
  $N = 3$, $r = 1/2$, $\dim_s S = \log 3/\log 2 = 1{,}584962500\ldots$
- Sierpiński carpet
  $N = 8$, $r = 1/3$, $\dim_s C = \log 8/\log 3 = 1{,}892789260\ldots$
- The Peano curve
  $N = 9$, $r = 1/3$, $\dim_s S = \log 9/\log 3 = 2$
- Sierpiński tetrahedron
  $N = 4$, $r = 1/2$, $\dim_s T = \log 4/\log 2 = 2$
- Menger sponge
  $N = 20$, $r = 1/3$, $\dim_s M = \log 20/\log 3 = 2{,}726833028\ldots$

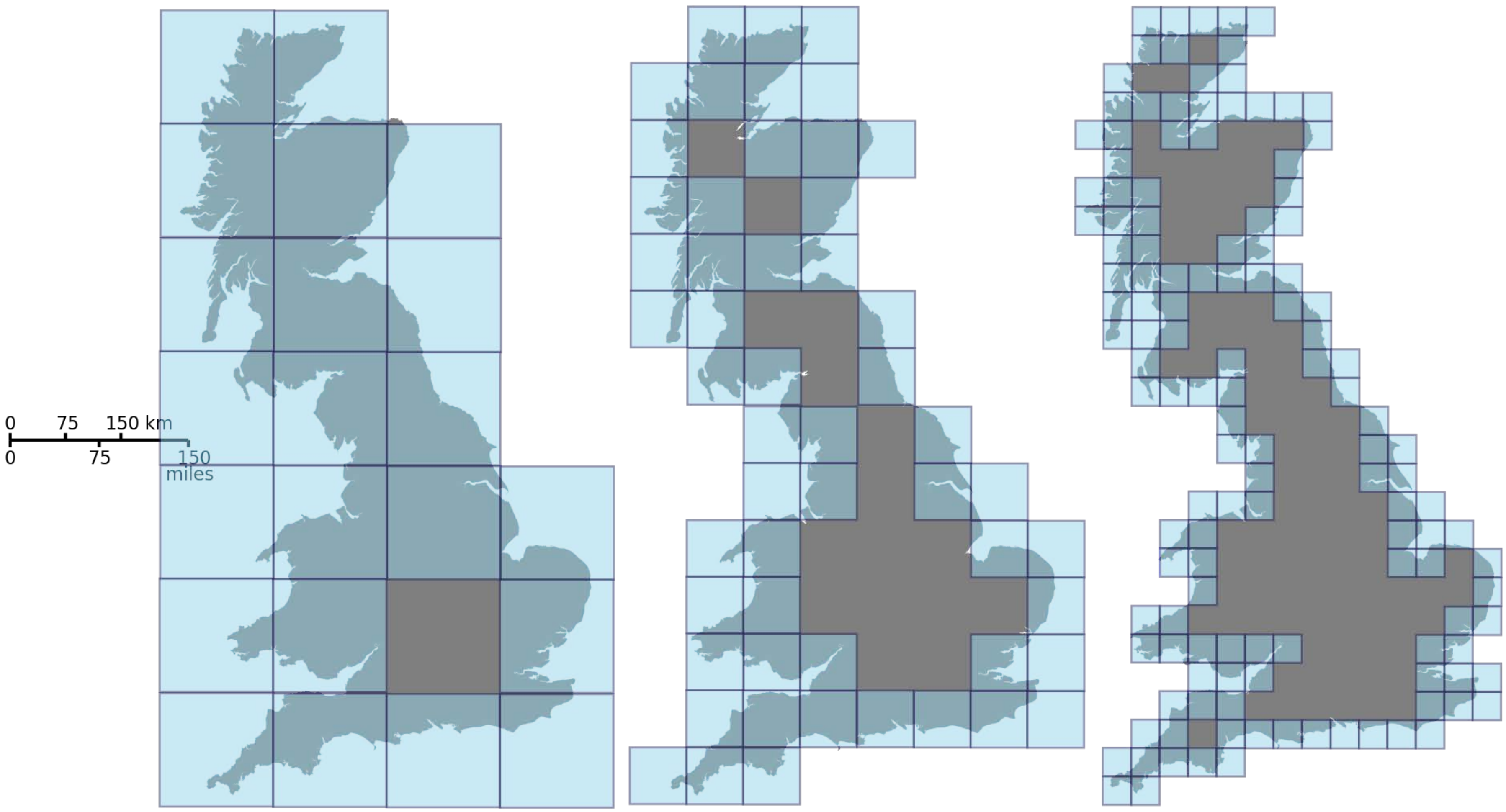
Box-counting dimension

- Let A be a set in a metric space.
- For each $\varepsilon > 0$, let $N(A, \varepsilon)$ denote the smallest number of closed balls of radius $\varepsilon > 0$ needed to cover A.
- If

$$D = \lim_{\varepsilon \to 0} \left\{ \frac{\ln(1/N(A,\varepsilon))}{\ln(\varepsilon)} \right\}$$

exists, then D is the box-counting dimension of A.

Estimating the box-counting dimension of the coast of Great Britain



Hausdorff-Besicovitch dimension

- Let *X* be a metric space. If $S \subset X$ and $d \in [0, +\infty)$, the *d*-dimensional Hausdorff content of *S* is defined by

$$C_H^d(S) = \inf\left\{\sum_i r_i^d : \text{there is a cover of } S \text{ by balls with radii } r_i > 0\right\}.$$

- In other words, $C_H^d(S)$ is the infimum of the set of numbers $\delta \geq 0$ such that there is some (indexed) collection of balls $\{B(x_i, r_i) : i \in I\}$ covering *S* with $r_i > 0$ for each $i \in I$ which satisfies

$$\sum_{i \in I} r_i^d > \delta.$$

- The Hausdorff dimension of *S* is defined by

$$\dim_H(S) = \sup\left\{d \geq 0 : C_H^d(S) = \infty\right\} = \inf\left\{d \geq 0 : C_H^d(S) = 0\right\}.$$

Examples

- Let F be a flat disc of unit radius in $\mathbb{R}^3$.
- From familiar properties of length, area and volume $C_H^1(F)$ = length (F) = ∞, $0 < C_H^2(F)$ = (4/π) × area (F) = 4 < ∞ and $C_H^3(F)$ = (6/π) × vol(F) = 0.
- Thus, $\dim_H F = 2$, with $C_H^d(S) = \infty$ if $d < 2$ and $C_H^d(S) = 0$ if $d > 2$.

Physical meaning

- Amount of variation in the object details
- A measure of roughness (fragmentation) of an object
- The concept was introduced in 1918 by the mathematician Felix Hausdorff.
- Many of the technical developments used to compute the Hausdorff dimension for highly irregular sets were obtained by Abram Samoilovitch Besicovitch.

What is a fractal?

- A fractal is by definition a set whose Hausdorff-Besicovitch dimension strictly exceeds its topological dimension.
- Since the dimension 1.2619 is greater than the dimension 1 of the lines making up the Koch curve, the curve is a fractal.

3. ITERATED FUNCTION SYSTEMS

- Preliminaries
- Distances between sets
- Transformations

The space where fractals live

- Let (X, ρ) be a metric space. Then, $\mathcal{H}(X)$ denotes the space whose points are the compact subsets of X, other than the empty set, i.e.

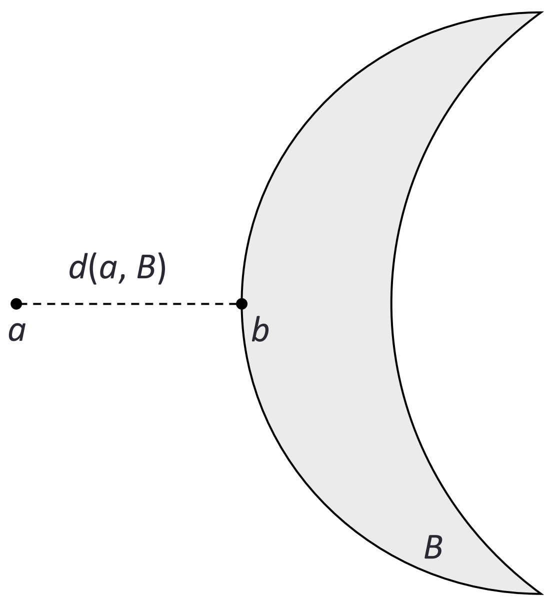
$$\mathcal{H}(X) = \{\varnothing \neq A \subset X : A \text{ is compact}\}.$$

- Sometimes $\mathcal{H}(X)$ is referred to as the 'space of fractals in X' (but note that not all members of $\mathcal{H}(X)$ are fractals).
- The difference between a subset of $\mathcal{H}(X)$ and a nonempty, compact subset of X is that $\mathcal{H}(X)$ is a set of sets, so every subset of it is a set of compact sets.

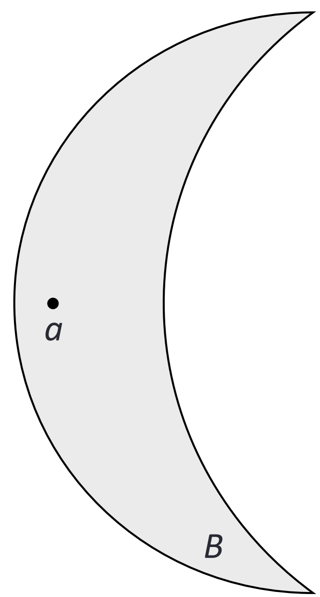
Distance between a point and a set

- The subset of real numbers $\{\rho(x, y) : y \in B\}$, where $x \in X$ and $B \in \mathcal{H}(X)$ has a smallest value.
- Then, as the distance of the point x from the subset B we consider

$$\min\{\rho(x, y) : y \in B\}.$$



$d(a, B) = \rho(a, b)$



$d(a, B) = 0$

Distances between sets

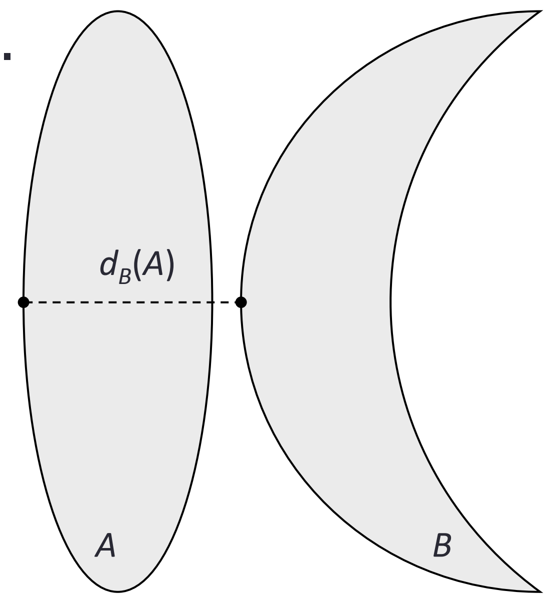
- Let A and B be two nonempty, compact subsets of a metric space (X, ρ). We define as

$$d_A(B) = \max\{d(x, A) : x \in B\}$$

and

$$d_B(A) = \max\{d(x, B) : x \in A\}.$$

- The function $d_B(A)$ is usually called the directed Hausdorff distance from A to B.

