Math 118, Spring 2,001

Shlomo Sternberg

March 13, 2001

Contents

Chapter 6

Hutchinson's theorem and fractal images.

6.1 The Hausdorff metric and Hutchinson's theorem.

Let X be a complete metric space. Let $\mathcal{H}(X)$ denote the space of non-empty compact subsets of X. For any $A \in \mathcal{H}(X)$ and any positive number ϵ, let

$$A_\epsilon = \{x \in X | d(x, y) \leq \epsilon, \text{for some } y \in A\}.$$

We call A_ϵ the ϵ-*collar* of A. Recall that we defined

$$d(x, A) = \inf_{y \in A} d(x, y)$$

to be the distance from any $x \in X$ to A, then we can write the definition of the ϵ-collar as

$$A_\epsilon = \{x | d(x, A) \leq \epsilon\}.$$

Notice that the infimum in the definition of $d(x, A)$ is actually achieved, that is, there is some point $y \in A$ such that

$$d(x, A) = d(x, y).$$

This is because A is compact. For a pair of non-empty compact sets, A and B, define

$$d(A, B) = \max_{x \in A} d(x, B).$$

So

$$d(A, B) \leq \epsilon \ \ iff \ \ A \subset B_\epsilon.$$

Notice that this condition is not symmetric in A and B. So Hausdorff introduced

$$\begin{aligned} h(A, B) &= \max\{d(A, B), d(B, A)\} & (6.1)\\ &= \inf\{\epsilon \mid A \subset B_\epsilon \ \text{ and } \ B \subset A_\epsilon\}. & (6.2) \end{aligned}$$

as a distance on $\mathcal{H}(X)$. He proved

Proposition 6.1.1 *The function h on $\mathcal{H}(X) \times \mathcal{H}(X)$ satsifies the axioms for a metric and makes $\mathcal{H}(X)$ into a complete metric space. Furthermore, if*

$$A, B, C, D \in \mathcal{H}(X)$$

then

$$h(A \cup B, C \cup D) \leq \max\{h(A, C), h(B, D)\}. \tag{6.3}$$

Proof. We begin with (6.3). If ϵ is such that $A \subset C_\epsilon$ and $B \subset D_\epsilon$ then clearly $A \cup B \subset C_\epsilon \cup D_\epsilon = (C \cup D)_\epsilon$. Repeating this argument with the roles of A, C and B, D interchanged proves (6.3).

We prove that h is a metric: h is symmetric, by definition. Also, $h(A, A) = 0$, and if $h(A, B) = 0$, then every point of A is within zero distance of B,and hence must belong to B since B is compact, so $A \subset B$ and similarly $B \subset A$. So $h(A, B) = 0$ implies that $A = B$.

We must prove the triangle inequality. For this it is enough to prove that

$$d(A, B) \leq d(A, C) + d(C, B),$$

because interchanging the role of A and B gives the desired result. Now for any $a \in A$ we have

$$\begin{aligned} d(a, B) &= \min_{b \in B} d(a, b) \\ &\leq \min_{b \in B} \; (d(a, c) + d(c, b) \;\; \forall c \in C \\ &= d(a, c) + \min_{b \in B} d(c, b) \;\; \forall c \in C \\ &= d(a, c) + d(c, B) \;\; \forall c \in C \\ &\leq d(a, c) + d(C, B) \;\; \forall c \in C. \end{aligned}$$

The second term in the last expression does not depend on c, so minimizing over c gives

$$d(a, B) \leq d(a, C) + d(C, B).$$

Maximizing over a on the right gives

$$d(a, B) \leq d(A < C) + d(C, B).$$

Maximizing on the left gives the desired

$$d(A, B) \leq d(A, C) + d(C, A).$$

We sketch the proof of completeness. Let A_n be a sequence of compact non-empty subsets of X which is Cauchy in the Hausdorff metric. Define the set A to be the set of all $x \in X$ with the property that there exists a sequence of points $x_n \in A_n$ with $x_n \to x$. It is straighforward to prove that A is compact and non-empty and is the limit of the A_n in the Hausdorff metric.

