Iterated Function Systems

Section 1. Iterated Function Systems

Definition. A transformation $f:R^m \to R^m$ is a **contraction** (or is a contraction mapping or is contractive) if there is a constant s with $0 \le s < 1$ such that

$|f(x)-f(y)| \le s\,|x-y|$ for all x and y in R^m.

We call any such number s a **contractivity factor** for f. The smallest s that works is called the **best contractivity factor** for f.

Example. Let $f:R \to R$ by $f(x) = 0.9x + 2$. Then f is a contraction mapping because $|f(x)-f(y)| = |0.9x - 0.9y| = 0.9|x-y|$. We may thus choose s = 0.9. [For that matter we could also choose s = 0.95 but not s = 1.05 or s = 0.85. We generally pick the smallest possible choice for s, which is 0.9 in this example. Thus s = 0.95 would not get full credit.]

Lemma. Let $f: R^2 \to R^2$ be a similarity, $f(x) = r\,x + t$ with $r > 0$. This combines a rescaling by r with a translation by the vector t = (t1, t2). Then f is a contraction provided $r < 1$. When this occurs, the best contractivity factor is r.
Proof. $|f(x) - f(y)| = |r\,x+t -(r\,y+t)| = |r\,x - r\,y| = r\,|x - y|$.
QED

This gives lots of examples of contractions on R^2.

Theorem. Suppose $f: R \to R$ is differentiable everywhere and $|f'(x)| \le s < 1$ for all $x \in R$. Then f is a contraction with contractivity factor s.
Example. $f(x) = 0.5 \cos x$ is a contraction on R.
Proof. By the Mean Value Theorem,

$|f(x) - f(y)| / |x-y| = |f'(c)| \le s$ for some c between x and y. But then $|f(x) - f(y)| \le s\,|x-y|$. QED

Definition. An **iterated function system** (IFS) is a set of continuous maps $w_1:R^m \to R^m$, $w_2:R^m \to R^m$, ..., $w_N: R^m \to R^m$ together with a set of positive numbers p_1, ..., p_N which add up to 1 (interpreted as probabilities). .

The IFS is **hyperbolic** if each w_i is a contraction map with contractivity factor s_i . Let $s = \max\{s_1, s_2, ..., s_N\}$.

Example. On R, $w_1(x) = (1/3)x$, $w_2(x) = (1/3)x+(2/3)$. Let $p_1 = 1/2$, $p_2 = 1/2$. This is a hyperbolic IFS.

Def. Given an IFS, let $H(R^m)$ be the collection of compact nonempty subsets of R^m. More explicitly, these are the subsets of R^m which are bounded, closed, nonempty. We define the **set transformation**

$W: H(R^m) \to H(R^m)$

by $W(B) = w_1(B) \cup w_2(B) \cup ... \cup w_N(B)$.
We say that the **contractivity factor** of W is $s = \max\{s_1, s_2, ..., s_N\}$.

Example. On R, $w_1(x) = (1/3)x$, $w_2(x) = (1/3)x+(2/3)$. Let $p_1 = 1/2$, $p_2 = 1/2$. Then $W(B) = w_1(B) \cup w_2(B)$ has contractivity factor (1/3).

Definition. If $A \in H(R^m)$ for a hyperbolic IFS is fixed under W (ie., $W(A) = A$), then A is called the **attractor** of the IFS.

Example. On R, $w_1(x) = (1/3)x$, $w_2(x) = (1/3)x+(2/3)$. The attractor is the Cantor set C, since $W(C) = C$.

Typically the attractor is a fractal.

Contraction Mapping Theorem. Let $w_1, ..., w_N$ be a hyperbolic IFS on R^m with contractivity factor s. Then the IFS possesses exactly one attractor A. Suppose B consists of a fixed point p of some w_i (ie, $B = \{p\}$). Then A consists of all the limits of all converging sequences $x_0, x_1, x_2, ...$ such that for all i, x_i lies in $W^i(B)$.

In fact, if B is any member of $H(R^m)$ then the sequence $\{W^n(B)\}$ converges to A. [The definition of what it means for a sequence of sets to converge to a set is complicated.]

Section 2. Drawing the attractor of an IFS

METHOD 1: THE "DETERMINISTIC ALGORITHM".

Choose any fixed point p of some wi. Let $A_0 = \{p\} \in H(R^2)$. Draw A_0. Draw $A_1 = W(A_0)$. Draw $A_2 = W(A_1)$. In general, we let $A_{j+1} = W(Aj)$ and we draw Aj's until we are satisfied.

Example. Consider again the Cantor set example. Let $w_1(x) = (1/3)x$, $w_2(x) = (1/3)x+(2/3)$. Choose $B = \{0\}$ and consider $W^n(B)$.