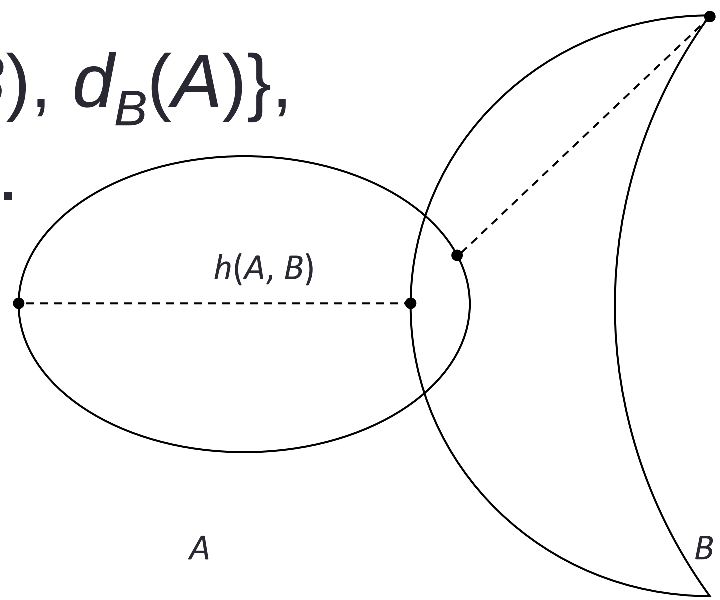


The Hausdorff metric

- It measures how far two subsets of a metric space are from each other.
- It turns the set of nonempty, compact subsets of a metric space into a metric space in its own right.
- If

$$h(A, B) = \max\{d_A(B), d_B(A)\},$$

then $(\mathcal{H}(X), h)$ is a metric space.



Iterated function

- In mathematics, an *iterated function* is a function which is composed with itself, possibly ad infinitum, in a process called iteration.
- *Iteration* means the act of repeating a process with the aim of approaching a desired goal, target or result.
- The formal definition of an iterated function on a set X follows.

Dynamic system

- Define f^k as the k-th iterate of f, where k is a non-negative integer, by $f^0 = \mathrm{id}_X$ and $f^{k+1} = f \circ f^k$, where id_X is the identity function on X and $f \circ g$ denotes function composition.
- Let S be a subset of $\mathbb{R}^n$ and let f: $S \to S$ be a continuous mapping. An iterative scheme $\{f^k\}$ is called a discrete dynamic system.
- We are interested in the behaviour of the sequence of iterates, or orbits, $\{f^k(x)\}$ for various initial points $x \in S$, particularly for large k.

A fixed point theorem

Let $f: X \to X$ be a continuous mapping, where (X, ρ) is a compact metric space. Then there exists a nonempty, closed set $A \subset X$ such that

$$f(A) = A.$$

Contraction mapping

- A contraction mapping, or contraction, on a metric space (X, ρ) is a function f from X to itself, with the property that there is a nonnegative real number $s < 1$ such that for all x and y in X,

$$\rho(f(x), f(y)) \leq s \cdot \rho(x, y).$$

- The smallest such value of s is called the Lipschitz constant of f.
- Contractive maps are sometimes called Lipschitzian maps.
- If the above condition is satisfied for $s \leq 1$, then the mapping is said to be non-expansive.

Banach fixed point theorem

- Also known as the contraction mapping theorem or contraction mapping principle.
- Let (X, ρ) be a nonempty, complete metric space. Let $T: X \to X$ be a contraction mapping on X.
- Then the map T admits one and only one fixed point x^* in X (this means $T(x^*) = x^*$).
- Furthermore, this fixed point can be found as follows: Start with an arbitrary element x_0 in X and define an iterative sequence by $x_n = T(x_{n-1})$ for $n = 1, 2, 3, \ldots$. This sequence converges and its limit is x^*.

The attractor

- We shall call a subset F of S an attractor for f if F is a closed set that is invariant under f (i.e. $f(F) = F$) such that the distance from $f^k(x)$ to F converges to zero as k tends to infinity for all x in an open set V containing F.
- The set V is called the basin of attraction of F.

Iterated Function Systems (IFS's)

A (hyperbolic) *Iterated Function System* (IFS) on the metric space $(\mathbb{R}^n, ||\cdot||)$ is defined as a pair $\{\mathbb{R}^n; w_{1-M}\}$, where

$$\{w_i : \mathbb{R}^n \to \mathbb{R}^n, i = 1, 2, \ldots, M\}$$

is a finite set of contractions with *contractivity factors* s_i, i.e. for every $i = 1, 2, \ldots, M$

$$||w_i(\mathbf{x}) - w_i(\mathbf{y})|| \leq s_i \, ||\mathbf{x} - \mathbf{y}|| \; \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

for some $0 \leq s_i < 1$.

Hutchinson operator

- A collection of functions on an underlying space *X*.
- Formally, let $\{\mathbb{R}^n; w_{1-M}\}$ be an IFS, or a set of *M* contractions from a compact set *X* into itself. We may regard this as defining an operator *H* on the power set 2^X as

$$H : A \mapsto \bigcup_{i=1}^{M} w_i(A),$$

  where *A* is any subset of *X*.
- The iteration of these functions gives rise to the attractor of an iterated function system, for which the fixed set is self-similar.

The attractor of an IFS

- The attractor of a (hyperbolic) IFS is the unique set

$$\mathcal{A}_\infty = \lim_{k\to\infty} H^k(A_0)$$

for every starting set A_0, where

$$H(A) = \bigcup_{i=1}^{M} w_i(A)$$

for all $A \in \mathcal{H}(\mathbb{R}^n)$.

- The map H is also called the collage map to alert us to the fact that $H(A)$ is formed as a union or 'collage' of sets.

Affine transformations

A transformation *w* is affine, if it may be represented by a matrix **A** and translation **t** as **w**(**x**) = **Ax** + **t**, or, if $X = \mathbb{R}^2$,

$$w\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & s \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d \\ e \end{pmatrix}$$

whereas if $X = \mathbb{R}^3$

$$w\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & g \\ h & k & s \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} l \\ m \\ r \end{pmatrix}.$$

Example

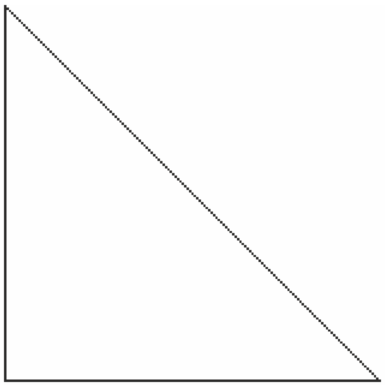
Consider an IFS of the form $\{\mathbb{R}^2; w_1, w_2, w_3\}$, where

$$w_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

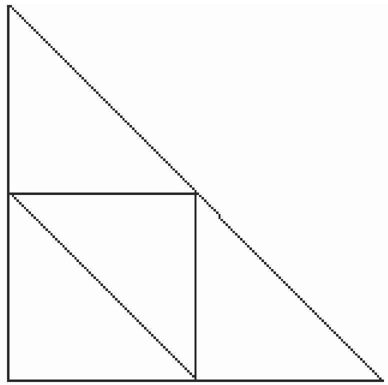
$$w_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$$

