CHAPTER 3

Iterated Function Systems

Throughout this chapter, X will denote a compact subset of $\mathbb{R}^n$ or $\mathbb{C}$ and $\mathcal{H}(X)$ will be its space of fractals, endowed with the Hausdorff metric. The reason for this stipulation is so that the space of fractals $\mathcal{H}(X)$ is itself a compact set under the Hausdorff metric, which is a fact that we present without proof. The most important thing this implies for our purposes is the following.

LEMMA 3.1. *If $\{A_n\}$ is a sequence in $\mathcal{H}(X)$ such that $\lim_{n\to\infty} A_n = A$, then $A \in \mathcal{H}(X)$. In particular, a convergent sequence of nonempty compact subsets of X converges to a nonempty compact subset of X.*

3.1. Collage maps as contractions on the space of fractals

For concreteness let us write down the definition of a contraction mapping in the setting of the space of fractals. In this definition consider $\mathcal{T}$ to be any transformation on $\mathcal{H}(X)$, most often but not always a collage map.

DEFINITION 3.2. We say $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ is a *contraction mapping* if there is a $c \in [0, 1)$ such that for all $A, B \in \mathcal{H}(X)$,

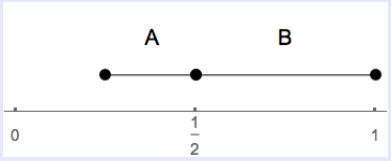
$$d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq c\, d_H(A, B)$$

An elementary contraction mapping on $\mathcal{H}(X)$ is one obtained by a contraction on X. To be precise, let $T : X \to X$ be a contraction mapping with contraction factor $c \in [0, 1)$. Then by definition 1.10, T defines a transformation on $\mathcal{H}(X)$ and we will soon prove that it is a contraction on $\mathcal{H}(X)$ with contraction factor c.

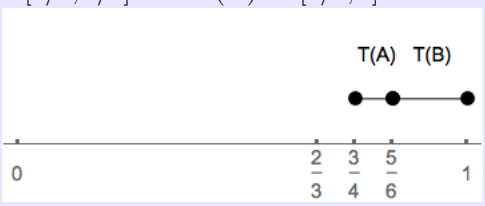
First we should note that whether or not $T : X \to X$ is a contraction, as long as it is continuous it extends to a transformation on $\mathcal{H}(X)$. To see this, consider $T(A)$ for some element $A \in \mathcal{H}(X)$. It is nonempty since A is. The fact that it is compact is a real analysis fact: we all learn the mantra "The continuous image of a compact set is compact". If T is a contraction on X, then it is automatically continuous and thus extends to a transformation of the space of fractals.

EXAMPLE 3.3. Let $X = [0,1]$, $A = [1/4, 1/2]$ and $B = [1/2, 1]$. Suppose that $T : X \to X$ is given by $T(x) = x/3 + 2/3$. We will prove that $d_H(T(A), T(B)) \leq 1/3\, d_H(A, B)$.

First let us compute $d_H(A, B)$. With this handy figure:



we can see that the furthest point in A from B is $1/4$, while the furthest point in B from A is 1, making the largest distance $1/2$ and so $d_H(A, B) = 1/2$. Next, we can compute $T(A) = [3/4, 5/6]$ and $T(B) = [5/6, 1]$. This next handy figure:



lets us compute $d_H(T(A), T(B)$ by noticing that the furthest point in $T(A)$ from $T(B)$ is $3/4$ while the furthest point in $T(B)$ from $T(A)$ is 1, making the largest distance $1/6$. Thus we see that $d_H(T(A), T(B)) = 1/6 = 1/3\, d_H(A, B)$.

Based on this evidence it would seem that the contracting factor for $\mathcal{H}(X)$ seems like it is going to equal the contracting factor on X itself, and that turns out to be true. The previous example is a special case of what happens for a general contraction on X.

LEMMA 3.4. *Suppose T is a contraction on X with contraction factor $c \in [0, 1)$. Then the transformation induced by T on $\mathcal{H}(X)$ is also a contraction mapping with contraction factor c.*

EXERCISE 3.5. The proof of Lemma 3.4 uses the following outline. Once you have figured out how all the steps work, write them into a formal proof.