$B = \{0\}$
$W(B) = \{0,2/3\}$
$W^2(B) = \{0, 2/9, 2/3, 8/9\}$
$W^3(B) = \{0,2/27,6/27,8/27,18/27,20/27,24/27,26/27\}$.

In practice this means that there is a finite set A_0 corresponding to pixels that are darkened. We draw A_0. For example, if A_0 consists of a single point p, then $A_0 = \{p\}$, $A_1= \{w_1(p), ..., w_N(p)\}$, $A_2 = \{w_i(w_j(p))\}$. In practice you make a linked list of the points to be drawn. This does not use the probabilities.

Example. The Sierpinski triangle S. Let $w_1(x) = (1/2)x$, $w_2(x) = (1/2)x + (0,1/2)$, $w_3(x) = (1/2)x + (1/2,0)$. Then W has contractivity factor 1/2. The attractor A satisfies $W(A) = A$, so $A = S$. Use program ListDraw for the Sierpinski triangle

Why does this work?
Lemma. A fixed point p of w_1 lies in A.
Proof. Let $B = \{p\}$.
Then $x_0 = p \in B$
$x_1 = w_1(p) = p \in W(B)$
$x_2 = w_1(x_1) = p \in W^2(B)$
Thus the sequence $x_i = p$ satisfies that $x_i \in W^i(B)$, whence p is a limit point.

Hence $p \in A$.
QED

Cor. For each n, $W^n(B) \subseteq A$.
Proof.
(1) For each i, $w_i(p) \in w_i(A) \subseteq W(A) = A$. Hence $W(B) \subseteq A$.
(2) Each point of $W^2(B)$ has the form $w_i(q)$ for $q \in W(B)$. But $q \in A$. Hence $w_i(q) \in w_i(A) \subseteq W(A) = A$. Hence $W^2(B) \subseteq A$.
(3) The general case follows by induction.
QED

By the Contraction Mapping Theorem, A is the set of limit points from $\cup W^n(B)$. On the computer you can't distinguish limit points. Hence the method draws a picture indistinguishable on the computer from A.

Method 1 leads to lots of redundancy since the same calculations are performed lots of times.

METHOD 2: THE "RANDOM ITERATION METHOD"

Give each map w_i a probability p_i with $0<p_i<1$ but $\sum p_i = 1$. Let x_0 be a point in A (for example, the fixed point of w_1). Draw x_0. Pick a map w_i at random (so w_i is chosen with probability p_i). Let $x_1 = w_i(x_0)$ and draw x_1. Pick a map wj at random and let $x_2 = wj(x_1)$; draw x_2. Repeat this.

This goes very fast. You don't need to draw many extraneous points. You don't have the overhead of keeping a long linked list.

Show MultiBarnsley for the Sierpinski triangle. For the Sierpinski triangle, we get good results if $p_1 = 0.33$, $p_2 = 0.33$, $p_3 = 0.34$. Note the weird results, however, if $p_1 = 0.66$, $p_2 = 0.30$, $p_3 = 0.04$. This means that the second and third maps are rarely used, so the detail doesn't fill in. Since p_3 is small, it is rare that we draw points in $w_3(A)$.

Frequently one sees the description to let x_0 be any point (not necessarily a point of A); but then one gets the additional complication of omitting the first few (maybe 10) points from the drawing since they will not be in A. (After a while, the points will be so close to A that it will not matter for the picture whether the point was actually in A; telling how many iterates to wait may be a bit complicated sometimes.)

Example. Consider again the Cantor set example. Let $w_1(x) = (1/3)x$, $w_2(x) = (1/3)x+(2/3)$. Let $p_1 = 1/2$, $p_2 = 1/2$.

Suppose instead we start with 0. Repeatedly flip a coin, and use w_1 if Heads, w_2 if Tails. The points we obtain are all in A and can get close to each point in A.

Why does the method work?

Lemma. Each x_i lies in A.
Proof. $x_0 \in A$ by definition.
$x_1 = wj(x_0)$ for some j, hence lies in $W(A) = A$, hence $x_1 \in A$.
$x_2 = wj(x_1)$ for some j, hence lies in $W(A) = A$, hence $x_2 \in A$.
etc.

QED

The convergence is a matter of probability (true only with high probability), but in any event every point drawn is in A.

Rule of Thumb for choosing the probabilities p_i: Make p_i proportional to the estimated area of $w_i(A)$. Thus for the Sierpinski triangle, the 3 pieces $w_i(A)$ should be the same size, so they should have the same probabilites. Hence $p_i = 1/3$.

Section 3. The Collage Theorem.