$$w_3\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

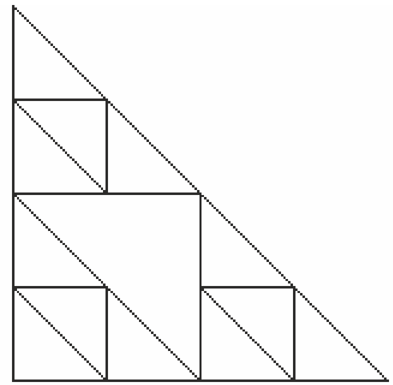
Photocopy machine 1



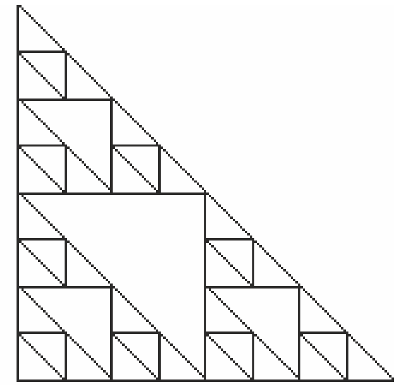
A_0



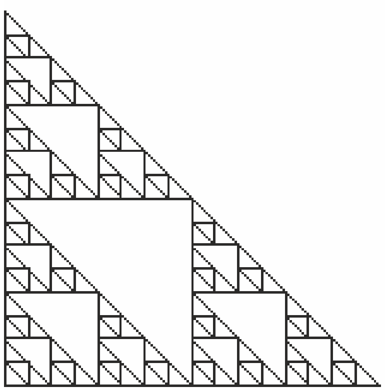
$W(A_0)$



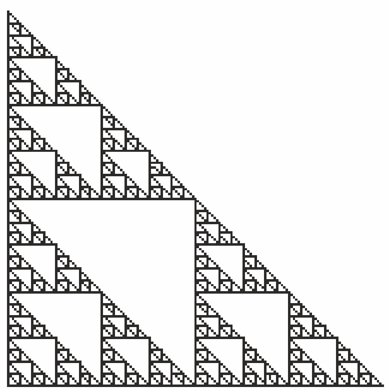
$W^2(A_0)$



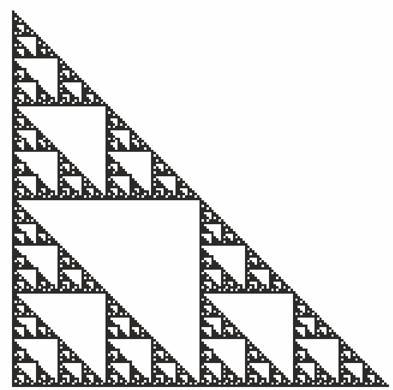
$W^3(A_0)$



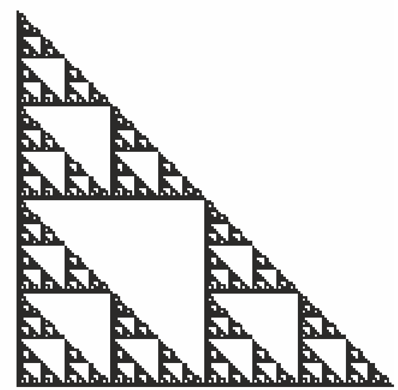
$W^4(A_0)$



$W^5(A_0)$

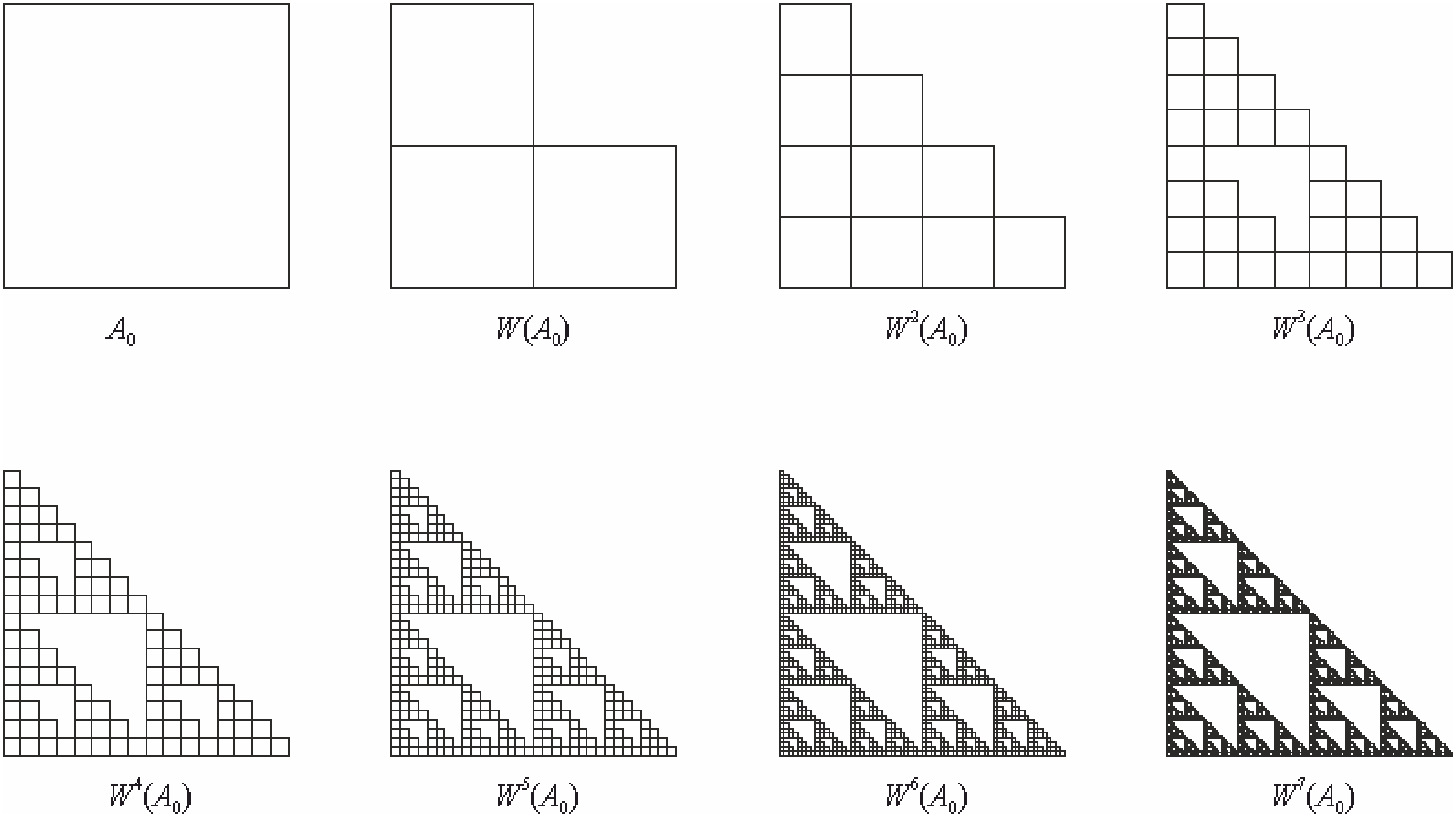


$W^6(A_0)$

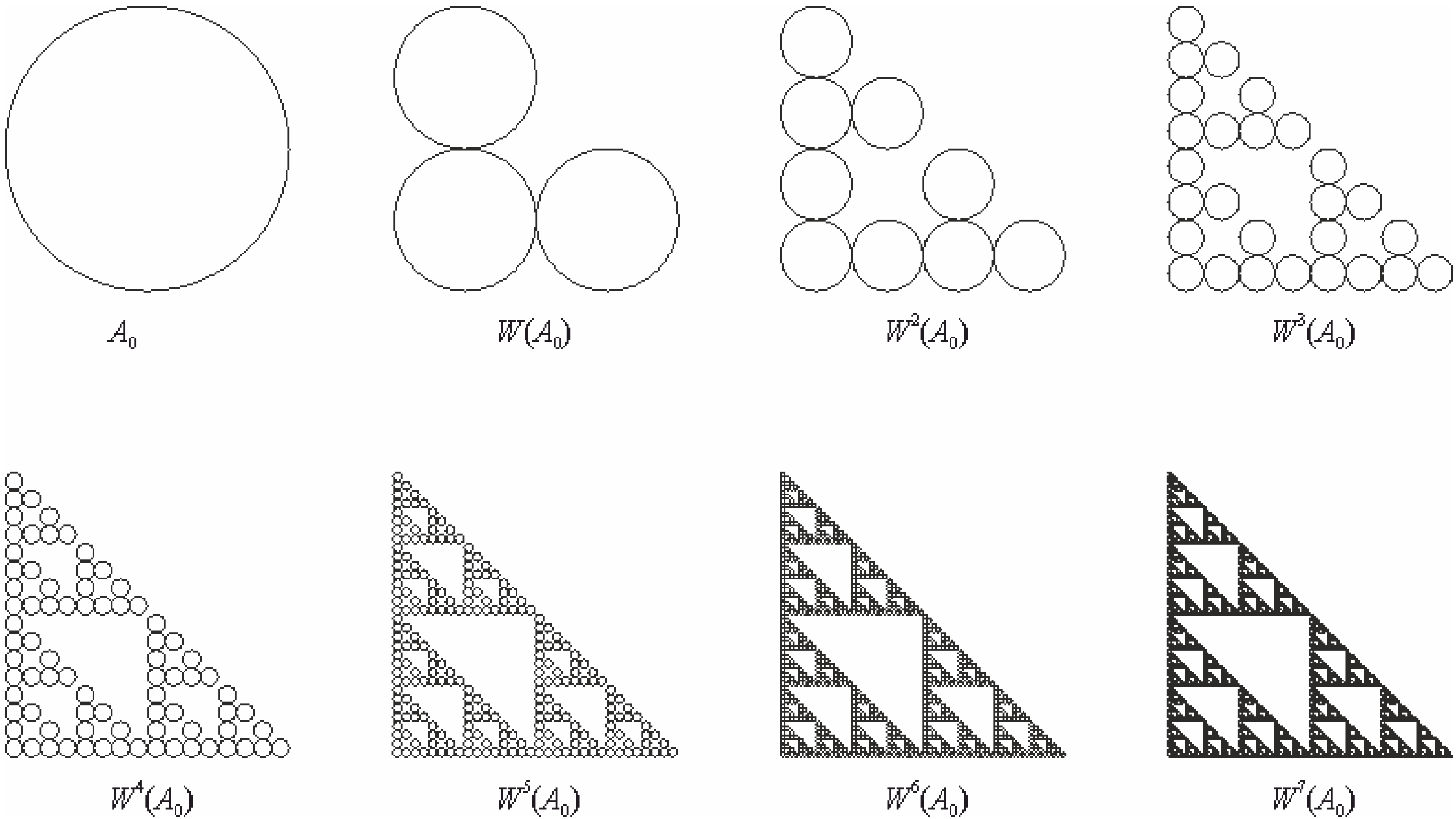


$W^7(A_0)$

Photocopy machine 2



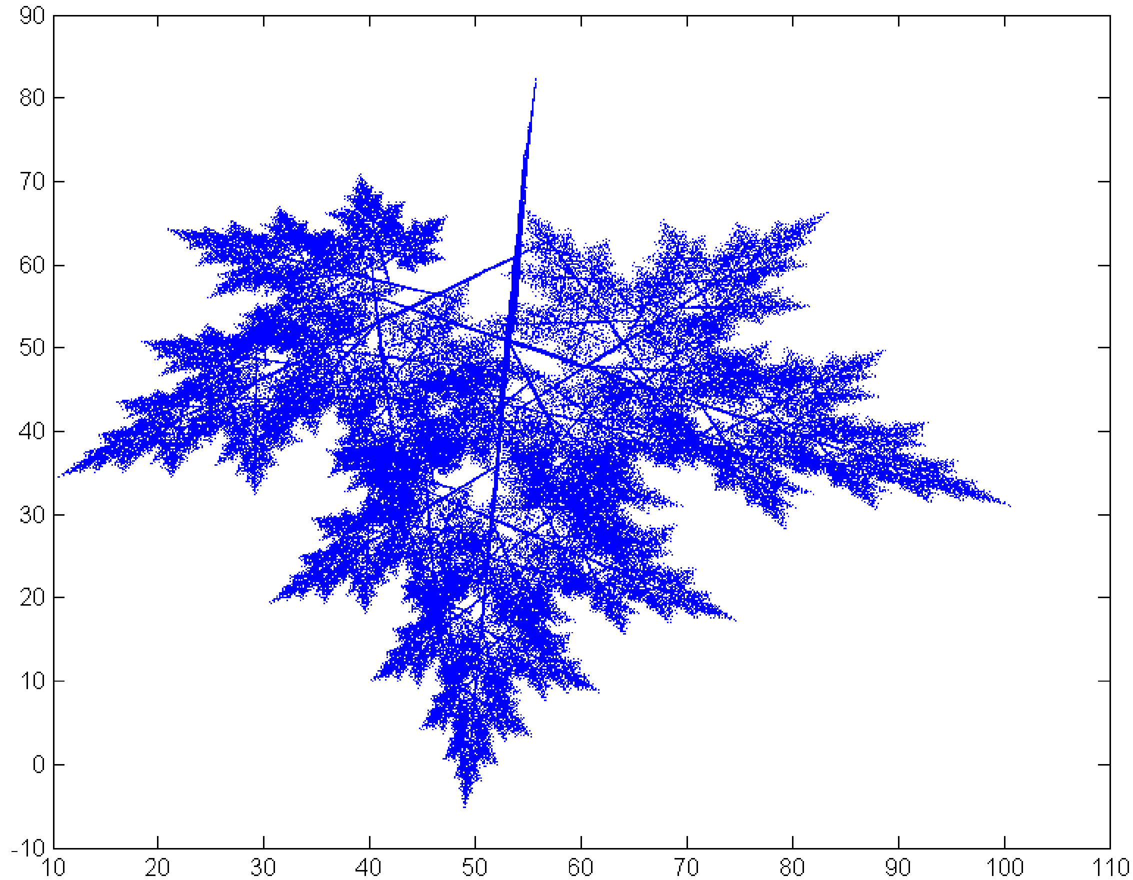
Photocopy machine 3



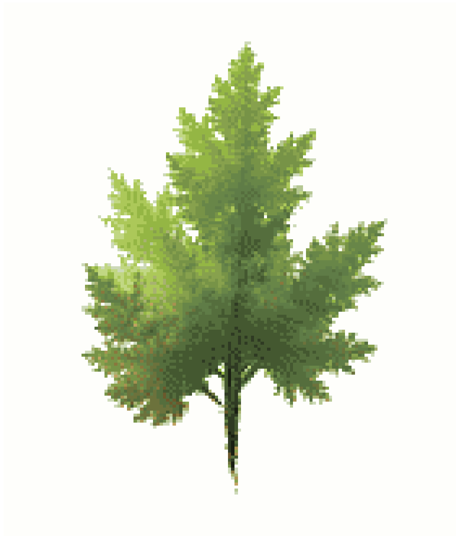
Recurrent IFSs

- An IFS with probabilities, written formally as $\{X; w_1, w_2, \ldots, w_M; p_1, p_2, \ldots, p_M\}$ or, somewhat more briefly, as $\{X; w_{1-M}; p_{1-M}\}$, gives to each transformation in H a probability or weight.
- If the weights of transformations differ, so do the measures on different parts of the attractor.
- A non-self-similar attractor, however, is more easily represented with a recurrent iterated function system, or RIFS for short.
- Each transformation has, instead of a single weight for the next iteration, a vector of weights for each transformation, $\{X; w_{1-M}; p_{i,j} \in [0, 1]; i, j = 1, 2, \ldots, M\}$, so that the matrix of weights is a recurrent Markov operator for the Hutchinson operator's transformation.

Complexity



The geometry of nature



Classification

- Self-similar fractals
  - Have parts that are scaled-down versions of the entire object
  - Can use different scaling factors for different parts
  - *Statistically self-similar*, if random variations are applied
  - Commonly used to model trees, shrubs, plants
- Self-affine fractals
  - Have parts that are formed with different scaling parameters (*sx*, *sy and sz*) in different coordinate directions.
  - *Statistically self-affine*, if random variations are used
  - Commonly used to model terrain, water and clouds
- Invariant fractal sets
  - Formed with nonlinear transformations
  - *Self-squaring fractals,* e.g. the Mandelbrot set
  - *Self-inverse fractals*

4. FRACTAL INTERPOLATION

- Introduction
- Functions
- Surfaces

Why interpolation functions?

- Euclidean geometry and elementary functions are the basis of the traditional methods for analyzing experimental data
- These functions can be expressed by simple mathematical formulas
- They can be stored in small files and computed by fast algorithms

Why fractal?

- Integral dimension
- Suitable for the design of man-made objects (e.g. circles, squares)

- Non-integral dimension
- Suitable for the design of natural objects (e.g. clouds, mountain ranges)
- Better fitting to experimental data (e.g. EEG, ECG, seismograph, image compression)

An example…

A fractal interpolation function.

Interpolation functions in $\mathbb{R}$

- Let the continuous function f be defined on a real closed interval $I = [x_0,x_M]$ and with range the metric space $(\mathbb{R}, |\cdot|)$, where

$$x_0 < x_1 < \cdots < x_M.$$

  It is not assumed that these points are equidistant.