(1) Consider any $A, B \in \mathcal{H}(X)$ and suppose $d_H(A, B) = \delta$. Explain why for every $a \in A$ there is a $b \in B$ for which $d(a, b) \leq \delta$.
(2) Compute an upper bound on the Euclidean distance between $T(a)$ and $T(b)$.
(3) Come up with an argument for why the previous step gives you a bound on the size of the ϵ-thickening of $T(B)$ required so that $T(A) \subseteq T(B)_\epsilon$. (Spoiler alert: that ϵ should be $c\delta$.)
(4) Make the symmetric argument that computes an epsilon for which $T(B) \subseteq T(A)_\epsilon$.
(5) Put the previous steps together to show that $d_H(T(A), T(B)) \leq c\delta = c\, d_H(A, B)$.

Of course when we are making fractals what we really care about is collages made from contractions. We would like to obtain a lemma similar to the previous lemma that tells us that such a collage is also a contraction, and what the contraction factor is. The following lemmas will help us get started.

LEMMA 3.6. *Let A, B, C, and D be elements of $\mathcal{H}(X)$. Then $d_H(A \cup B, C \cup D) \leq \max\{d_H(A, C), d_H(B, D)\}$.*

PROOF. Let $d_H(A, C) = \delta$ and $d_H(B, D) = \gamma$, and let $\nu = \max\{\delta, \gamma\}$. By definition of $d_H(A, C)$ we know that $A \subseteq C_\delta \subseteq C_\nu$; we also know that $C_\nu \subseteq C_\nu \cup D_\nu$. By Exercise 3.22 we know that $C_\nu \cup D_\nu = (C \cup D)_\nu$, and thus we have that $A \subseteq (C \cup D)_\nu$. Similarly, we see that $B \subseteq D_\gamma \subseteq D_\nu \subseteq C_\nu \cup D_\nu = (C \cup D)_\nu$. Putting these facts together we have that $A \cup B \subseteq (C \cup D)_\nu$.

In the opposite direction we know that $C \subseteq A_\delta \subseteq A_\nu \subseteq (A \cup B)_\nu$, and that $D \subseteq B_\gamma \subseteq B_\nu \subseteq (A \cup B)_\nu$. This shows us that $\nu \geq d_H(A \cup B, C \cup D)$ by the definition of Hausdorff metric using ϵ-thickening. □

This result extends to any finite unions of elements of $\mathcal{H}(X)$ as the next corollary shows. It can be proved by induction on the number of elements in the union.

COROLLARY 3.7. *Let $A_i, B_i \in \mathcal{H}(X)$ for $i = 1, 2, ..., n$ and let $\nu = \max_{1 \leq i \leq n} \{d_H(A_i, B_i)\}$. Then*

$$d_H\left(\bigcup_{i=1}^{n} A_i\,,\, \bigcup_{i=1}^{n} B_i\right) \leq \nu.$$

EXERCISE 3.8. Let $T_1 : X \to X$ be a contraction with contraction factor c_1 and let $T_2 : X \to X$ be a contraction with contraction factor c_2. In this exercise, fill in the blanks to complete a formal proof that the collage map $\mathcal{T} = T_1 \cup T_2$ is a contraction on $\mathcal{H}(X)$.

PROOF. Let $A, B \in \mathcal{H}(X)$. Then $\mathcal{T}(A) =$ _____ and $\mathcal{T}(B) =$ _____ by _____ . For $i = 1, 2$ we know that $d_H(T_i(A), T_i(B)) \leq$ _____ by Lemma _____. Applying Lemma _____ we know that

$$d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq \max\{\text{_____}\}$$
$$\leq \max\{\text{_____}\}$$

This means that $d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq c\, d_H(A, B)$, where c is defined to be _____. Since $c \in [0, 1)$ we have completed the proof that $\mathcal{T}$ is a contraction mapping on $\mathcal{H}(X)$. □

EXERCISE 3.9. Consider the transformations in Exercise 1.16, Exercise 1.17, Exercise 1.30, and Exercise 1.37. Using analysis similar to the proof in Exercise 3.8, determine the contraction factor for each collage map. Write the proof that your collage is a contraction following the form in Exercise 3.8; there are small details that will need to change (for instance, "Lemma" might become "Corollary" in a key place, depending on your number of maps).

Your work in the previous two exercises would replicate nearly exactly to prove the general case, stated next. Note that while in most of our examples the contraction maps have been affine, that is not needed. All that is required is that the transformations be contractions.