Suppose that $K : X \to X$ is a contraction. Then K defines a transformation on the space of subsets of X (which we continue to denote by K):

$$K(A) = \{Kx | x \in A\}.$$

Since K continuous, it carries $\mathcal{H}(X)$ into itself. Let c be the Lipschitz constant of K. Then

$$\begin{aligned} d(K(A), K(B)) &= \max_{a\in A}[\min_{b\in B} d(K(a), K(b))] \\ &\leq \max_{a\in A}[\min_{b\in B} cd(a, b)] \\ &= cd(A, B). \end{aligned}$$

Similarly, $d(K(B), K(A)) \leq c\ d(B, A)$ and hence

$$h(K(A), K(B)) \leq c\ h(A, B). \tag{6.4}$$

In other words, a contraction on X induces a contraction on $\mathcal{H}(X)$.

The previous remark together with the following observation is the key to Hutchinson's remarkable construction of fractals:

Proposition 6.1.2 *Let $T_1, \ldots, T_n$ be a collection of contractions on $\mathcal{H}(X)$ with Lipschitz constants $c_1, \ldots, c_n$, and let $c = \max c_i$. Define the transformation T on $\mathcal{H}(X)$ by*

$$T(A) = T_1(A) \cup T_2(A) \cup \cdots \cup T_n(A).$$

Then T is a contraction with Lipschitz constant c.

Proof. By induction, it is enough to prove this for the case $n = 2$. By (6.3)

$$\begin{aligned} h(T(A), T(B)) &= h(T_1(A) \cup T_2(A), T_1(B) \cup T_2(B)) \\ &\leq \max\{h(T_1(A), h(T_1(B)), h(T_2(A), T_2(B))\} \\ &\leq \max\{c_1 h(A, B), c_2 h(A, B)\} \\ &= h(A, B) \max\{c_1, c_2\} = c \cdot h(A, B) \end{aligned}$$

.

Putting the previous facts together we get Hutchinson's theorem;

Theorem 6.1.1 *Let $K_1, \ldots, K_n$ be contractions on a complete metric space and let c be the maximum of their Lifschitz contants. Define the Hutchinson operator, K, on $\mathcal{H}(X)$ by*

$$K(A) = K_1(A) \cup \cdots \cup K_n(a).$$

Then K is a contraction with Lipschtz constant c.

6.2 Affine examples

We describe several examples in which X is a subset of a vector space and each of the T_i in Hutchinson's theorem are affine transformations of the form

$$T_i : x \mapsto A_i x + b_i$$

where $b_i \in X$ and A_i is a linear transformation.

6.2.1 The classical Cantor set.

Take $X = [0, 1]$, the unit interval. Take

$$T_1 : x \mapsto \frac{x}{3}, \quad T_2 : x \mapsto \frac{x}{2} + \frac{2}{3}.$$

These are both contractions, so by Hutchinson's theorem there exists a unique closed fixed set C. This is the Cantor set.

To relate it to Cantor's original construction, let us go back to the proof of the contraction fixed point theorem applied to T acting on $\mathcal{H}(X)$. It says that if we start with any non-empty compact subset A_0 and keep applying T to it, i.e. set $A_n = T^n A_0$ then $A_b \to C$ in the Hausdorff metric, h. Suppose we take the interval I itself as our A_0. Then

$$A_1 = T(I) = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1].$$

in other words, applying the Hutchinson operator T to the interval $[0, 1]$ has the effect of deleting the "middle third" open interval $(\frac{1}{3}, \frac{2}{3})$. Applying T once more gives

$$A_2 = T^2[0, 1] = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1].$$

In other words, A_2 is obtained from A_1 by deleting the middle thirds of each of the two intervals of A_1 and so on. This was Cantor's original construction. Since $A_{n+1} \subset A_n$ for this choice of initial set, the Hausdorff limit coincides with the intersection.