- The function f is called an interpolation function corresponding to the generalised set of data

$$\{(x_m, y_m) \in K = I\times\mathbb{R} : m = 0,1,\ldots,M\},$$

  if $f(x_m) = y_m$ for all $m = 0,1,\ldots,M$ and $K = I\times\mathbb{R}$.

- The points $(x_m, y_m) \in \mathbb{R}^2$ are called the interpolation points. We say that the function f *interpolates* the data and that (the graph of) f passes through the interpolation points.

Affine fractal interpolation

Let us represent our, real valued, set of data points as

$$\{(u_n, v_n) : n = 0, 1, \ldots, N;\ u_n < u_{n+1}\}$$

and the interpolation points as

$$\{(x_m, y_m) : m = 0, 1, \ldots, M;\ M \le N\},$$

where u_n is the sampled index and v_n the value of the given point in u_n.

The affine fractal interpolation function (AFIF) is constructed with M affine mappings of the form

$$w_i\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_i & 0 \\ c_i & s_i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d_i \\ e_i \end{pmatrix}$$

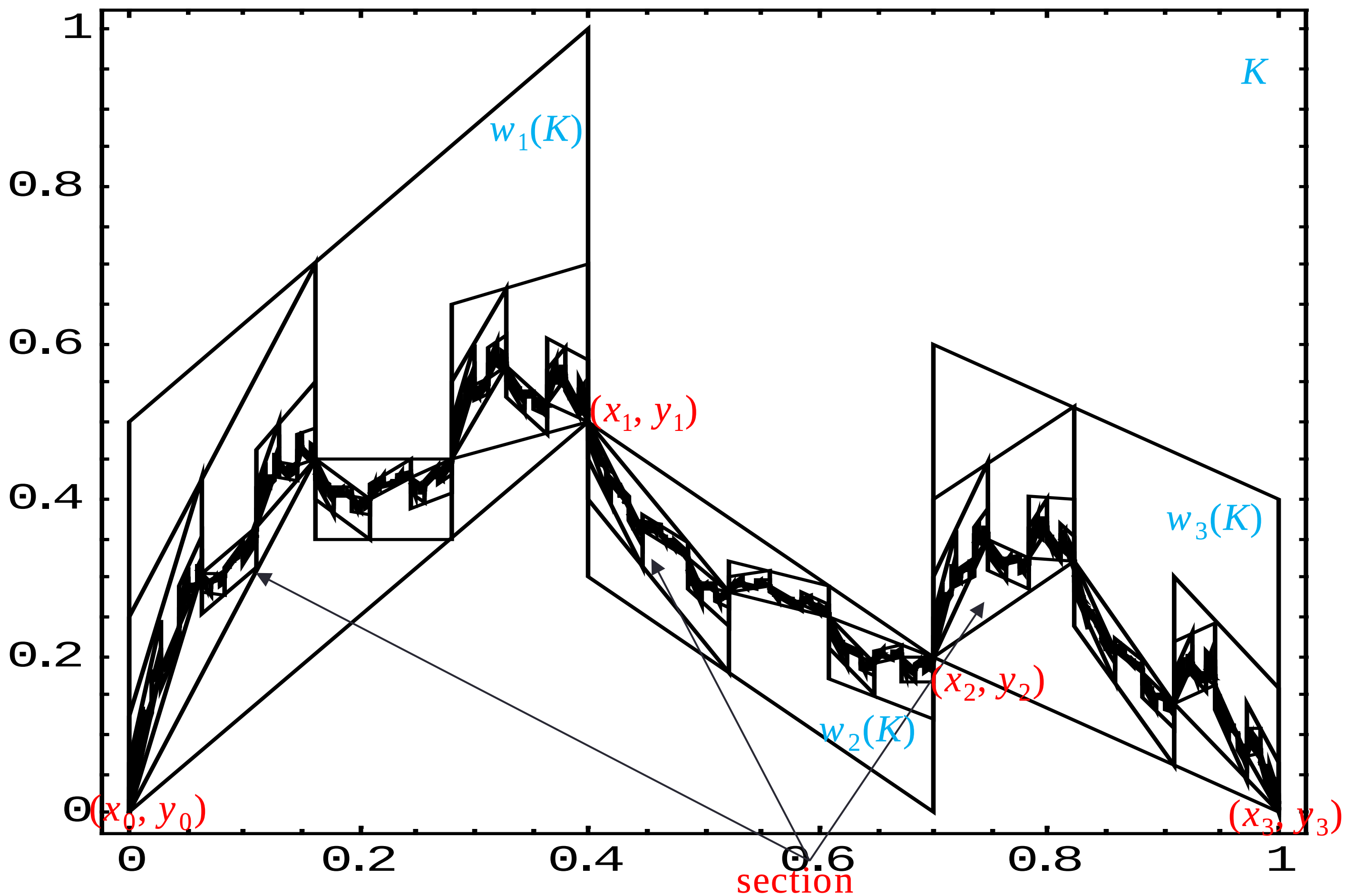
where $s_i \in (-1,1)$ is the (free) vertical scaling factor, whereas the coefficients a_i, c_i, d_i, e_i arise from the constraints

$$w_i\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} \quad \text{and} \quad w_i\begin{pmatrix} x_M \\ y_M \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} \quad i = 1, 2, \ldots, M.$$

IFS and interpolation functions

- The IFS $\{\mathbb{R}^2; w_{1-M}\}$ has a unique attractor, that is the graph of some continuous function which interpolates the data points.
- This function is called a fractal interpolation function (FIF), because its graph usually has non-integral dimension.

1D fractal interpolation



We map the entire (graph of the) function to each section of it.

Piecewise affine fractal interpolation

A pair of data points, which are called addresses, is now associated with each w_i

$$\{(\tilde{x}_{i,j}, \tilde{y}_{i,j}) : i = 1, 2, \ldots, M;\ j = 1, 2\}.$$

The domain is now the pair of addresses.

The constraints of the above mentioned case become

$$w_i \begin{pmatrix} \tilde{x}_{i,1} \\ \tilde{y}_{i,1} \end{pmatrix} = \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} \quad \text{and} \quad w_i \begin{pmatrix} \tilde{x}_{i,2} \\ \tilde{y}_{i,2} \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

subjected to

$$\tilde{x}_{i,2} - \tilde{x}_{i,1} > x_i - x_{i-1} \qquad i = 1, 2, \ldots, M.$$

Constraints

- For practical reasons suppose that the distance between the interpolation points along the horizontal and vertical direction is δ.
- We mapped the entire (graph of the) function to each section of the function. Now we map domains of the function to sections of the function. Suppose that each domain has size Δ.
- Points within a given interpolation section are not necessarily contained within any domain.

Interpolation functions in $\mathbb{R}^2$

- Let the discrete data

$$\{(x_i, y_j, z_{ij} = z(x_i, y_j)) \in \mathbb{R}^3 : i = 0, 1, \ldots, N;\ j = 0, 1, \ldots, M\}$$

  be known.
- Each affine mapping that comprises the hyperbolic IFS $\{\mathbb{R}^3; w_{1-N,\,1-M}\}$ is given by

$$w_{nm}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_{nm} & b_{nm} & 0 \\ c_{nm} & d_{nm} & 0 \\ e_{nm} & g_{nm} & s_{nm} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} h_{nm} \\ k_{nm} \\ l_{nm} \end{pmatrix},$$

  with $|s_{nm}| < 1$ for every $n = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, M$. The condition

$$\left\| \begin{matrix} a_{nm} & b_{nm} \\ c_{nm} & d_{nm} \end{matrix} \right\| < 1$$

  ensures that

$$u_{nm}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_{nm} & b_{nm} \\ c_{nm} & d_{nm} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} h_{nm} \\ k_{nm} \end{pmatrix}$$

  is a similitude and the transformed surface does not vanish or flip over.

5. FRACTAL-BASED IMAGE ENCODING AND COMPRESSION

- Introduction
- Encoding
- Decoding

Introduction

- Fractal compression is a lossy image compression method using fractals.
- It is best suited for textures and natural images, relying on the fact that parts of an image often resemble other parts of the same image.
- It differs from pixel-based compression schemes such as JPEG, GIF and MPEG since no pixels are saved.
- Special algorithms convert these parts into mathematical data called ‘fractal codes’ which are used to recreate the encoded image.
- These codes can be decoded to fill any screen size without the loss of sharpness that occurs in conventional compression schemes.

What is Fractal Image Compression?

- A set of contractive (affine) transformations can approximate a real image.
- Instead of storing the whole image, it is enough to store the relevant parameters of the transformations reducing memory requirements and achieving high compression ratios.

Why is it *Fractal* Image Compression?

- The decoder uses the same procedure for generating self-similar fractals
- Achieved by approximating each segment of the image by applying a (contractive) transformation on some larger segments in the image.

Why is it Fractal Image *Compression*?

- Standard image compression methods can be evaluated using their compression ratio
- It takes advantage of across-scale redundancies presented in the images

FIC using IFS

aided by

- Wavelets
- Genetic Algorithms
- Fractal Interpolation Functions

Fractal Image Encoding and Compression

- Based directly on the Collage Theorem
- Based on the Fractal Transform
- Based on Local (Partitioned) IFS
- Based on RIFS or on FIF

The inverse problem

- The fundamental principle of FIC consists of the representation of an image by an IFS of which the *fixed point* is ‘close’ to that image.
- The encoding process is first to find an IFS and then a suitable transformation *W* whose fixed point is ‘close’ to the given image.

Concerns

- How ‘close’ is the approximate to the real image?
- Is the convergence ensured;

The $\{1/n\}$ is a Cauchy sequence in the Euclidean subspace $T = (0, 1]$ of $\mathbb{R}$, but it does not converge in T.

The Collage theorem

If $B \in (\mathcal{H}(\mathbb{R}^n), h)$, where h is a metric, obeys

$$h\left(B, H\left(B\right)\right) \leq \varepsilon,$$

then

$$h\left(B, \mathcal{A}_{\infty}\right) \leq \frac{\varepsilon}{1-s},$$

where $s = \max\{s_i : i = 1, 2, \ldots, M\}$.

Interpretation

- The closer the union is to the given set, the closer the attractor of the IFS will be to the given set.
- To test the closeness of an attractor to a given set, one need not compute the attractor itself.
- The theorem is not constructive, it does not indicate how to find a set of proper mappings.

Collage

- In terms of w_i, we have

$$\left.\begin{aligned} &W(C) \approx C \\ &W(C) = \bigcup_{i=1}^{N} w_i(C) \end{aligned}\right\} \Rightarrow \bigcup_{i=1}^{N} w_i(C) \approx C$$

- This can be done by partitioning C into parts C_i,

$$C = \bigcup_{i=1}^{N} C_i$$

such that each part C_i can be closely approximated by applying a contractive affine transformation w_i on the whole C, i.e.,

$$C_i = w_i(C).$$

Fractal transform

- A technique invented by Michael Barnsley *et al.* to perform lossy image compression.
- This first practical fractal compression system for digital images resembles a vector quantization system using the image itself as the codebook.

Fractal transform compression

- Start with a digital image A_1.
- Downsample it by a factor of 2 to produce image A_2.
- Now, for each block B_1 of 4x4 pixels in A_1, find the corresponding block B_2 in A_2 most similar to B_1 and then find the grey-scale or RGB offset and gain from A_2 to B_2.
- For each destination block, output the positions of the source blocks and the colour offsets and gains.