THEOREM 3.10. *Let $T_i : \mathcal{H}(X) \to \mathcal{H}(X)$ be a contraction mapping with contraction factor $c_i \in [0, 1)$ for all $i = 1, 2, ..., n$. Then the collage map $\mathcal{T} = \bigcup_{i=1}^{n} T_i$ is also a contraction mapping on $\mathcal{H}(X)$, with contraction factor $c = \max\{c_i \text{ such that } i = 1, 2, ..., n\}$.*

DEFINITION 3.11. Let (X, d) be a compact metric space. A *(hyperbolic) iterated function system (IFS)* is given by $\{X; T_1, T_2, ..., T_K\}$, where each T_j is a contraction from X to itself with contractivity factor c_j. The *contractivity factor* of the IFS is given by $c = \max\{c_1, c_2, ..., c_K\}$.

3.2. Existence of Attractors for Iterated Function Systems

We are ready to prove the existence theorem for fractals generated by contraction maps on $\mathcal{H}(X)$, which includes those given by collages of contractions (i.e., iterated function systems). To prove the theorem we will rely on a lemma from real analysis that we state without proof. It says that if $\{B_i\}_{i=1}^{\infty}$ is a nested sequence of nonempty compact subsets of X, then the limit of the sequence is a nonempty compact subset of X also.

LEMMA 3.12. *Let $B_1, B_2, ...$ be a sequence of elements of $\mathcal{H}(X)$ such that $B_i \subseteq B_{i-1}$ for all $i = 1, 2, ...$ Then $\lim_{i\to\infty} B_i$ is an element of $\mathcal{H}(X)$.*

Since $B_1 \supseteq B_2 \supseteq B_3...$ it is clear that the sets in the sequence are getting smaller; the key fact to note is that this lemma says that they cannot disappear entirely (although they certainly could converge to a set containing only a point!).

EXERCISE 3.13. Let $X = [-10, 10]$. Make up an example of a nested sequence of elements of $\mathcal{H}(X)$, and find its limit.

And now for the big theorem. It is a special case of the *contraction mapping theorem*, which in a broad range of cases proves that a contraction map has a unique fixed point to which the orbit of every point is attracted.

THEOREM 3.14 (Existence theorem for fractals). *Let X be a compact subset of $\mathbb{R}^n$ or $\mathbb{C}$ and let $\mathcal{H}(X)$ be its space of fractals endowed with the Hausdorff metric. Let $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ be the collage of contractions $T_1, T_2, ..., T_K$ on X. Then*

1. *There exists a unique $A \in \mathcal{H}(X)$ for which $\mathcal{T}(A) = A$, and*
2. *For any $B \in \mathcal{H}(X)$,*

$$\lim_{n\to\infty} \mathcal{T}^n(B) = A \tag{3.1}$$

Before we launch into the proof, notice a few things. First, that part (1) promises us that a fixed point A exists, and that it is unique. The notation A is used because it is the "attractor" in the sense of (2): that the orbit of every $B \in \mathcal{H}(X)$ converges to it.

PROOF. Consider the sequence of elements of $\mathcal{H}(X)$ given by $\{\mathcal{T}^n(X)\}_{n=1}^{\infty}$. It is a nested sequence since $\mathcal{T}(X) \subseteq X$, and hence $\mathcal{T}^2(X) \subseteq \mathcal{T}(X)$, implying that $\mathcal{T}^3(X) \subseteq \mathcal{T}^2(X)$, and so on. By Lemma 3.12 we know that $\lim_{n\to\infty} \mathcal{T}^n(X)$ is an element of $\mathcal{H}(X)$ which we will call A. We must show that A is a fixed point of T, that it is unique, and that all orbits are attracted to it.

Since $\mathcal{T}$ is a contraction mapping, it is a continuous function and so

$$\mathcal{T}(A) = \mathcal{T}(\lim_{n\to\infty} \mathcal{T}^n(X)) = \lim_{n\to\infty} \mathcal{T}(\mathcal{T}^n(X)) = \lim_{n\to\infty} \mathcal{T}^{n+1}(X) = A,$$

and thus A is a fixed point of $\mathcal{T}$. The proof that A is unique is the next exercise. This will prove A is the only fixed point of $\mathcal{T}$ and thus completes the proof of part (1).

EXERCISE 3.15. Consider a set $C \in \mathcal{H}(X)$ for which $\mathcal{T}(C) = C$ and let $c \in [0, 1)$ represent a contraction factor for $\mathcal{T}$, which exists by Theorem _____. Then $d_H(\mathcal{T}(A), \mathcal{T}(C)) \leq$ _____. But A and C are both _____ by $\mathcal{T}$ and so that inequality becomes _____ $\leq c\, d_H(A, C)$. This is only possible if $d_H(A, C) =$ ___ . Since d_H is a _____ on $\mathcal{H}(X)$, we know that this is only possible if _____, proving that A is the unique fixed point.

