Complex Systems **3** (1989) 37–62

Chaotic Optimization and the Construction of Fractals: Solution of an Inverse Problem

Giorgio Mantica
Department of Physics, Atlanta University, Atlanta, GA 30314, USA
and
*School of Physics, Georgia Institute of Technology, Atlanta, GA 30332-0430, USA**

Alan Sloan
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
and
Iterated Systems Incorporated, 5550 P'tree Parkway, Suite 545, Atlanta, GA 30092, USA

Abstract. An *inverse problem* in fractal set construction is introduced in this paper, according to the theory of *iterated function systems* (IFS). This theory allows the construction of a class of fractals depending on a finite number of parameters. Finding a set of parameters which reconstructs a given fractal is the goal of the inverse problem. As the solution of the inverse problem generally involves a compression of the information encoded in the fractal, complexity theory is here applied. In particular, we define the *IFS-entropy* to characterize the class of fractals for which the problem can be profitably solved.

The inverse problem can be reduced to the *minimization* of a suitable function in parameter space. We describe a new algorithm to obtain a reliable minimum solution, which originates from the theory of dynamical systems. We suggest that this algorithm should greatly improve *simulated thermal annealing* à *la* Kirkpatrick-Szu when a metric structure can be given to an optimization problem.

1. Introduction

Fractal sets are mathematical fireworks thriving on the Dedekind continuum of real numbers. While their mere existence worried the founders of modern

*Computer address (Bitnet): ph262gm@gitvm1.bitnet.

analysis, in recent times a more sympathetic light has been shed on these entities: they can be a geometrical representation of ubiquitous natural objects like clouds, rivers, and forests [1], they unexpectedly arise in the dynamics of simple dynamical systems [2,3]. Yet, the usefulness of fractals as *bona fide* geometrical objects has not been fully exploited : algorithmical tools to *compute* fractals need to be developed; on the contrary real analysis and analytical geometry provide an efficient way to deal with smooth curves and manifolds.

A convenient construction of fractals is effected via *iterated function systems* (IFS) [4,5]. These consist of a finite collection of affine contractive maps over a compact space. Iterating these maps and considering the unique attractor of the process, one gets in most instances a complicated set with fractal structure. As this latter is uniquely and completely defined by the parameters of the affine maps, IFS are candidates for a basis on which to develop an algorithmic theory of fractals.

A crucial issue in this plan is discussed in this paper, namely the approximation or inverse problem. In analytical geometry one is often asked to find a curve interpolating a set of data, or the lines that best delimit a given set in the plane, thereby providing a concise representation of the geometrical object via a small number of parameters — say, the coefficients of a polynomial. Similarly, finding the IFS parameters that can best reconstruct an arbitrary set is crucial to a viable fractal geometry.

An "arbitrary" set is an "uncomputable" object, in the language of algorithmic complexity theory [6]. Now, IFS-generated fractals are efficiently computable: we want to use these latter to approximate the former! It is apparent that the information content of the IFS representation poses necessary restriction on the solutions of the inverse problem. We will make these ideas precise by defining an appropriate *IFS-entropy* which characterizes the class of "well approximable" sets. In this class, the inverse problem may be put in the form of a non-convex minimization problem.

To solve the minimization problem, we construct a deterministic dynamic on the parameter space which by sampling values tries to "learn" the structure of the function to be minimized, and to "predict" its minimum value. The study of forecasting models has recently been undertaken with the techniques of dynamics [7,8]. Yet, the strategy we present here appears to be original, in that we construct a dynamical system whose global attractor corresponds to the locus of the minima of the function to be minimized. The exact solution of the problem can then be obtained by evolving trajectories with initial conditions in the basin of attraction of the global minimum. This method provides an efficient solution of the inverse problem.

This paper is structured as follows: in the next section we review the general formalism of IFS and we apply it to a finite lattice space. In section 3 we discuss the inverse problem in fractal construction. The concept of approximate solution of an inverse problem is defined in section 4, which also contains a few results in approximation theory by IFS. The inverse problem is translated into a minimum problem in section 5. The method of "chaotic"

minimization is described and applied in sections 6 and 7. As a further example, section 8 treats the fractal approximation of arbitrarily complex data originating from experimental results. The Conclusions briefly summarize this work and the possible developments of the research here exposed.

2. Iterated functions systems and fractals

According to Mandelbrot [1], a fractal is a set $T \subset R^n$ with non-integer Hausdorff dimension. The iteration of affine linear maps offers an interesting way to generate a class of fractals: we review here this approach, as exposed in references [5,24].

We will generally work in C, a compact metric space with distance function d. In most of our derivations C will be the unit interval $I = [0, 1]$ with Euclidean distance, or its discretization $\mathcal{L} = \{0, 1, 2, \ldots, L\}$, that is the set of the first $L + 1$ integer numbers, with distance $d(i, j) = |i - j|/L$.

Let $\mathcal{P}(C)$ denote the set of all non-empty closed subsets of C. Fractal sets are objects within this space. Of course, the usual definition of fractal does not apply on the discrete space $\mathcal{P}(\mathcal{L})$ because the proper limits cannot be taken. On the other hand, it is customary to deal with finite-resolution computer images, whose "would be" fractal dimension can be numerically computed. Any set $B \in \mathcal{P}(\mathcal{L})$ can be represented in an obvious way as a finite sequence of binary digits $\{b_0 \ldots b_L\}$, the 1's corresponding to lattice points in B: $b_j = 1$ *iff* $j \in B$. The set $\mathcal{P}(C)$ becomes a compact metric space when endowed with the Hausdorff distance $h(A, B)$

$$h(A, B) = \max\{\sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{x \in B} \inf_{y \in A} d(x, y)\} \quad A, B \subset C. \tag{2.1}$$

In words, this distance gives the maximum separation of the furthest element (either in A or in B) from the opposite set. This distance is not of customary use in coding theory ($C = \mathcal{P}(\mathcal{L})$), where it is usually preferred to count the number of differing "digits" in the two sequences. Our choice is dictated on one hand by the fact that Hausdorff distance interlaces very nicely with IFS theory; on the other hand two fractals "very much alike" to the eye are usually very close according to Hausdorff, but may significantly differ according to "digits counting."[1]

Let's consider now a finite collection of mappings $\{w_i\}$

$$w_i : C \to C, \quad i = 1, \ldots, M, \tag{2.2}$$

where w_i are contracting transformations: for any i it exists a $\delta_i < 1$ such that

$$d(w_i(x), w_i(y)) \leq \delta_i \, d(x, y), \quad \forall x, y \in C. \tag{2.3}$$

[1]Consider as an example the following sets in C: $A = \{$ all even lattice points $\} = \{10101\ldots\}$, $B = \{$ all odd lattice points $\} = \{010101\ldots\}$. The "digits counting" distance between the two is L, while $h(A, B) = 1/L$.

The triple $(C, d, \{w_i\}_{i=1,\ldots,M})$ is called a *hyperbolic iterated function system.*

The collection of functions $\{w_i\}$ may be used to define a mapping W from $\mathcal{P}(C)$ into itself by

$$W(A) = \bigcup_{i=1,\ldots,M} w_i(A),\ A \in \mathcal{P}(C). \tag{2.4}$$

We denote by $\bar{\delta}$ the maximum of the δ_i. It can be shown that W is a contractive mapping in the metric given by the Hausdorff distance

$$h(W(A), W(B)) \leq \bar{\delta}\ h(A, B),\ \forall A, B \in \mathcal{P}(C). \tag{2.5}$$

Consequently, the contraction principle guarantees that it exists a unique *attractor* $\bar{A}$ of W, defined by $W(\bar{A}) = \bar{A}$. The convergence of the iterative construction of $\bar{A}$

$$\bar{A} = \lim_{m \to \infty} W^m(D),\ \forall D \in \mathcal{P}(C) \tag{2.6}$$

is geometrical. This attractor is the product of the IFS representation. As an illustration, the ternary Cantor set is the attractor of the following family of maps on I

$$\begin{array}{rcl} w_1(x) & = & \frac{x}{3} \\ w_2(x) & = & \frac{2}{3} + \frac{x}{3}. \end{array} \tag{2.7}$$

For simplicity, we will from now on consider only maps of the form (2.7), that is affine linear transformations.