There remains the important problem, given a proposed attractor A, of finding an IFS that has A as its attractor. This is accomplished by the Collage Theorem:

Let $B \in H(R^m)$. Let $\varepsilon > 0$.
$B+\varepsilon = \{x \in R^m$: there exists $b \in B$ with $|x - b| < \varepsilon \}$ = B fattened by ε
Example. B = line segment
Example. B = edge of a square

Collage Theorem.
Let $L \in H(R^m)$. [This is what you hope to be the attractor.]
Let $\varepsilon \geq 0$ be given. [This is the allowed resolution.]
Suppose [by hook or crook] there is an IFS $w_1, w_2, \ldots, w_N$ with contractivity factor s, $0 \leq s < 1$, such that

$W(L) \subseteq (L + \varepsilon)$
$L \subseteq (W(L) + \varepsilon)$

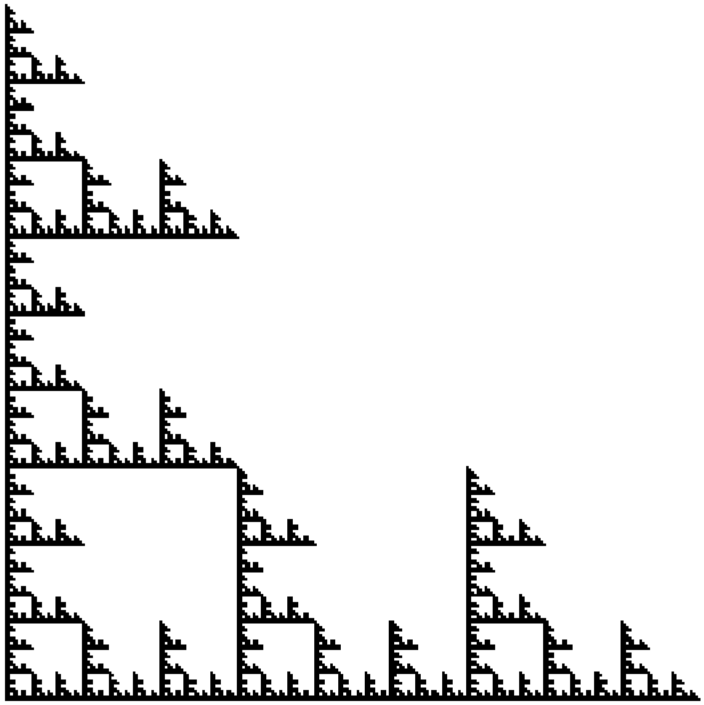
Then the attractor A of the IFS satisfies:

(1) $A \subseteq (L + \varepsilon/(1-s))$
(2) $L \subseteq (A + \varepsilon/(1-s))$

Hence, if ε is small, A looks very much like L; A is contained in a fattened version of L and L is contained in a fattened version of A.

Example 1. Let L be the Sierpinski triangle. (Use overhead.) Let $w_1(x) = (1/2)x$, $w_2(x) = (1/2)x + (0,1/2)$, $w_3(x) = (1/2)x + (1/2,1/2)$. Then A = L.

Example.
Tell the maps for a hyperbolic IFS that draws the fractal below. Assume the bottom is [0,1].



Answer:
w1(x) = (1/3) x
w2(x) = (1/3)x + (1/3,0)
w3(x) = (1/3)x + (2/3,0)
w4(x) = (1/3) x + (0, 1/3)
w5(x) = (1/3)x + (0, 2/3)

Section 4. Transformations on R^2

Some types of transformations are used so much that they have names.

A map w: $R^m \to R^m$ by w(x) = x + t where t = is a fixed vector. The effect of w is to move the vector over by t. We say w **translates** by t.

A map w: $R^m \to R^m$ by w(x) = r x **rescales** by the scale factor r.

A map w: $R^m \to R^m$ by w(x) = rx + t first rescales, then translates by t.

A map w: $R^2 \to R^2$ by $w(x_1,x_2) = (x_1, -x_2)$. We say that w **reflects** in the x_1 axis.
We can rewrite w using matrices: Let R denote the matrix corresponding to reflection in the x_1 axis:

R = [1 0]
[0 -1]

Then w(x) = R x using matrix multiplication.

A map w: $R^2 \to R^2$ by

$$w(x) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} x = R_\theta\, x$$

w corresponds to counterclockwise **rotation** about the origin by the angle θ.
Here R_θ denotes the matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Ex. Rotate by 45 degrees. Show the unit square so rotated.

Ex. To rotate clockwise 45 degrees, instead you rotate counterclockwise by -45 degrees.

A **similitude** on R^2 is a transformation w: $R^2 \to R^2$ that has either of the forms:

$$w(x) = r\, R_\theta\ (x) + v$$

or $$w(x) = r\, R_\theta\, R\ (x) + v$$

where we shall insist that the constant r satisfies r >0.

In the first, this says take x, rotate by θ, then rescale by r, then translate by v. The second says take x, reflect in the x axis, rotate by θ, rescale by r, then finally translate by v.

Example. Find the formula for the similitude that first rotates 45 degrees and then translates by (1,2)T.

Solution. $$f(x) = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} x$$

and g(x) = x + (1,2). Hence what we want is g(f(x))

$$= \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$