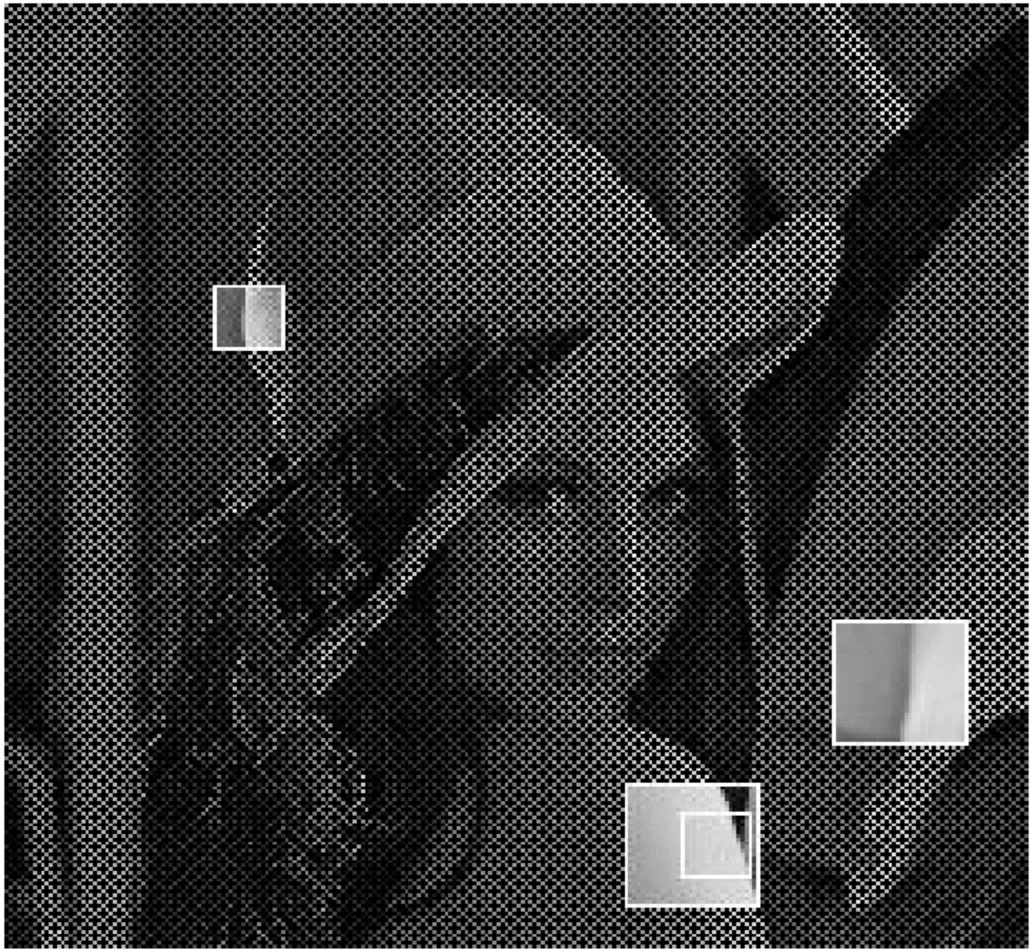
Fractal transform decompression

Starting with an empty destination image A_1, repeat the following algorithm several times:

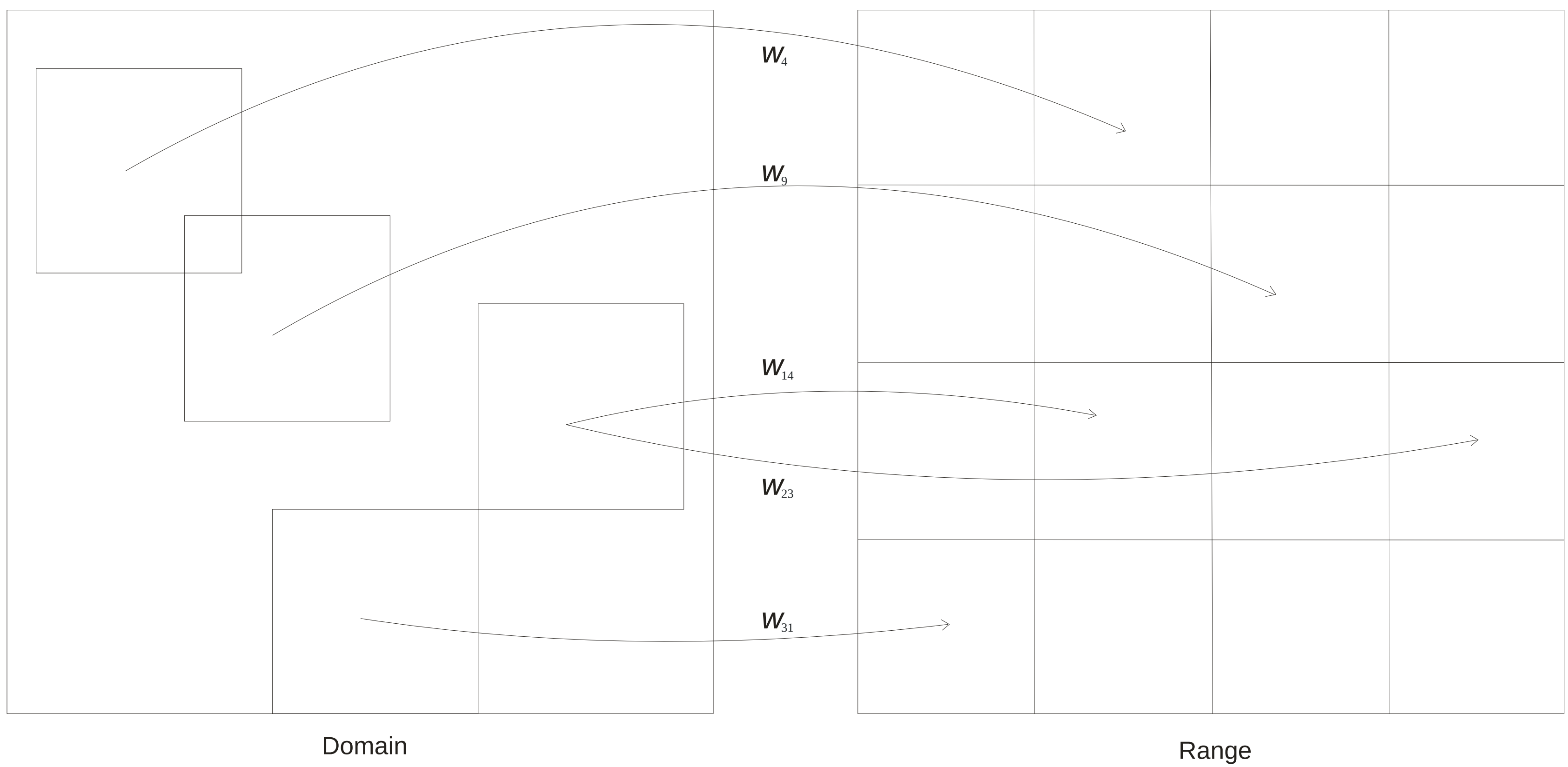
- Downsample A_1 down by a factor of 2 to produce image A_2.
- Then copy blocks from A_2 to A_1 as directed by the compressed data, multiplying by the respective gains and adding the respective colour offsets.

Intuition

- Real-world images, generally, do not contain parts that are affine transforms of the whole image.
- Different parts of the image may become similar under certain affine transformation.



Partitioned (local) IFS



Each range block is constructed by a transformed domain block.

Local IFS

- IFS
  - Approximates each part of the image by a transformed version of the *whole* image
- Local IFS
  - Approximates each part of the image by a transformed version of *another part* of the image
  - The image C is partitioned into range segments C_i, where

$$C = \bigcup_{i=1}^{N} C_i$$

  - Then each range segment C_i is approximated by a transformed version of a bigger domain segment D_i

$$\left.\begin{aligned} w_i(D_i) &\approx C_i \\ W(C) &= \bigcup_{i=1}^{N} w_i(D_i) \end{aligned}\right\} \Rightarrow \bigcup_{i=1}^{N} w_i(D_i) \approx C$$

Implementation issues

- How to segment the image?
- What kind of transformations to use?
- How to find the parameters of the transformations?
- Where to find the matching segments?

Encoding Images

- Suppose we are given an image f we wish to encode.
- We want to find $w_1, w_2, \ldots, w_N$ such that f is the fixed point of W.
- Partition f into N range blocks R_i.
- Find the domain blocks D_i and $w_i(\cdot)$ that minimize the distance $d(R_i, w_i(D_i))$, $i = 1, 2, \ldots, N$.
- The best matching domain D_i is said to *cover* the range R_i.

An illustrative example

- Original image 128 × 128 pixels
- Range blocks 4 × 4 ➔ 1024 blocks (*non-overlapping*)
- Domain blocks 8 × 8 ➔ 121 × 121
  = 14641 (*overlapping*)
- Need to compare 14641 squares with each of the 1024 range blocks
- Since the size of domain block is 4 times the size of range block, we need to downsample.

The basic idea

The image of Lena divided into domains of size 128×128 pixels.

The image of Lena divided into regions of size 64×64 pixels.

An illustrative example

w_i include

- translation and downsampling

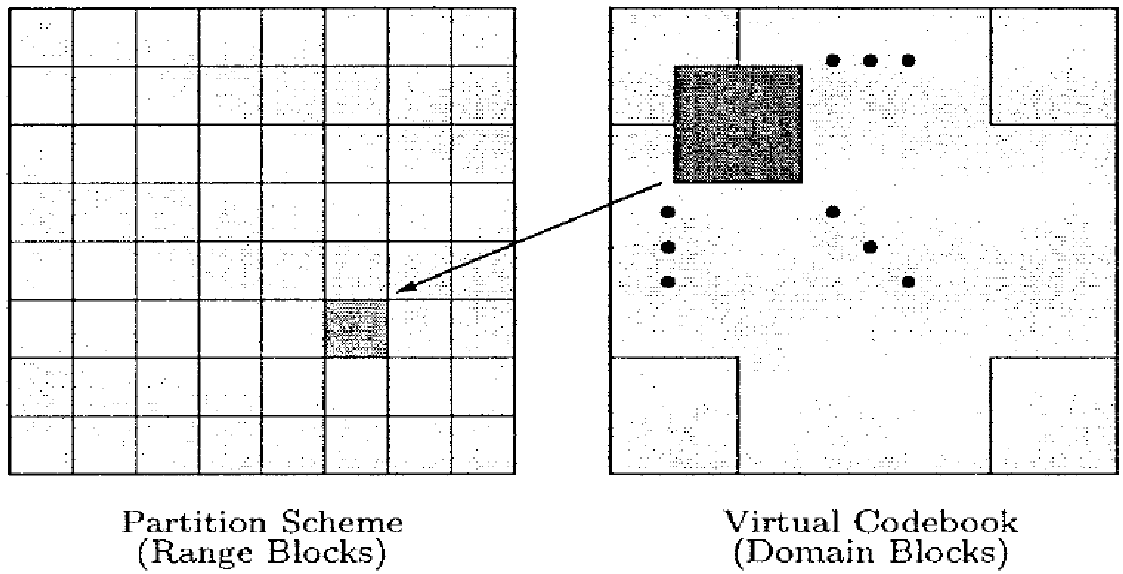
$$w_i\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}\left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \mathrm{col}_i \\ \mathrm{row}_i \end{bmatrix}\right)$$

- adjust contrast *a* and brightness *b*

$$w_i\left(\begin{bmatrix} x \\ y \\ I(x,y) \end{bmatrix}\right) = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & a \end{bmatrix}\begin{bmatrix} x \\ y \\ I(x,y) \end{bmatrix} + \begin{bmatrix} \mathrm{col}_i/2 \\ \mathrm{row}_i/2 \\ b \end{bmatrix}$$

Encoding Images = Finding w_i

- Search for best spatial transformation



One of the block mappings in a PIFS representation.

- Search for best grey-scale transformation

a = {0, 0.2, 0.4, 0.6, 0.8}

b = mean (range1D) – a*mean(p);

Things you can do for extra credits

- Add rotation and flip
  - Eight types of spatial transformations:
  - 1 ---> Rotate counterclockwise 0 degree.
  - 2 ---> Rotate counterclockwise 0 degree and flip.
  - 3 ---> Rotate counterclockwise 90 degree.
  - 4 ---> Rotate counterclockwise 90 degree and flip.
  - 5 ---> Rotate counterclockwise 180 degree.
  - 6 ---> Rotate counterclockwise 180 degree and flip.
  - 7 ---> Rotate counterclockwise 270 degree.
  - 8 ---> Rotate counterclockwise 270 degree and flip.