The maps defining an hyperbolic iterated functions system on the lattice space $\mathcal{P}(\mathcal{L})$ are the natural discretization of (2.7):

$$w_i(x) = \frac{1}{L} Int\left[(p_i - q_i)x\right] + \frac{q_i}{L}\ ;\ x = \frac{l}{L}, \tag{2.8}$$

where p_i, q_i and l are integers less than or equal to L. The maps w_i bring the unit interval into the subintervals

$$w_i(I) = \left[\frac{q_i}{L}, \frac{p_i}{L}\right], \tag{2.9}$$

extrema included. We allow the possibility $p_i = q_i$, which is analogous to a continuous map with contraction factor $\delta_i < 1/L$. The discretized ($L = 243$) ternary Cantor set is shown in figure 1. Being only concerned with the attractor of an IFS and not with the so-called balanced measure on it (see reference [5]) we do not allow for overlapping intervals $[q_i, p_i]$. As the set of values $[q_i, p_i]_{i=1,\ldots,M}$ define our parameter space, we will implicitly assume them to be increasingly ordered. Equivalently, the unrestricted parameter space can be endowed with a permutation symmetry, since interchanging any pair of parameters is assumed to define the same maps w_i and hence the same attractor.

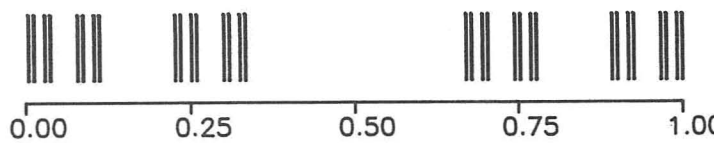

Figure 1: Finite resolution Cantor set. The set has been discretized on a $L = 243$ lattice. The set has been produced as the attractor of a 2 maps IFS, with parameters q_i, p_i, $i = 1, 2$: $[0, 81, 162, 243]$. The dark stripes correspond to points in the set.

3. An inverse problem in fractal construction

The IFS theory exposed in the previous section permits us to introduce an *inverse problem* in fractal sets construction: let us suppose that we are given a set T (target) in C. We want to find an IFS whose attractor is T.

We start by examining the case when $C = \mathcal{L}$. In this case, a trivial solution is readily obtained. Let $s_i \in \mathcal{L}$, $i = 1, \ldots, p$ be the elements of T, and consider the collection of discrete maps w_i defined by

$$w_i(m) = s_i, \ \forall m \in \mathcal{L}, \ i = 1, \ldots, p. \tag{3.1}$$

These are a (non-unique) solution of the inverse problem.

Equation (3.1), together with the "prescriptions" given by equation (2.4), (2.6) are a *coding* of the set T. It is apparent that this coding merely enumerates the points in T. What we would like to have is a higher "compression" of the information encoded in T, that is to say an efficient geometrical description.

To make these ideas precise, we utilize the binary representation $\{b_l\}_{l=1,\ldots,L}$ of T. The theory of algorithmical complexity [9] defines complexity (or information content) K_L of $\{b_l\}$ the length of the shortest sequence which programmed on a general Turing machine is capable to output $\{b_l\}$ and stop; in other words, the complexity is the length of the most concise coding of the sequence $\{b_l\}_{l=1,\ldots,L}$.

The task of finding a shorter representation of a particular sequence is delicate: no general criteria can be given other than actually trying a particular procedure. This is a consequence of the undecidable character of algorithmical complexity, and an equivalent form of the Turing Halting Theorem [10]. The IFS representation is a particular class of algorithms to implement this attempt, which makes use of the self-similarity structure (when present) of the set.

An IFS $(C, d, \{w_i\}_{i=1,\ldots,M})$ is an algorithm that can be translated into a sequence of binary digits of length $k(M)$,

$$k(M) \simeq M[2\log_2(L)] + \log_2(L) + Q + O(1). \tag{3.2}$$

The logarithmic contributions come from specifying the parameters q_i, p_i of the M maps in equation (2.8), (2.9). The additional $\log_2(L)$ term is a "self-delimiting" specification needed to input the size of the lattice space. The constant Q is the length of the digitized algorithm needed to "reconstruct" the fractal, equation (2.4), (2.6).[2] This last part of the coding does not depend on the specific sequence $\{b_l\}$, and in some practical applications (e.g. the transmission of digitized images) need to be transmitted or recorded only once. The measure of the efficiency of the representation is then the compression ratio $r = (2M+1)\log_2(L)/L$.

Let us denote by $\mathcal{P}_M(\mathcal{L})$ the subset of $\mathcal{P}(\mathcal{L})$ consisting of all sets that may be generated by M-maps IFS. Since we have proven the existence of a solution at least of the discretized inverse problem, we can claim that $\mathcal{P}(\mathcal{L}) \subset \bigcup_M \mathcal{P}_M(\mathcal{L})$. Nevertheless, the IFS representation is only meaningful when $K_L \sim k(M) \ll L$, being otherwise a mere "reshuffling" of information. Indeed, this is not only a practical instance, but it has deep implications in the limit of infinite L, as we shall see. The ternary Cantor is an example for which the IFS representation attains a high compression, as it belongs to $\mathcal{P}_2(\mathcal{L})$. On general grounds, we can estimate the cardinality of $\mathcal{P}_M(\mathcal{L})$ by

$$\#(\mathcal{P}_M(\mathcal{L})) = \left(\begin{array}{c} L \\ M \end{array} \right), \tag{3.3}$$

thanks to the permutation symmetry in the parameters. When $L \gg M > 1$ the information compression is significant. In this case we have that

$$\#(\mathcal{P}_M(\mathcal{L})) \simeq L^M \, M^{M-1/2}. \tag{3.4}$$

The cardinality of the full set is obviously $\#(\mathcal{P}(\mathcal{L})) = 2^L$. We immediately see that the ratio

$$\frac{\#(\mathcal{P}_M(\mathcal{L}))}{\#(\mathcal{P}(\mathcal{L}))} \simeq M^{M-1/2} \, 2^{M\log_2(L)-L} \tag{3.5}$$

tends exponentially fast to zero as L goes to infinity. This is consistent with the well known fact that most discrete sequences are random [6] and cannot be compressed by any algorithm.

On the other hand, the ratio (3.5) is a fast increasing function of M, for fixed L: in going from M to $M+1$ it is multiplied by the factor $(L-M)/(M+1)$. At the same time, the coding bit length $k(M)$ increases by $2\log_2(L)$: a larger number of maps permits a better approximation of a general fractal set, at the expense of a longer coding. These considerations specify the range and theoretical limitations of the IFS representation. In the next section we will allow L to increase indefinitely, in order to approximate a general fractal set. To do this, we need a few results in approximation theory by iterated functions systems.

[2]The statements of algorithmic complexity are all formulated modulo constants $O(1)$, depending only on the particular implementation or Turing machine.

4. Approximation theory by IFS

When dealing with a general space $C \subset R^n$, an approximation theory for fractals needs to be developed around the concept of approximate solution of the inverse problem. An ε-solution of the inverse problem is an IFS whose attractor $\bar{A}$ is closer than ε to the target fractal: $h(\bar{A}, T) \leq \varepsilon$.