But of course Hutchinson's theorem (and the proof of the contractions fixed point theorem) says that we can start with *any* non-empty closed set as our initial "seed" and then keep applying T. For example, suppose we start with the one point set $B_0 = \{0\}$. Then $B_1 = TB_0$ is the two point set

$$B_1 = \{0, \frac{2}{3}\},$$

B_2 consists of the four point set

$$B_2 = \{0, \frac{2}{9}, \frac{2}{3}, \frac{8}{9}\}$$

and so on. We then must take the Hausdorff limit of this increasing colletion of sets. To describe the limiting set c from this point of view, it is useful to use triadic expansions of points in $[0, 1]$. Thus

$$\begin{aligned} 0 &= .0000000\cdots \\ 2/3 &= .2000000\cdots \\ 2/9 &= .0200000\cdots \\ 8/9 &= .2200000\cdots \end{aligned}$$

and so on. Thus the set B_n will consist of points whose triadic exapnsion has only zeros or twos in the first n positions followed by a string of all zeros. Thus a point will lie in C (be the limit of such points) if and only if it has a triadic expansion consisting entirely of zeros or twos. This includes the possibility of an infinite string of all twos at the tail of the expansion. for example, the point 1 which belongs to the Cantor set has a triadic expansion $1 = .222222\cdots$. Simialrly the point $\frac{2}{3}$ has the triadic expansion $\frac{2}{3} = .0222222\cdots$ and so is in the limit of the sets B_n. But a point such as $.101\cdots$ is not in the limit of the B_n and hence not in C. This description of C is also due to Cantor. Notice that for any point a with triadic expansion $a = .a_1a_2a_2\cdots$

$$T_1a = .0a_1a_2a_3\cdots, \quad \text{while} \quad T_2a = .2a_1a_2a_3\cdots.$$

Thus if all the entries in the expansion of a are either zero or two, this will also be true for T_1 and T_2a. This shows that the C (given by this second Cantor description) satisfies $TC \subset C$. On the other hand,

$$T_1(.a_2a_3\cdots) = .0a_2a_3\cdots, \qquad T_2(.a_2a_3\cdots) = .2a_2a_3\cdots$$

which shows that $.a_1a_2a_3\cdots$ is in the image of T_1 if $a_1 = 0$ or in the image of T_2 if $a_1 = 2$. This shows that $TC = C$. Since C (according to Cantor's second description) is closed, the uniqueness part of the fixed point theorem guarantees that the second description coincides iwth the first.

The statement that $TC = C$ implies that C is "self-similar".

6.2.2 The Sierpinski Gasket

Consider the three affine transformations of the plane:

$$T_1 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}, \quad T_2 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$T_3 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The fixed point of the Hutchinson operator for this choice of T_1, T_2, T_3 is called the Sierpinski gasket, S. If we take our initial set A_0 to be the right triangle with vertices at

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

then each of the T_iA_0 is a similar right triangle whose linear dimensions are one-half as large, and which shares one common vertex with the original triangle. In other words,

$$A_1 = TA_0$$

is obtained from our original triangle be deleting the interior of the (reversed) right triangle whose vertices are the midpoints of our origninal triangle. Just as in the case of the Cantor set, successive applications of T to this choice of original set amounts to sussive deletions of the "middle" and the Hausdorff limit is the intersection of all them: $S = \bigcap A_i$.

We can also start with the one element set

$$B_0 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\}$$

Using a binary expansion for the x and y coordinates, application of T to B_0 gives the three element set

$$\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ .1 \end{pmatrix} \right\}.$$

The set $B_2 = TB_1$ will contain nine points, whose binary expansion is obtained from the above three by shifting the x and y exapnsions one unit to the right and either inserting a 0 before both expansions (the effect of T_1), insert a 1 before the expansion of x and a zero before teh y or vice versa. Proceding in this fashion, we see that B_n consists of 3^n points which have all 0 in the binary expansion of the x and y coordinates, past the n-th position, and which are further constrained by the condition that at no earler point do we have both $x_i = 1$ and $y_1 = 1$. Passing to the limit shows that S consists of all points for which we can find (possible inifinite) binary expansions of the x and y coordinates so that $x_i = 1 = y_i$ never occurs. (For example $x = \frac{1}{2}, y = \frac{1}{2}$ belongs to S because we can write $x = .10000\cdots, y = .011111\ldots$). Again, from this (second) description of S in terms of binary expansions it is clear that $TS = S$.