Things you can do for extra credits

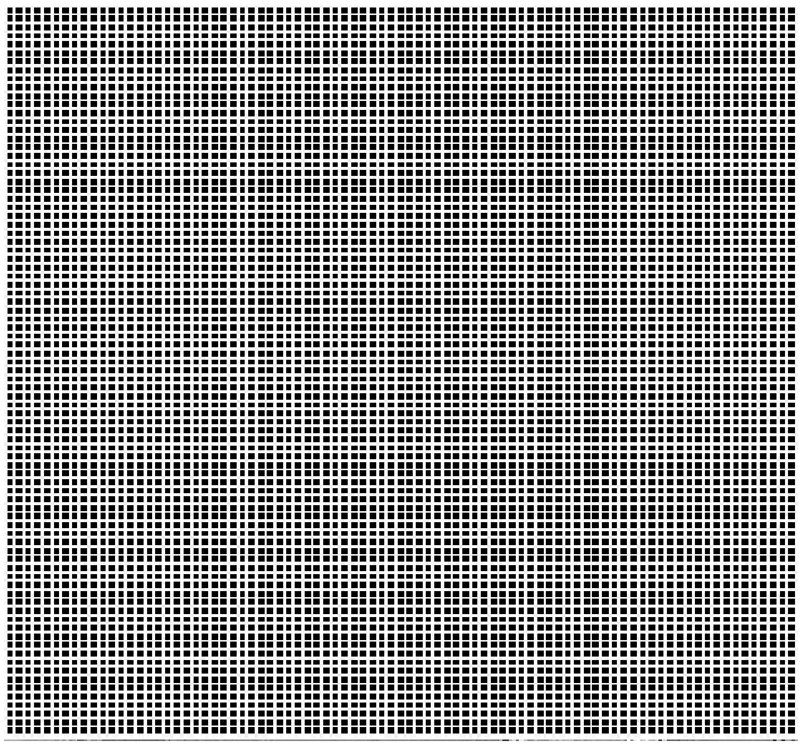
- Solve both *a* and *b* analytically
  - Minimize $$R = \sum_{i=1}^{n} (a \cdot p_i + b - q_i)^2$$
  - By setting the partial derivatives to zero we have

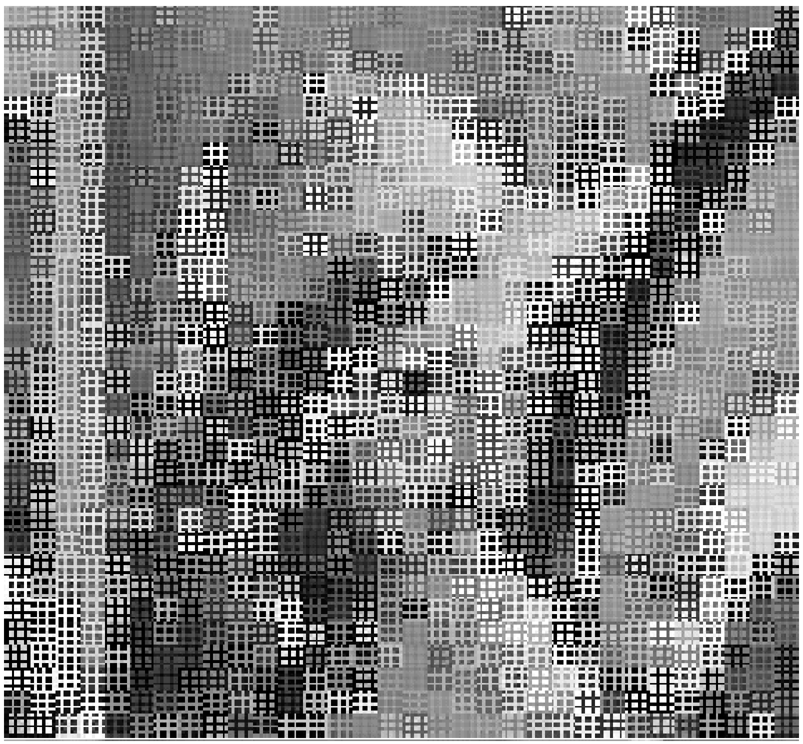
$$a = \frac{n^2 (\sum_{i=1}^{n} p_i q_i) - (\sum_{i=1}^{n} p_i)(\sum_{i=1}^{n} q_i)}{n^2 \sum_{i=1}^{n} p_i^2 - (\sum_{i=1}^{n} p_i)^2}$$

$$b = \frac{1}{n} \left[\sum_{i=1}^{n} q_i - a \sum_{i=1}^{n} p_i \right]$$

Results

- Left: Original
- Right: After first iteration





Results

- Left: After the second iteration
- Right: After the tenth iteration

Partition schemes

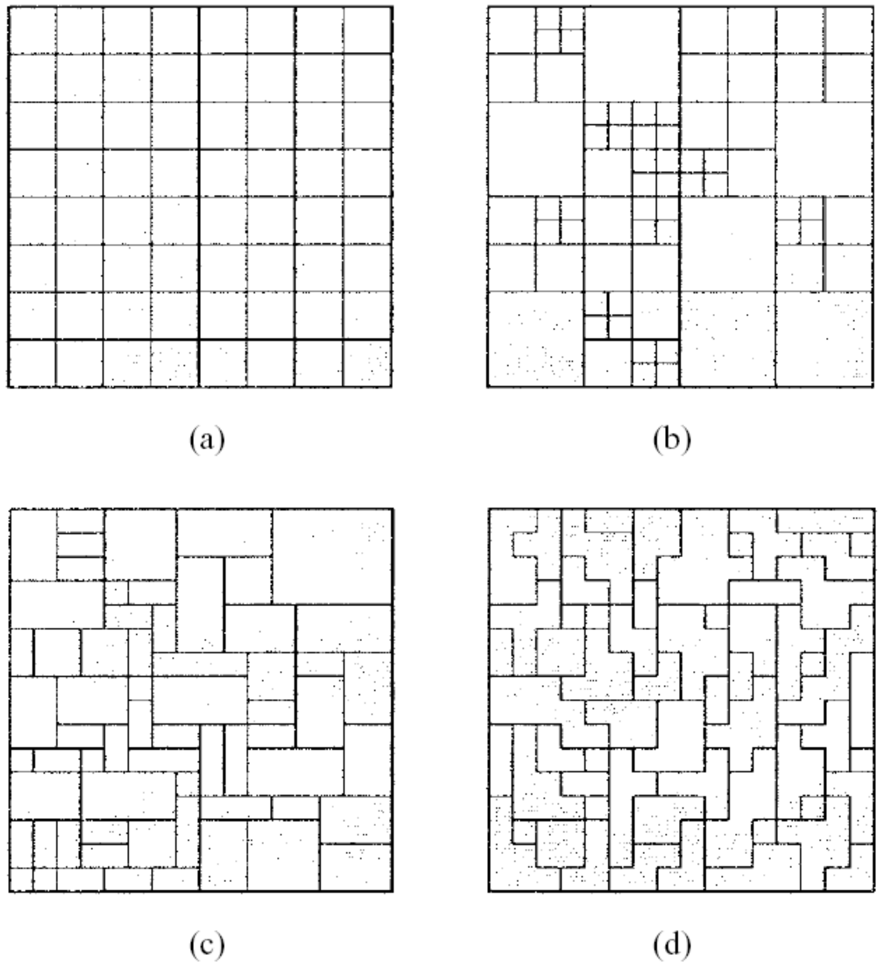


Fig. 2. Right-angled range partition schemes. (a) Fixed block size. (b) Quadtree. (c) Horizontal-vertical. (d) Irregular partition.

Partition schemes

- Motivation
  - Different regions should be covered by different sizes of range blocks.
- Quadtree partitioning
  - Divide a square into 4 equally sized subsquares.
  - Repeat divisions recursively until the squares are small enough.

Partition schemes

- Motivation
  - Use rectangular instead of square
- HV-Partitioning
  - A rectangular image is recursively partitioned either horizontally or vertically to form two new rectangles.
  - More flexibility than Quadtrees
  - Can make the partitions share certain similar structures.

HV-Partitioning

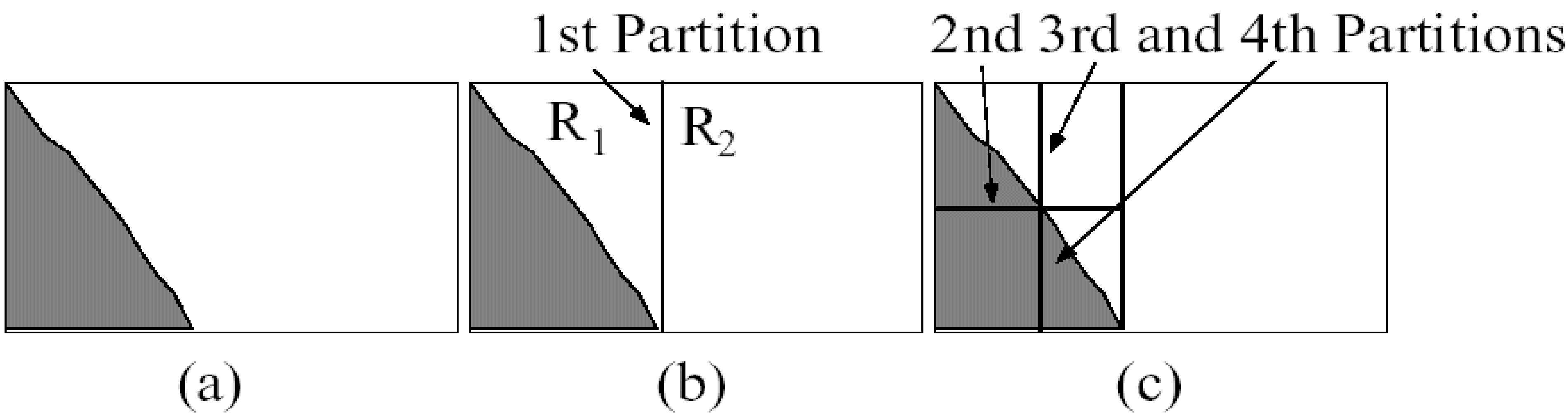
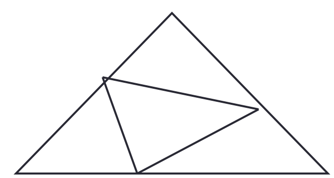


Figure 11. The HV scheme attempts to create self similar rectangles at different scales.

Results using HV-Partitioning

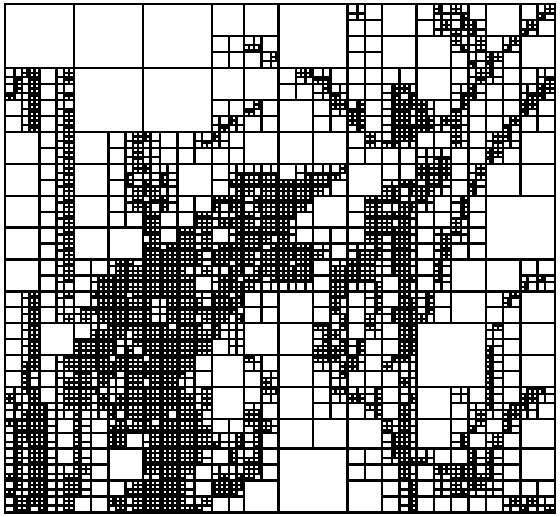
Triangular partitioning

- A rectangular image is divided diagonally into two triangles.
- Each triangle is recursively subdivided into 4 triangles by joining 3 partitioning points on the sides of the original triangle.