A significant ε-solution of the inverse problem is easily obtained [11]. For simplicity, let us consider a target set $T \in \mathcal{P}(I)$, the generalization to higher dimensions being straightforward. An IFS with M affine maps is defined by the intervals $w_j(I)$,

$$w_j(I) = (\alpha_j, \alpha_j + \delta_j), \quad j = 1, \ldots, M, \tag{4.1}$$

where α_j and δ_j are affine constants and compression rates respectively. Again defining $W(I) = \bigcup_j w_j(I)$ as in equation (2.4), we may write

$$h(T, \bar{A}) \leq h(T, W(I)) + h(W(I), \bar{A}). \tag{4.2}$$

Taking the freedom to fix M in a suitable way, we can construct a covering of T by $W(I)$ with equal $\delta_j = \delta$, with the consequence that $h(T, W(I)) \leq \delta/2$. It is also easy to prove [12] that the last term at rhs of (4.2) may be written

$$h(W(I), \bar{A}) \leq \frac{1}{1-\delta}\ h(W(I), W^2(I)) \leq \frac{\delta}{1-\delta}. \tag{4.3}$$

From equations (4.2) and (4.3) it then follows that

$$h(T, \bar{A}) \leq 2\delta + O(\delta^2). \tag{4.4}$$

Choosing $\delta \simeq \varepsilon/2$ insures that the IFS specified by the "covering" maps (4.1) provides an ε-solution of the inverse problem.

The above result is also valid for the lattice space $\mathcal{L}$, if we assume that $\varepsilon \gg L^{-1}$, and the theoretical results of the previous section can now be applied. The number of maps M needed for a δ-covering of T is related to a topological quantity, the *capacity* or *fractal dimension* $C(T)$ [13] of the set T: $M \simeq \delta^{-C(T)}$, in the limit $\delta \to 0$. The set T is then known up to $b(\varepsilon) = -\log_2(\varepsilon)$ bits of precision by giving the parameters of the $M(\varepsilon) = 2^{C(T)(b(\varepsilon)+1)}$ maps. Following equation (3.2), the length of the IFS coding of the set T at precision ε is

$$k(\varepsilon) \simeq 2\log_2(L)\, 2^{C(T)[-\log_2(\varepsilon)+1]} + Q. \tag{4.5}$$

It is clear that in many instances the coding specified by the IFS (4.1) is far from being optimal ($k \gg K$), however it has a noteworthy implication: if we let L go to infinity while keeping εL constant, we are constructing a finer and finer discretization of the target fractal set. Now, equation (4.5) implies that the limit $K_\infty = \lim K_L/L$ for L tending to infinity, is null if $C(T)$ is less than one. This limit represents the complexity of the (infinite) discretization of the set [10], and hence *the discretization of a fractal set (whose fractal dimension*

is strictly less than one) does not produce a random digits sequence.[3] We are so assured to be dealing with computable objects.

We did not yet take into account any self-similarity structure of the set T. Aiming at this, Barnsley [12] has proven the following estimate

$$h(T, \bar{A}) \leq \frac{1}{1-\bar{\delta}}\, h(T, W(T)), \tag{4.6}$$

where again $\bar{\delta}$ is the maximum of the contraction rates of the affine maps: $\bar{\delta} = \max_j\{\delta_j\}$. In words, if a set T is a fixed point of W with tolerance ε, the IFS attractor $\bar{A}$ of W is close within $\varepsilon/(1-\bar{\delta})$ to the target set. This theorem permits a "collage" construction of IFS maps [14], but it requires the human eye to detect the self-similarity structure of T.

Focusing more deeply on the process of construction of an IFS, we now characterize the second iterate $W^2(I)$: it consists of M^2 intervals of length less than $\bar{\delta}^2$

$$W^2(I) = \bigcup_{i,j=1,\ldots,M} w_{i,j}(I), \tag{4.7}$$

$$w_{i,j}(I) = (\alpha_i + \delta_i\alpha_j, \alpha_i + \delta_i\alpha_j + \delta_i\delta_j). \tag{4.8}$$

Were we able to obtain a non-redundant covering of T by these sets (that is, a covering such that each $w_{i,j}(I)$ contains at least a point of T), then the general relation

$$h(T, \bar{A}) \leq h(T, W^n(I)) + \frac{\bar{\delta}^n}{1-\bar{\delta}} \tag{4.9}$$

would imply a $\bar{\delta}^2$-solution of the inverse problem.

It is simple combinatorics to see that the set $W^n(I)$ consists of the union of M^n intervals of length less than $\bar{\delta}^n$: the process has a geometric nature based on self-similarity. The main question now is how accurately a general target fractal T will be fit by this structure. To this aim, it is relevant to quantify the self-similarity of T appearing at successive scales via the IFS-*entropy* $\mathcal{S}(T)$ defined as the limit

$$\mathcal{S}(T) = \limsup_{\varepsilon\to 0}\{\frac{\log_2 M(\varepsilon)}{\log_2(1/\varepsilon)}\}, \tag{4.10}$$

where again $M(\varepsilon)$ is the minimum number of affine maps needed to obtain an ε-solution.

An upper bound to the IFS-entropy of a set T is given by its capacity: $\mathcal{S}(T) \leq C(T)$, as follows from the derivation of equation (4.5). The solution

[3]This result could be puzzling at first sight. To get an intuitive picture, it is sufficient to recall that a random infinite sequence must pass all conceivable tests for randomness (one should not be able to bet on it and win.) A non-integer dimension is an anomalous distribution of gaps (0's) which is a failed statistical test in the same way as an anomalous occurrence of 16's is a disaster for a roulette.

of the inverse problem enables one to experimentally determine the IFS-entropy of a set: one could look for the best ε-approximation obtainable with variable number of maps m, thereby defining a function $\varepsilon(m)$. For sets belonging to $\mathcal{P}_M(C)$ the function $\varepsilon(m)$ is exactly zero for $m \geq M$. This entails that the IFS-entropy of all these families is null. It is also null for all sets for which $\varepsilon(m)$ decays faster than any inverse power of m. We are now ready to *define as the ambient of the IFS representation the subset of* $\mathcal{P}(C)$ *characterized by a null IFS-entropy.*

A *positive* IFS-entropy entails that on a logarithmic scale the number of maps required to represent the set T is proportional to the resolution. Completely new patterns appear at successive scales when looking at fractals with positive IFS entropy, rendering the IFS representation quite ineffective. In the next sections we embrace the task of actually solving for the IFS parameters of ε-solutions.

5. Fractal reconstruction as a minimization problem

Even when the set T belongs to $\mathcal{P}_M(C)$, the actual determination of the $2M$ parameters of the affine transformations $\{w_i\}$ is a challenging problem. In fact, one would try to best match the M^n intervals obtained by $W^n(I)$ (see equation (4.8)), or minimize $h(T, W(T))$ according to the "collage" method. Both these methods are difficult and computationally heavy to implement, nor is any "direct" method of solution available in the literature at the present time. We are so left with the task of actually computing the distance function $h(\bar{A}, T)$ and trying to determine its minimum in parameter space. Again focusing the discussion on $\mathcal{P}(\mathcal{L})$ we immediately notice that an exhaustive search and comparison technique is jeopardized by the number ($\simeq L^M/M^{M+\frac{1}{2}}$) of possible choices, which gives to this approach a typical non-polynomial time signature [15].

The set $\mathcal{P}_M(\mathcal{L})$ can be mapped onto the lattice $\mathcal{Z}$ of allowable parameter values q_i, p_i of the affine maps (see section 2). A point $z \in \mathcal{Z}$, $z = [q_1, p_1, \ldots, q_M, p_M]$ identifies a particular IFS with attractor $\bar{A}_z$. The distance function of this attractor from an exact solution will be denoted by $F(z)$:

$$F(z) = h(\bar{A}_z, T), \quad z \in \mathcal{Z}. \tag{5.1}$$

The level curves of $F(z)$ on a plane crossing the parameter space (figure 2) show a complicate structure of local minima, which renders this minimization problem non - convex, resembling well known solid state analogies. Nevertheless, a few more results can be derived about $F(z)$.