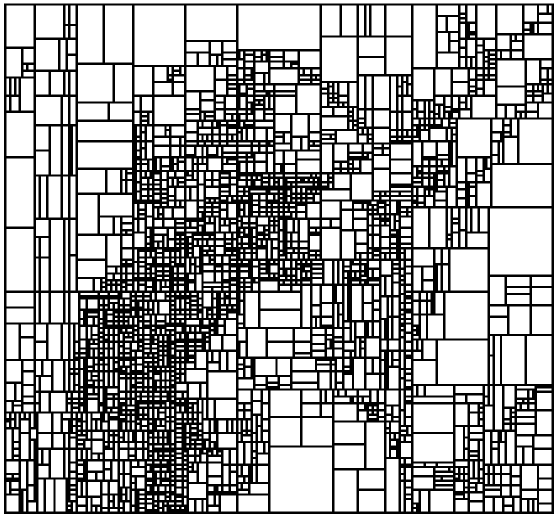


- More flexible: triangles can have self-similarities.
- The artifacts do not run horizontally and vertically.

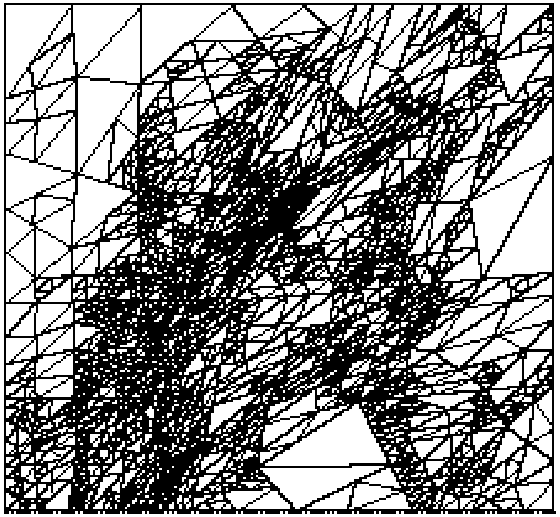
Ways to partition images



A quadtree
(5008 squares)

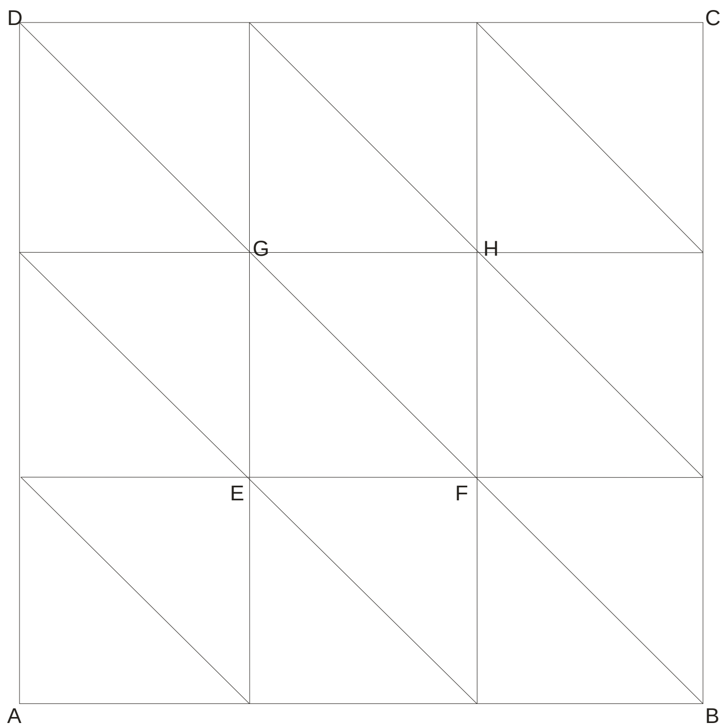


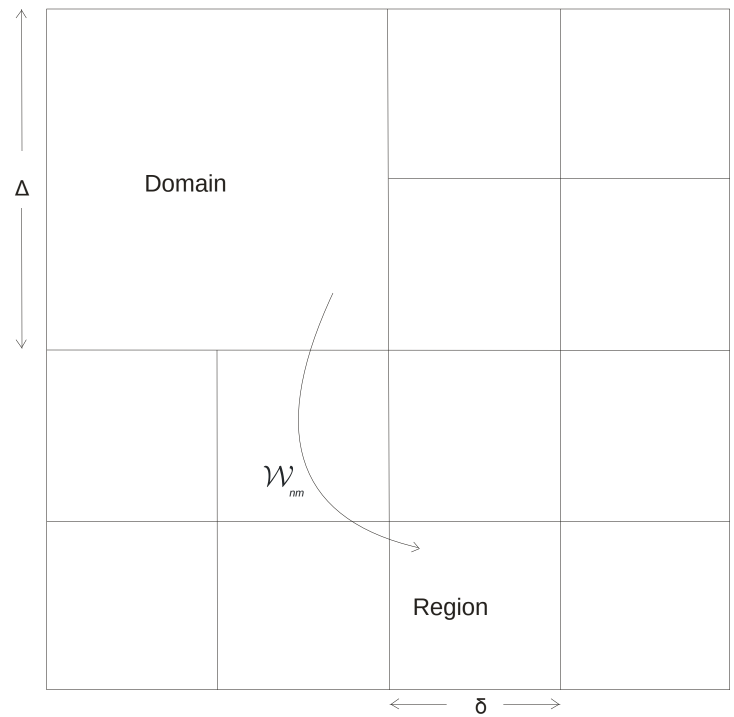
An HV-partition
(2910 rectangles)



A triangular partition
(2954 triangles)

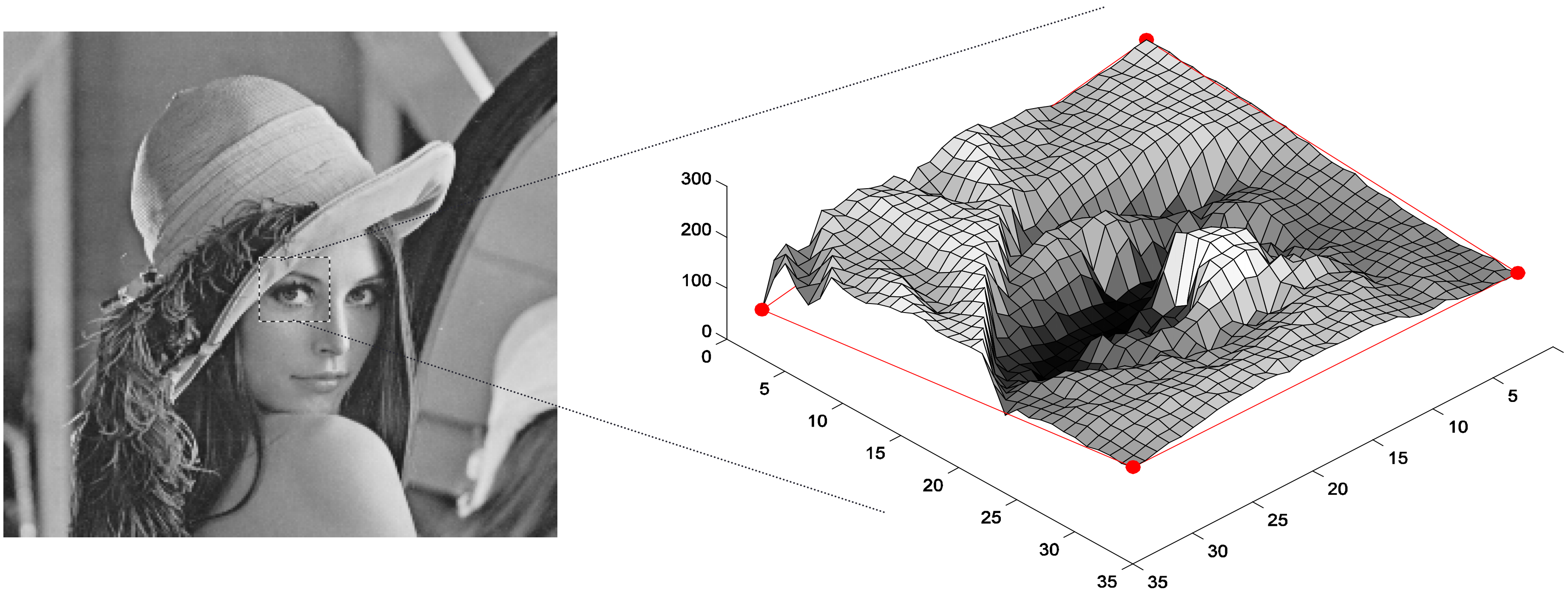
Rectangular lattices



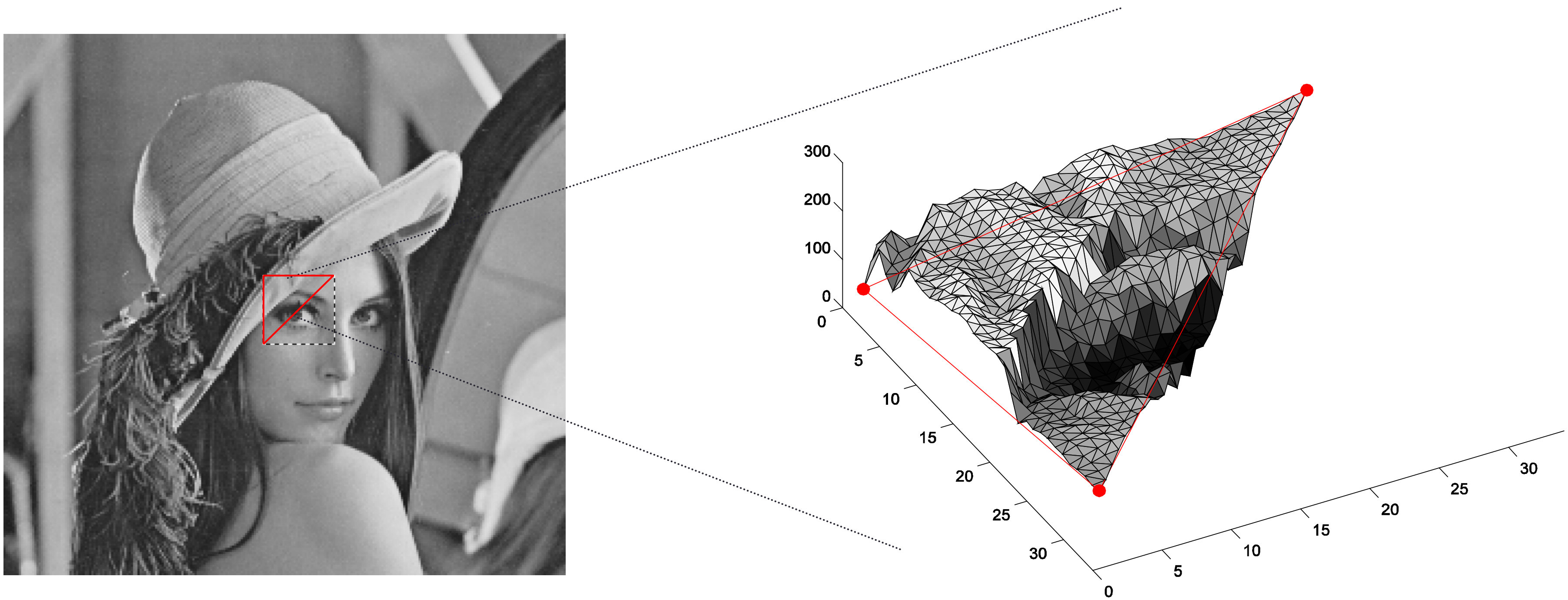


Domains for fractal interpolating surfaces over rectangular lattices using RIFS on (a) triangular tiling, (b) rectangular tiling.

Rectangular tiling



Triangular tiling



Comparative results of the 1D methods

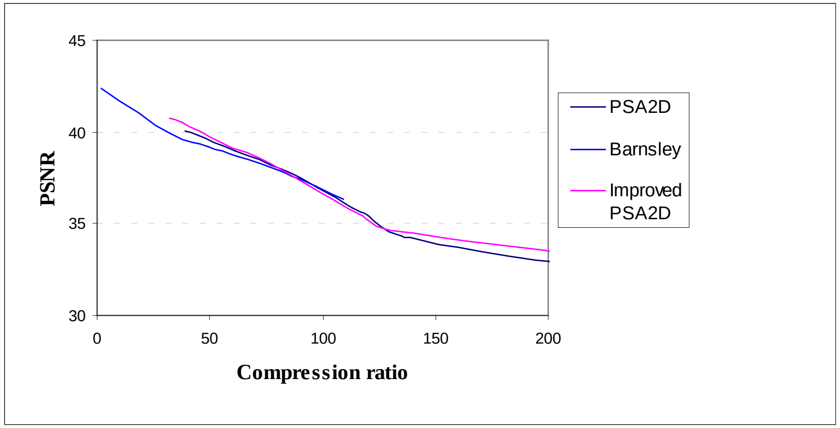
$$PSNR = 20\log_{10}\left(\frac{b}{rms}\right)$$

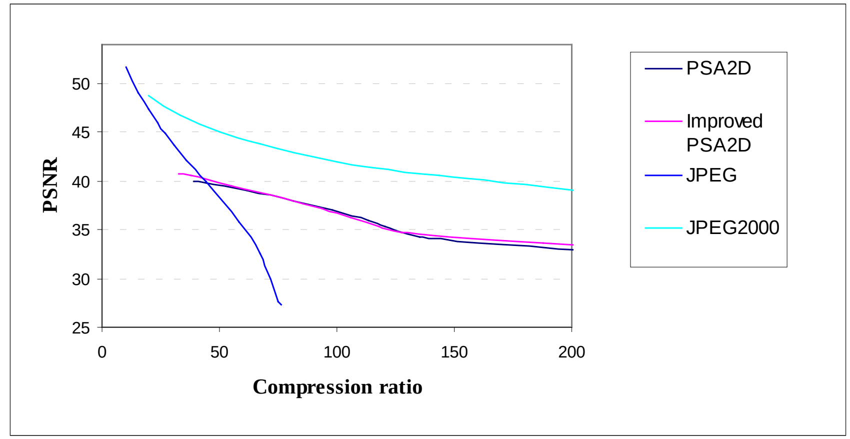
where *b* is the largest possible value of the signal (typically 255) and *rms* is the root mean square difference between two images.

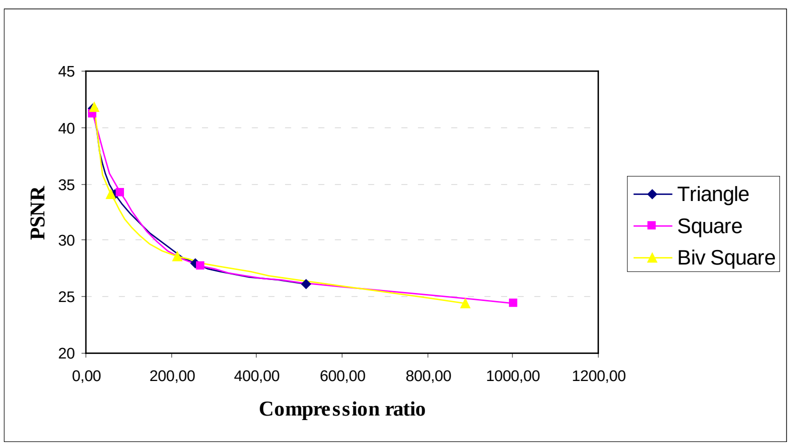
Method (*δ*,*Δ*)	**Encoding time (sec)**	**PSNR (dB)**	**Compression ratio**
Self-affine	194	20,87	1,33:1
Piecewise (4, 8)	0,1	42,8	1,35:1
Piecewise (8, 16)	0,1	32,31	2,69:1

Table 1
Image size: 433×433×8 bits, JPEG compression ratio: 6,07 (56,78), JPEG SNR: 30,05 (13,32) dB

Comparative results of the 2D methods







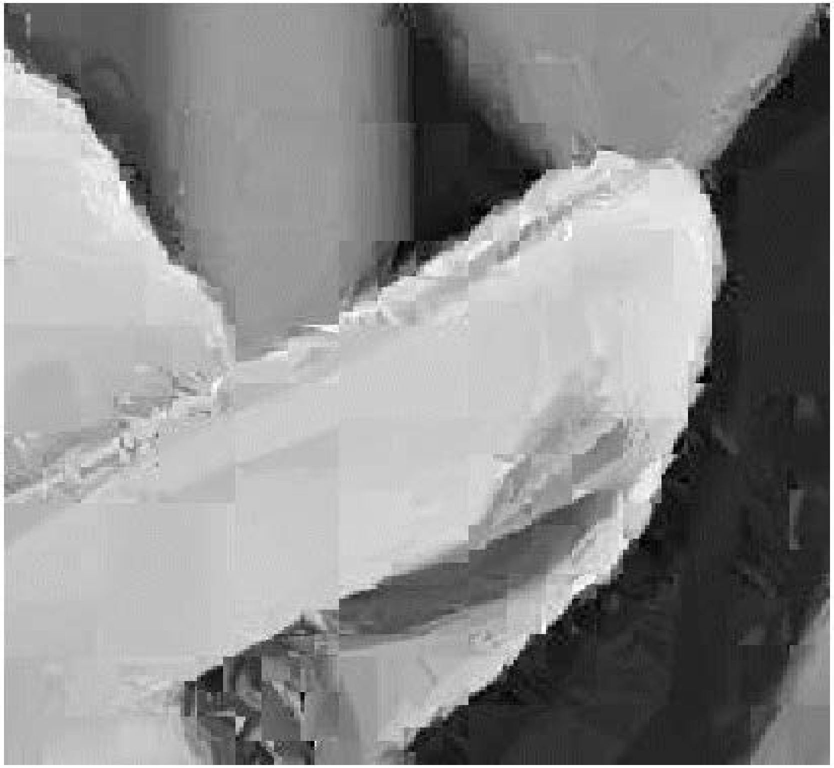
- (Over) Comparison of results using rectangular tiling
- (Left) Comparison between triangular and rectangular tiling

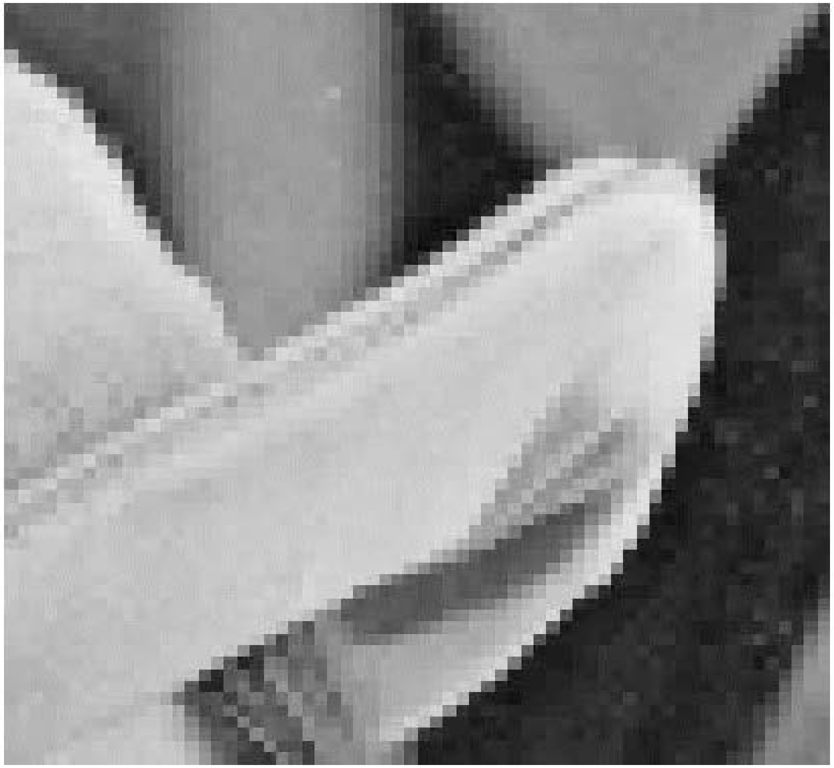
Fractal zoom

- Resolution Independence
  - Decoded image can have higher resolution than the original image.
- The additional resolution is generated because the domain block is larger than its range block.
- Assumption: details of the domain block is also similar to details of the range block,
  - although details of the range block are not given in the original image.

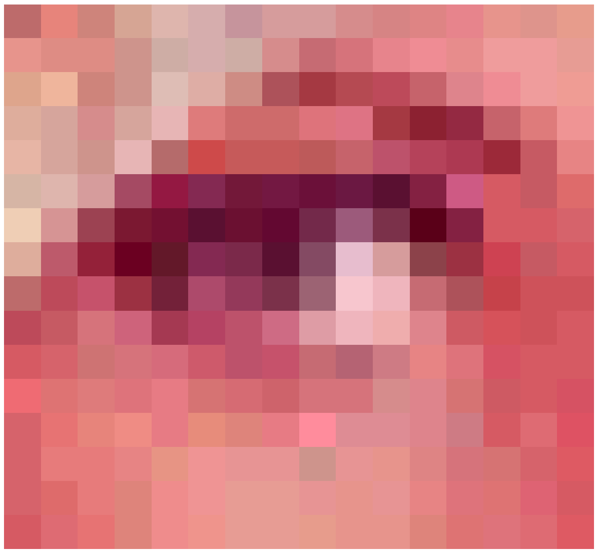
Fractal zoom

- Left: Decoding at 4 times its encoding size
- Right: Original image enlarged to 4 times the size

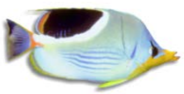


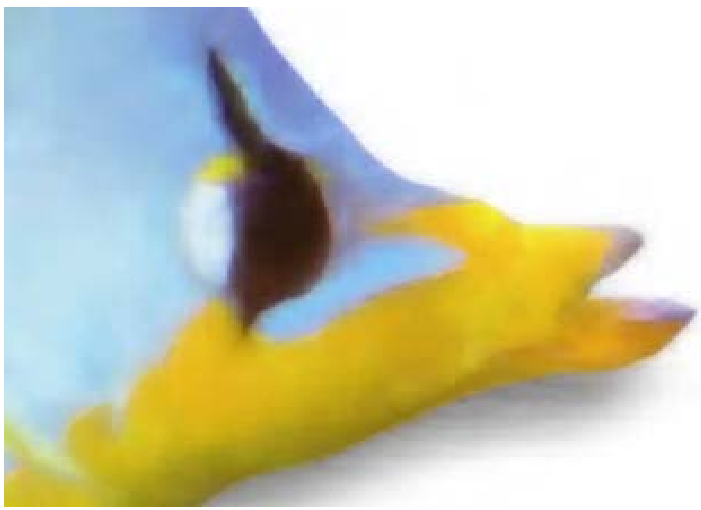


Fractal zoom



Fractal zoom





Complexity

- Fractal-based image encoding is asymmetric, i.e., encoding complexity is much higher than decoding complexity.
- Encoding complexity is much higher than that of transform coding (e.g. JPEG) and vector quantization
- Heuristics to speed encoding
  - Limiting Search
  - Limited number of transforms

Conclusions

- Fractal image compression is also called as fractal image encoding because a compressed image is represented by contractive transformations and mathematical functions required for reconstructing the original image.
- Fractal image compression enables an incredible amount of data to be stored in highly compressed data files.
- An inherent feature of fractal compression is that images become resolution independent after being converted to fractal code.
- This is because the iterated function systems in the compressed file scale indefinitely. This indefinite property, known as *fractal scaling* or *fractal zooming*, leaves no trace of the original pixel structure.