Following equation (4.1) α_j and δ_j may be also used to parametrize the IFS maps. Consider two parameter vectors z, $z' \in \mathcal{Z}$, where z' is obtained by z incrementing one contraction rate δ_j by Δ while keeping all the other parameters fixed. A proper generalization of equation (4.8) leads to the estimate

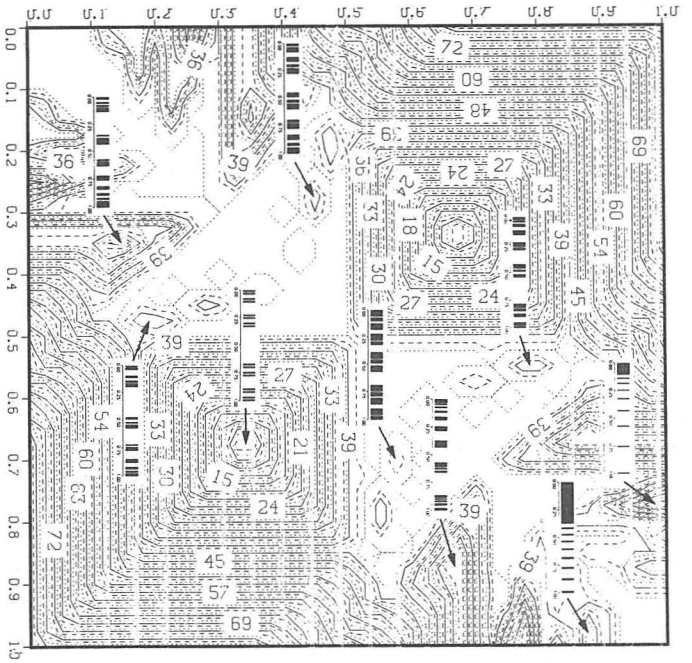

Figure 2: Hausdorff distance between the target Cantor set and attractors in $\mathcal{P}_2(\mathcal{L})$. For convenience, the distance is in integer units of $1/L$, $L = 243$. The level curves refer to a two-dimensional section of the four-dimensional parameter space, defined by $q_1 = 0$, $p_2 = 243$. The permutation symmetry in the two coordinates of the graph, $p_1 \leftrightarrow q_2$ is used here and in the following to produce full portrait pictures. The coordinates are normalized by $1/L$. The exact solution (minimum at zero distance) corresponds to the point $(1/3, 2/3)$ (or to the mirror $(2/3, 1/3)$). The attractors corresponding to this and to a number of local minima are shown superimposed.

$$h(\bar{A}_z, \bar{A}_{z'}) \leq \frac{1}{(1-\bar{\delta})^2}|\Delta|, \tag{5.2}$$

where $\bar{\delta}$ is now the largest contraction factor, before and after the variation $\delta_{\bar{j}} \to \delta_{\bar{j}} + \Delta$. We can also prove that a variation in one of the affine constants $\alpha_{\bar{j}}$ of the amount Δ results in

$$h(\bar{A}_z, \bar{A}_{z'}) \leq \frac{1}{(1-\bar{\delta})}|\Delta|. \tag{5.3}$$

These relations are important. They express continuity between the Hausdorff metric and the metric $d(z, z') \equiv \sum_j |\alpha_j - \alpha'_j| + |\delta_j - \delta'_j|$ in the parameter space. They imply that around the exact minimum of $F(z)$ (when $T \in \mathcal{P}_M(C)$) there exists a region of volume at least $V(\Delta) \propto \Delta^{2M}$ where $F(z) \leq \Delta$. The region around a ε-minimum can be similarly described. These facts are exemplified in figure 2, where we notice that the estimates (5.2) and (5.3) are sharp in the monotonically decreasing region around the minimum. Preliminary computations on the target set can be performed in order to exclude the consideration of a finite region around $\bar{\delta} = 1$. In this case the above estimates imply that the distance function is also Lipschitz.

This constraints imposed by the nature of the problem on the distance function $F(z)$ are by no means stringent, and are indeed present in other "complex" minimization problems, yet they conspire to render the "chaotic" minimization procedure exposed in the next section quite effective.

6. Chaotic optimization of complex problems

The problem of non-convex minimization has been solved by the introduction of stochastic algorithms, of which *simulated thermal annealing* (STA) is the basic model [16]. The application of STA to fractal set reconstruction was first proposed in reference [17]. We introduce here a conceptually different and new algorithm to efficiently compute an optimal solution. As we claim that this algorithm can overcome some of the limitations of typical stochastic algorithms, we open now a parenthesis to discuss these latter methods.

Our criticism of standard STA is two-fold: first, STA requires a high computational price, in terms of the unpredictable number of iterations needed to find a satisfactory minimum. Its performance is usually gauged accordingly to the cooling schedule [18], but like the convergence of the method itself, having a faster cooling schedule is only meaningful in the infinite time limit. Second, STA is a general-purpose algorithm: it works with probability one for *all* functions $F(z)$. As a consequence, it doesn't take any advantage from their global structure, as will be apparent from the following considerations.

We can think of any minimization procedure as composed of a sequence of two elementary steps: S1, evaluation of the function F in a lattice point z; S2, comparison of $F(z)$ with a set of m values previously found and choice of a new point z'. The integer m quantifies the *memory* of the process. In

simulated thermal annealing random element is introduced in S2. This step is in fact effected by a family of operators $U(\tau)$[4]

$$U(\tau): \mathcal{Z} \times \Omega \to \mathcal{Z} \tag{6.1}$$

where Ω is a probability space, and τ a parameter called *temperature*. The iterative application of $U(\tau_n)$ (τ_n, $n=1,\ldots$ is called an *annealing schedule*) generates a sequence of random variables

$$z_n(\omega) = U(\tau_n)(z_{n-1}, \omega),\ \omega \in \Omega. \tag{6.2}$$

Any iteration of U provides new information on the structure of the function F: not only $F(z_n)$ is exactly known, but sometimes this infers information on neighboring points as well. Nevertheless, in STA only the position at time n influences the next move ($m = 1$): a point z' generated starting from z_n is "accepted" only considering the difference $F(z_n) - F(z')$ [16]. Once z_{n+1} is so determined, the information contained in $z_n, F(z_n)$ is discarded. Convergence of $z_n, n \to \infty$ to the absolute minimum is achieved thanks to the infinite amount of complexity contained in the probability space Ω. With probability one, a realization of the random sequence (6.2) will eventually remedy all its "errors" and sit in the absolute minimum. Standard convergence theorems [19] demand in fact that a state in a neighborhood of any given one may be generated infinitely often in time.[5] The same principles inform a "Cauchy" STA, which has been proposed in order to optimize local sampling versus acquisition of new information by non-local moves [21].

The new minimization procedure we now propose tries to take advantage from the information gained in the step S1 (which is typically time-consuming). Moreover, it tries optimize this advantage over a *finite* number of iterations. To show this, let us consider a *dynamic* on $\mathcal{Z}$ given by a family of evolution operators $B(\lambda)$

$$B(\lambda): \mathcal{Z}^m \to \mathcal{Z}. \tag{6.3}$$

$B(\lambda)$ produces a new lattice state z deterministically as a function of m previous "positions" and of the parameter λ. Typical non-random minimization schemes can be formalized in this way, for constant λ. Like the steepest descent method, they perform poorly when dealing with complicated functions, since they get very easily trapped in local minima. On the contrary, let us suppose that the following facts hold:

> The mappings $B(\lambda)$ utilize the information in $z_{n-1}, \ldots, z_{n-m}$ and in $F(z_{n-1}), \ldots, F(z_{n-m})$ in order to compute z_n.

[4]The usual formulation of STA via a non-stationary Markov chain [18] is equivalent to the one proposed in equation (6.1), (6.2). We want to stress here the dynamical aspect of the algorithm, to offer a comparison with our method.

[5]When dealing with finite numerical simulations a heuristic criterium for the "time scale" of the annealing is usually provided by an *ad hoc* prescription called the *equilibrium condition* [20]. The temperature is lowered when it appears that the random trajectory of the Markov chain has provided a good sample of the equilibrium distribution, at that temperature. Many iterations are usually required to achieve this goal.

For any finite λ greater than zero the motion generated by $B(\lambda)$ is exponentially unstable (positive Lyapunov exponents).

As λ tends to zero, the motion is more and more attracted by the lowest F-value points in the m memory.

Determining the *dynamical attractor* of the system $B(\lambda)$ is the goal of chaotic optimization, since it contains the global minimum of $F(z)$. This can be done numerically evolving a few randomly initiated trajectories. The part of the "phase space" which originate trajectories ending in the global minimum is the "basin of attraction" of this latter. This may not extend to the whole parameter space due to the presence of local minima in the global attractor of the system $B(\lambda)$. Yet, if it is wide enough, a few trials are able to provide a solution of the optimization problem. Also, by increasing the "computation time" of the method (i.e. the typical time interval needed for a trajectory to reach the attractor) the basin of attraction of the global minimum can be widened.

We can provide a heuristic argument to support this conjecture. The complexity of the motion is determined by the function $F(z)$: "simple" functions originate stable trajectories, while more complicated optimization problems give rise to unstable dynamics. We expect that enough information will be encoded in the sequence z_n and in the sample $F(z_n)$ to provide a good approximation to the global minimum of the function. Indeed, it may be proven in the theory of dynamical systems that an "unstable" trajectory has maximal complexity [22]: as a consequence it represents the "best" possible "sampling" of $F(z)$.

It is not possible to know *a priori* what dynamic would provide the fastest convergence to a good solution, not necessarily the optimal one. For this reason we allow the dynamical system $B(\lambda)$ to be tuned according to a parameter λ. This can be varied externally, or by an internal feed-back (a primitive form of "learning").

A particular realization of these ideas is used in this paper: we consider the integer-time dynamical system $B(\lambda_n)$ acting in D-dimensional space, defined as follows: a Coulomb charge is placed in each of m initial locations $z_1, \ldots, z_m$. The intensity of the charge is denoted by $q(z_j)$ and is

$$q(z_j) = \begin{cases} 1 & \text{if} \quad j = m \\ 0 & \text{if} \quad m - p < j < m \\ -\exp\{\lambda_n^{-1} |F(z_j) - F(z_m)|\} & \text{if} \quad F(z_j) < F(z_m), \\ & \qquad j = 1, \ldots, m - p \\ \gamma(F(z_j) - F(z_m)) & \text{if} \quad F(z_j) \geq F(z_m), \\ & \qquad j = 1, \ldots, m - p \end{cases} \tag{6.4}$$

where γ is a positive constant.[6] The Coulomb force on the unit positive charge located at z_m is then calculated. A step in the D-dimensional space is

[6]The exclusion of $z_{m-p+1}, \ldots, z_m$ provides a delay p on the effect of newly generated states and reduces the time spent into any local ball.

Set	M	$N_{\max}$	λ_0	λ_N	γ	h_0	h	h_N	h_{opt}
A	2	380	243.	.05	.01	49.3	16.7	5.0	0.
B	3	1200	"	"	"	45.4	18.2	7.4	1.
C	4	1200	"	"	"	42.4	15.6	5.3	1.
D	4	1200	"	"	"	44.5	15.7	4.0	1.
E	5	2000	"	"	"	31.2	13.6	5.2	2.

Table 1: Inverse problem solution for the sets described in figure 4.

then taken from z_n in the direction of the force, thereby determining a new lattice point z_{m+1}. This provides m data $(z_2, \ldots, z_{m+1})$ for the successive iteration, which is performed in the same way. We remark that the computer implementation of this procedure can be fully vectorized, and a large amount of information can be processed at a speed competitive with STA; the computation of the Hausdorff distance is here the rate-limiting process.

A discussion of the few external parameters occurring in our formulation is in order. The schedule $\lambda(n)$ is chosen accordingly to the expected oscillations of F, in such a way that at the beginning of the computation only high cost differences will have a definite effect, while at the end the system will greatly overprice the smallest values of F found. This is the only external information required, and it may be arbitrarily specified in a first computation. The parameter γ weights the balance between attractive and repulsive forces, which turns in favor of the former when λ is sufficiently small.

7. Solution of the inverse problem

We now present the results of the numerical simulations obtained following the theory of the previous paragraph. We first reproduce the ternary Cantor set of figure 1. The minimization problem depends on the four independent parameters $[q_1, p_1, q_2, p_2]$. In figure 3 we report the data concerning a set of $N = 10$ randomly initiated trajectories. After approximately 350 iterations, all the trajectories are trapped in the immediate neighborhood of the absolute minimum. Figure 3 shows the average Hausdorff distance versus iteration-time. The parameter λ_n is shown in the frame. An exponentially decreasing function has been chosen. The solution of the inverse problem is exactly accomplished.

One might object that our solution has been facilitated by the relatively simple structure of the example. To solve the inverse problem for more complex fractals, and to investigate the performance of our method as a function of the computational complexity of the problem, we increased the number of generating maps. Figure 4 shows the examples we investigated. The exponentially increasing cardinality of $\mathcal{Z}$ according to equation (3.3) are (a) 141,722,460; (b) 2.6×10^{11}; (c,d) 2.6×10^{14}; (e) 1.6×10^{17}. For each of these problems we performed a chaotic optimization as previously described, evolving 10 independent trajectories. Table 1 reports the corresponding parameters and results. Each set A-E has been reconstructed with M maps,

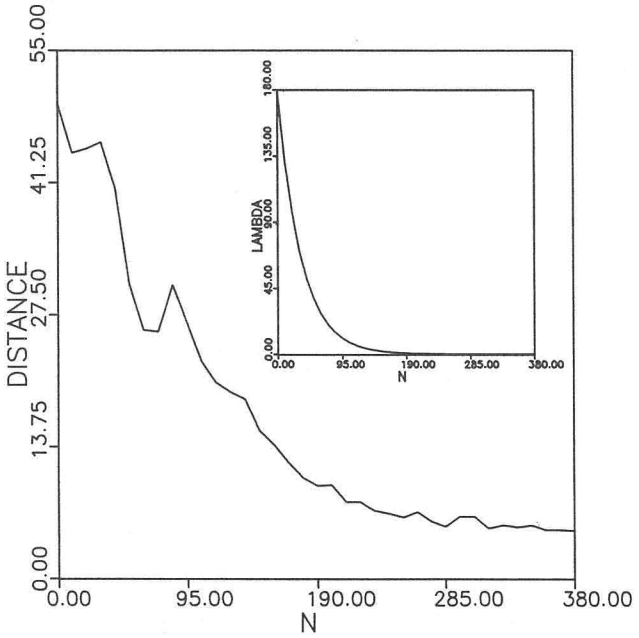

Figure 3: Average Hausdorff distance (in units of $1/L$) versus "time" in the chaotic optimization process consisting of 380 iterations. The initial average distance is 49.3, which corresponds to the expected value of a "blind guess". Ten different trajectories contribute to the average. The value of λ decreases from 243 to .001, as plotted in the frame. The value of γ is .01.

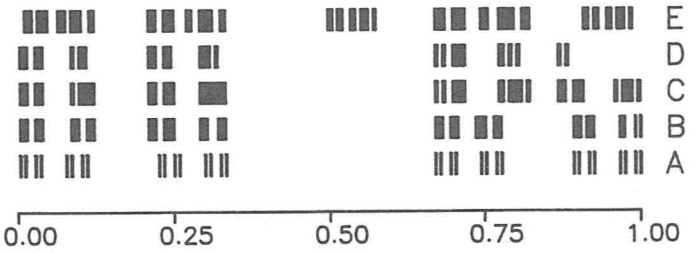

Figure 4: Target fractal sets obtained by IFS with increasing number of maps. (A) Cantor set of figure 1 (reported for comparison). (B) 3-maps IFS: $[q_1, \ldots, p_3] = [0, 30, 50, 81, 162, 243]$. (C) 4-maps: [0,-30, 50, 81, 162, 200, 210, 243]. (D) 4-maps: [0, 30, 50, 81, 162, 200, 210,-215]. (E) 5-maps: [2, 30, 50, 81, 120, 140, 162, 200, 220, 240].

via a chaotic optimization process consisting of $N_{\max}$ iterations, with initial parameter $\lambda = \lambda_0$, and final $\lambda = \lambda_N$. The average distance of the optimization process from the exact solution in the first 50 iterations is h_0, which decays to h_N in the last 50, while its global average is $\bar{h}$; h_{opt} is the final state (solution) distance from the target. All distances are in units of the lattice constant $1/L$. When h_{opt} is different from zero, the exact solution has not been found, yet the process has come so close to it that any conventional method can easily home in.

Figure 5 shows pictorially how the method "works" in the case of the set D: the points sampled by the method are projected onto a Hausdorff distance section of the parameter space. The cloud of sampling points rapidly focuses around the minimum. It is to be remarked that the points appearing on a plane in the figure are indeed floating in an eight-dimensional space.

The complexity of the inverse problem may be also fully appreciated in the eight-dimensional parameter space corresponding to the set D. The structure of the Hausdorff distance function is rather complicated and allows for the presence of a deep local minimum in proximity of the global (figure 6).

Each of these minima is surrounded by a region of low values of the cost function of linear dimension approximately $l = .05$. We can estimate the average number of iterations needed to randomly sample a point inside such region, as the inverse ratio of the volume of the region, l^8, to the total volume of the parameter space, $V \sim 1/8!$, and we obtain $N \propto 10^8$. It is of course more difficult to give analogue estimates for STA, since its performance can be enhanced by properly modifying the "move" strategy S1 [17]. We implemented a simulated annealing procedure according to references [20,21]. By

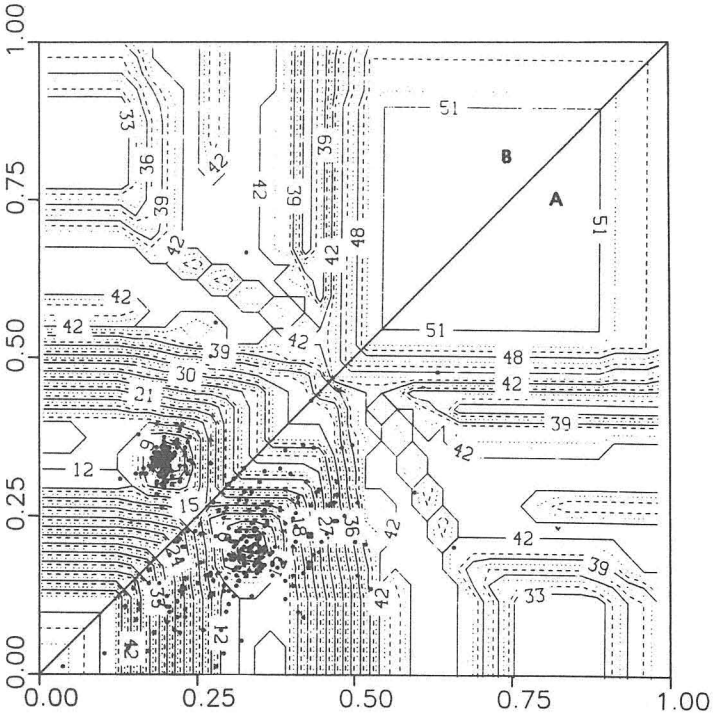

Figure 5: Dynamics of the optimization process for the set D (Table 1). The projections (*not* the intersections !) of the sampling points over the plane $q_1 = 0$, $p_1 = 30$, $q_3 = 162$, $p_3 = 200$, $q_4 = 210$, $p_4 = 215$ are shown as dots over the "equipotential" lines of the Hausdorff distance. The graph is symmetric in the coordinates q_2, p_2. Thanks to this symmetry, we can partition the figure, and plot in A) samples at any 5 iterations from $n = 1$ to $n = 300$, and in B) samples from $n = 900$ to $n = 1200$.

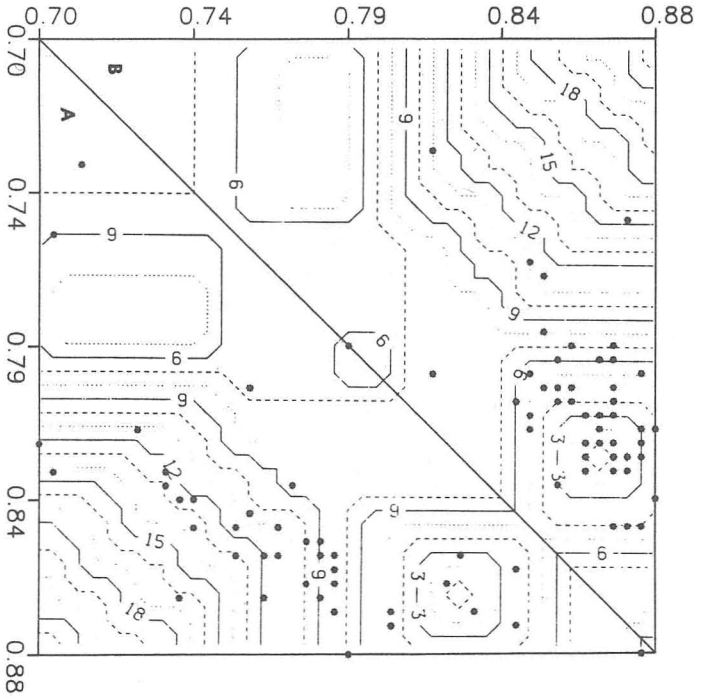

Figure 6: Particular of the dynamics in the case of figure 5. The coordinates are now $p_3 \leftrightarrow q_4$. The section plan is defined by: $q_1 = 0$, $p_1 = 30$, $q_2 = 50$, $p_2 = 81$, $q_3 = 162$, $p_4 = 215$. The exact minimum corresponds to the point $p_3 = 200$, $q_4 = 210$, (in normalized units (.823,.864)). A deep local minimum around (.72,.76) is evident. The trajectory scans both regions in the first 300 iterations (A), and eventually singles out the absolute minimum (B) in the last 300.

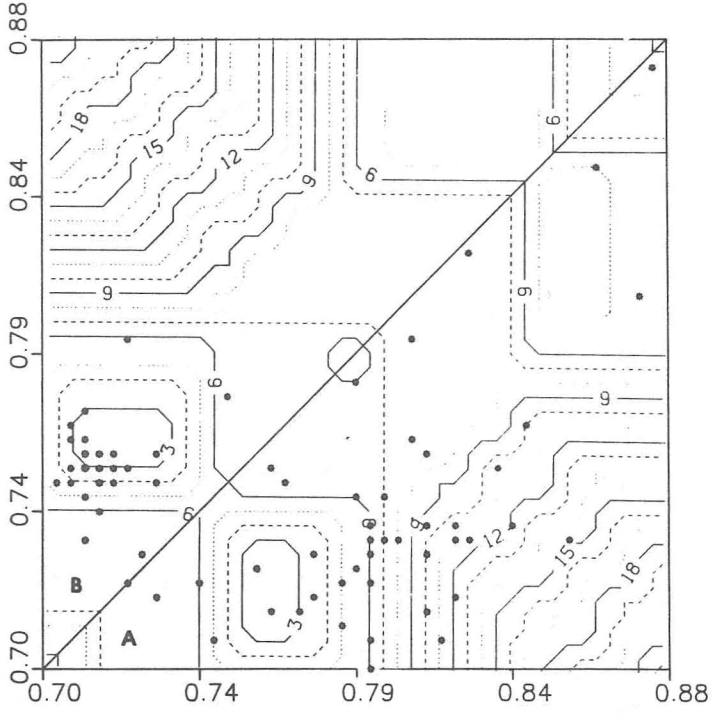

Figure 7: The section plane of figure 6 is here slightly moved varying $q_3 = 164$, $p_4 = 220$, to evidentiate the bottom of the local minimum at (.71,.76). The absolute minimum has almost disappeared. A trajectory in the basin of attraction of the local minimum is shown. The other parameters are the same as in figure 6.

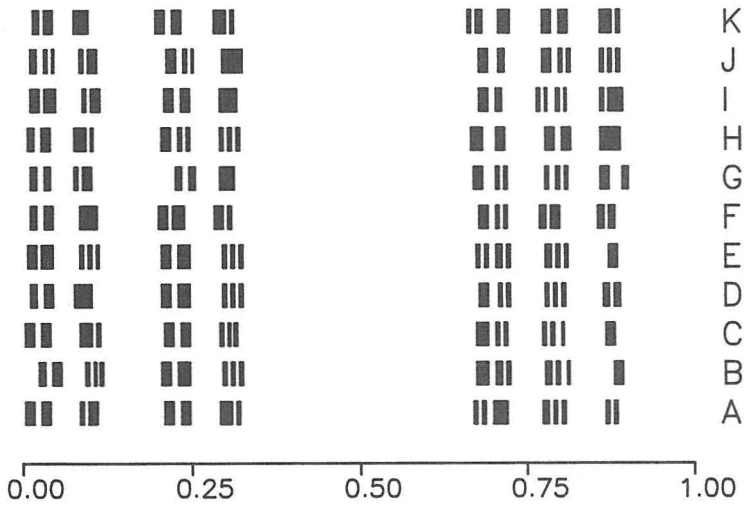

Figure 8: The solution attractors from table 2 B-K are shown, and compared with the exact solution here denoted with A.

fixing the parameters of the first two maps to their actual values we initially searched in the four-dimensional subspace of the remaining parameters. To achieve convergence to the global minimum 6000 iterations were required. We then freed the frozen parameters and ran the annealing algorithm in the full, eight-dimensional parameter space. Convergence of the process to the exact solution was not observed in 50,000 iterations.

On the contrary, the method of chaotic minimization is not affected in a significant way, as table 1 shows eloquently; moreover, no attempt has been made to further reduce $N_{\max}$. The only remarkable consequence is the fact that the basins of attraction of the global and local minima share a comparable amount of phase space: 60% and 40% approximately, according to our particle sampling; a small number of iterations is nonetheless sufficient to identify the two. In figures 6 and 7 we see the dynamics of two trajectories attracted in the different minima. In our limited (though effective) particle-tracking, six trajectories were found to end in the neighborhood of the absolute minimum. Their attractors are shown in figure 8 as $B - G$; the four solutions in the "wrong" minimum correspond to the sets $H - K$.

A further interesting feature of the IFS representation may be seen via table 2, which reports more extensively the data of the attractors in figure 8. Iterated function systems can be used to compute the fractal dimension of digitized images. One may in fact prove [23] that the fractal dimension (or capacity) C of the attractor of an IFS is the unique solution of the equation

Set	q_1	p_1	q_2	p_2	q_3	p_3	q_4	p_4	h	C
A	0	30	50	81	162	200	210	215	0	.601
B	5	31	49	82	163	201	213	217	5	.589
C	0	31	50	80	163	199	210	214	1	.590
D	2	27	49	82	164	200	209	216	2	.597
E	1	30	49	82	163	201	211	215	1	.598
F	2	30	48	78	164	198	207	214	3	.593
G	2	27	54	78	162	201	208	220	4	.600
H	1	28	49	81	161	175	188	219	2	.615
I	2	31	50	80	164	174	185	220	2	.611
J	2	30	51	82	164	175	187	219	2	.607
K	3	26	47	79	160	178	187	219	3	.620

Table 2: Results of chaotic optimization for the set A of figure 8. Reported are the the IFS map parameters $[q_j, p_j]$, the Hausdorff distance from the target set h, and the fractal dimension (capacity) C of the solution attractors. B-G are the results of six trajectories ending close to the absolute minimum, while H-K correspond to attractors in the neighborhood of the local minimum.

$$\sum_{j=1}^{M} \delta_j^C = 1. \tag{7.1}$$

In table 2 we see that the estimated capacity C from approximate solutions of the inverse problem is a very good approximation to the actual.

8. Fractal interpolation functions

As classical analysis provides algorithms to interpolate smooth functions, IFS theory and the method of chaotic optimization can be translated into an algorithm to reproduce in a concise way sets of irregular data. The theory of IFS may be developed over $C = [0,1]^2$ by the class of affine maps $w_i : C \to C$ of the form

$$w_i \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_i & 0 \\ b_i & \delta_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c_i \\ d_i \end{pmatrix}. \tag{8.1}$$

The real parameters $a_i, \ldots, d_i$ $i = 1, \ldots, N$ can be algebraically determined in such a way as to exactly interpolate a set of data in $N+1$ points. The contraction factors δ_i, $0 \leq \delta_i < 1$ determine the fractal dimension of the interpolating graph: setting all δ_i to zero results in a piecewise linear interpolation. See reference [24] for further details. The number of free parameters is hence $2 \times N + 1$: $N+1$ interpolating points and N contraction factors.

Choosing the l^2 metrics as approximation control, we face a problem similar to the one we previously discussed. We tested the IFS scheme and the chaotic minimization algorithm on a set of data of natural origin: the sea floor topography of the Mathematician Ridge (!) in the East Pacific Rise

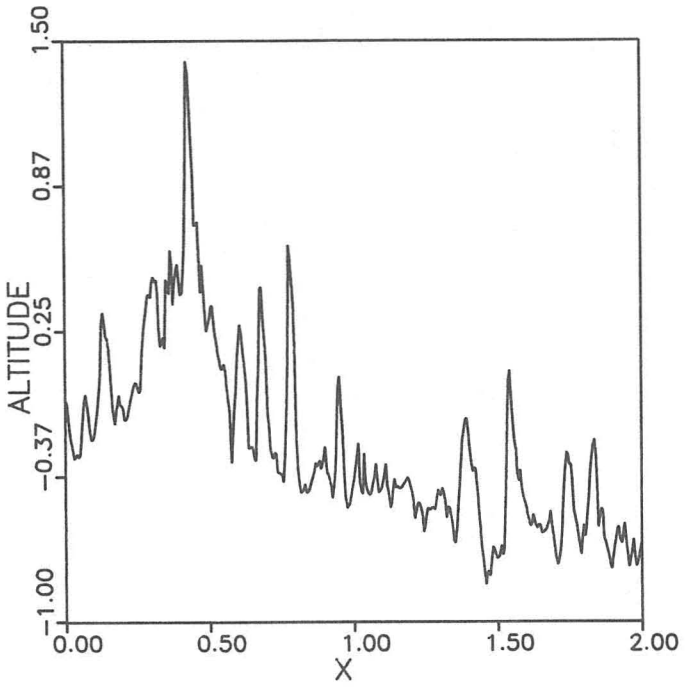

Figure 9: Sea floor topography profile. The unit on abscissa is 100 Km, while the altitude is in units of 1 Km. The reference level has been set at −3.0 Km, under sea level. Recent geological processes have ruggedly shaped this profile.

[25]. Figure 9 shows the one-dimensional profile on its full expansion: the horizontal scale runs over 200 kilometers. The discretization is made on 400 intermediate locations.

Fractal interpolation with 11 points (corresponding to a 95% information compression rate) is shown in figure 10. Ten maps have been employed, implying a 21-dimensional parameter space. The parameters of the chaotic minimization are: $N_{\max} = 2000$ iterations have been computed, with $\lambda_0 = 3.$, $\lambda_N = 10^{-4}$, $\gamma = .01$. The average l^2 distance in the initial steps was .68, decaying to .38 in the last 20 iterations. Only two different trajectories have been evolved, with equivalent results. The fractal dimension of the curve as estimated by IFS is $C = 1.613$.

9. Conclusions

We have shown in this paper that the combination of two major techniques, iterated function systems, and chaotic minimization of non-convex functions can be effectively used in the representation of fractals.

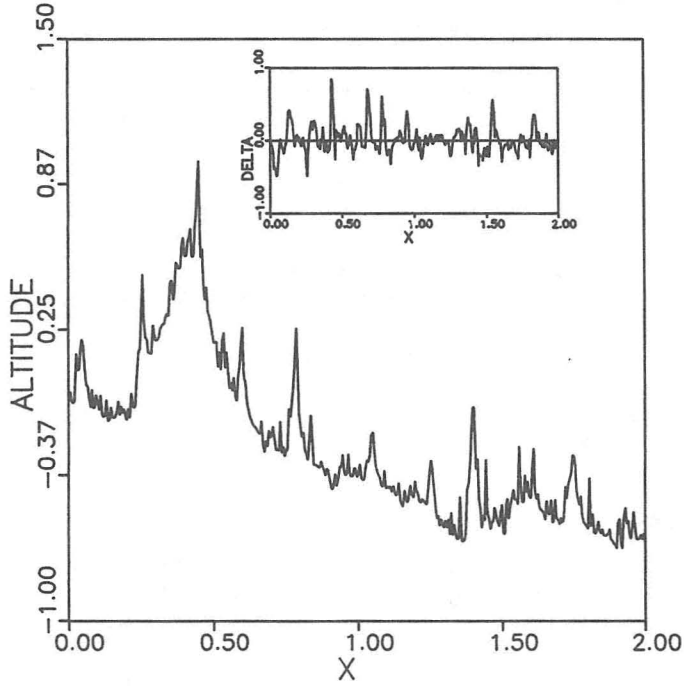

Figure 10: Fractal interpolation of the data of figure 10. The absolute difference original-interpolation is shown in the frame. The l^2 distance between the two is .38, while the l^∞ distance is .85.

We tried to argue that the adoption of an indirect way of solution of the inverse problem is motivated by its inherent complexity. We are nonetheless convinced that further progress can be made in the direct solution process, most likely by providing estimates which would restrict the region of parameter space where to look for an optimal solution.

The method of chaotic optimization, which we introduced here, might also be improved introducing more profitable dynamical models than the one we employed. We are confident that in many instances this method can efficiently supercede STA.

The advantage of fractal geometry rests upon the wide variety of geometrical objects that it can describe, such as high-complexity structures. For instance, the discrete formulation developed in this paper could be used to code genetic characters or even economical patterns — structures traditionally lacking a convenient mathematical representation. Also, the parameters of the "compressed IFS code" may be of significance in detecting the general characteristics of experiment-generated fractals (for instance, spatio-temporal sequences of a complex dynamical system). Finally, the theory exposed in this paper can be of practical utility in coding the information to be transmitted from a remote detector, or stored in data-banks with high storage requirements.

More than all this, though, we think that the general theoretical problem of "laying bridges" over the still unexplored river of computable structures with high information content is a basic challenge in many branches of contemporary science.

Acknowledgments

G.M. acknowledges support from N.S.F. under grant RII-8604334. The Pittsburgh Supercomputing Center provided computational facilities for part of this research.

References

[1] B. Mandelbrot, *The fractal geometry of nature*, (San Francisco: W.H. Freeman, 1982).

[2] E.N. Lorenz, "Deterministic non-periodic flows", *J. Atmos. Sci.*, **20** (1963) 130–141.

[3] M. Henon, "A two-dimensional mapping with a strange attractor", *Commun. Math. Phys.* **50** (1976) 69–77.

[4] J. Hutchinson, "Fractals and self-similarity", *Indiana J. Math.* **30** (1981) 713–747.

[5] M.F. Barnsley and S.G. Demko, "Iterated function systems and the global construction of fractals", *Proc. R. Soc. London* **A 399** (1985) 243–275.

[6] P.M. Löf, "The definition of random sequence", *Info. and Control* **9** (1966) 602–619.

[7] M. Casdagli, "Nonlinear prediction of chaotic time series", (1988) (to appear in *Physica* **D**).

[8] J.P. Crutchfield and B.S. McNamara, "Equations of motion from a data series", *Complex Systems*, **1** (1987) 417–452.

[9] A.K. Zvonkin, and L.A. Levin, "The complexity of finite objects", *Russ. Math. Surv.* **25** (1970) 83–124.

[10] G.J. Chaitin, *Information, Randomness and Incompleteness*, (Singapore: World Scientific, 1988).

[11] P. Diaconis, M. Shahshahani,"Products of random matrices and computer image generation", in *Proc. AMS-IMS-SIAM Conf. on Random Matrices 1984.* (Providence, RI: American Math. Soc., 1986).

[12] M.F. Barnsley, V. Ervin, D. Hardin, and J. Lancaster, "Solution of an inverse problem for fractals and other sets", *Proc. Nat. Acad. Science USA* **83** (1986) 1975–1977.

[13] P. Grassberger and I. Procaccia, "Dimensions and entropies of strange attractors from a fluctuating dynamics approach", *Physica* **13D** (1984) 34.

[14] M.F. Barnsley and A. Sloan, "A better way of compressing images", *Byte*, **13** (1988) 215–223.

[15] M.R. Garey, D.S. Johnson, *Computers and intractability: A guide to the theory of NP-completeness* (New York: W.H. Freeman, 1979).

[16] S. Kirkpatrick, C.D. Gelatt and P.M. Vecchi, "Optimization by simulated annealing", *Science*, **220** (1983) 671–680.

[17] M.F. Barnsley, "Image compression by simulated annealing", Proposal to DARPA (1986).

[18] B. Gidas, "Nonstationary Markov chains and convergence of the annealing algorithm", *J. Stat. Phys.* **39** (1984) 73–131.

[19] S. Geman and G. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images", *IEEE PAMI*, **6** (1984) 771–741.

[20] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing", *Proc.* 24^{th} *Conf. on Decision and Control* (1985) 761–767.

[21] H. Szu and R. Hartley, "Fast simulated annealing", *Phys. Lett.* **A 122** (1987) 157.

[22] A.A. Brudno, "The complexity of the trajectory of a dynamical system", *Russ. Math. Surv.* **33** (1978) 197–198.

[23] K.J. Falconer, "The Hausdorff dimension of some fractals and attractors of overlapping construction", *J. Stat. Phys.* **47** (1987) 123–132.

[24] M.F. Barnsley, *Fractals Everywhere*, (New York: Academic Press, 1988).

[25] J.C. Mareschal, "Fractal reconstruction of sea floor topography", (1988) (to appear in *J. Pure and App. Geoph.*).