The proof of equation (3.1) is the next exercise.

EXERCISE 3.16. Let $B \in \mathcal{H}(X)$ and let c be a contraction factor of $\mathcal{T}$. Prove equation (3.1) by proving that $\lim_{n\to\infty} d_H(\mathcal{T}^n(B), A) = 0$. As part of the proof, use induction to establish that $d_H(\mathcal{T}^n(B), A) \leq c^n\, d_H(B, A)$.

□

To summarize our findings, we now know the following.

(1) If $\{X; T_1, T_2, ..., T_K\}$ is an iterated function system with contraction factor $c = \max\{c_1, c_2, ..., c_K\}$, then it has a unique fixed point $A \in \mathcal{H}(X)$.
(2) That fixed point looks like $\mathcal{T}^n(X)$ for a large enough value of n, and so
(3) In chapter 1, the images for C_5, K_5, and S_5, which were the 5th iterations of X, look pretty much like the fractal does.

3.3. Two computer algorithms for IFS fractals

3.3.1. Deterministic algorithm. The theory is pretty clear: start with any initial compact set B and plot $\mathcal{T}^n(B)$ for as large of an n as your computer can handle. Here is some "pseudocode" for how you might actually do this; details will depend on the software you are using.

- Input the functions $T_1, T_2, ..., T_K$ into the computer.
- Input some initial set B_0, which could be as simple as a single point.
- Decide how many iterations you want to do and call it something like *numits*.
- Compute $\mathcal{T}^{numits}(B_0)$.
- Display $\mathcal{T}^{numits}(B_0)$.

For this course we will carry out the algorithm in Mathematica, and you will get some sample code for that soon.

3.3.2. Probabilistic Algorithm. The probabilistic algorithm for making fractals on the computer is quicker and more efficient than the deterministic one. It capitalizes on the fact that any set converges to A by applying the individual maps at random instead of using the entire collage map. We claim (but do not prove)

that the orbit of any point under a random application of the T_j's will trace out the entire fractal A. This may be believable since the collage map applied to any one-element set converges to A. Let us present the algorithm and then revisit this issue and the issue of relative efficiency.

Suppose we have the iterated function system $\{X; T_1, T_2, ..., T_K\}$. The algorithm begins with any initial point $x_0 \in X$ and computes its orbit a little differently than in Chapter 1, since $\mathcal{T}$ is not a map on X and we need to account for using all the T_js. For each $n \in 1, 2, 3, ...$ we select an index $j_n \in \{1, 2, ..., K\}$ randomly and apply that transformation to compute $x_n = T_{j_n}(x_{n-1})$. We compute and plot the entire orbit of x_0 up to $numits2$ iterations, where $numits2$ will be an extremely large number.

In advance of writing code there are two decisions to be made. It is necessary to decide on probabilities $p_1, p_2, ..., p_K$ to use to select the transformations; of course we need $\sum_{j=1}^{K} p_j = 1$. Higher probabilities mean more visits to the region that transformation covers, so the probabilities do affect the final images. That can be used as an advantage, for instance by giving smaller regions lower probabilities.

The other decision is what initial point x_0 to use. If the initial point is chosen from inside of A (for instance if it is the fixed point of a T_j), then the entire orbit of x_0 will be in A as you will prove in exercise 3.29. However it is not necessary to choose x_0 carefully since after a few iterations it must be close to A anyway, as you will prove in exercise 3.30.

Here is some pseudocode for the probabilistic algorithm.

- Input the T_j's, p_j's, and $numits2$.
- Generate a sequence $\{j_n\}_{n=1}^{numits2}$ of elements of $\{1, 2, ..., K\}$ according to the probabilities using your software's random number generator.
- Input x_0 and compute the set $\{x_0, x_1, x_2, ..., x_{numits2}\}$ according to $x_n = T_{j_n}(x_{n-1})$.
- Plot the subset $\mathcal{O} = \{x_{nmin}, x_{nmin+1}, x_{nmin+2}, ..., x_{numits2}\}$, where $nmin$ ensures the orbit is close enough to A. We may use $nmin = 10$ or 20.

One thing that is clear is that all of the points in $\mathcal{O}$ are close to points in A, as long as $nmin$ is chosen to be moderately large. You can decide on the size of $nmin$ by looking at the contraction factor of $\mathcal{T}$ and the maximum distance M in X. If you want the points to be within ϵ of A, simply choose $nmin$ so that $c^{nmin}M \leq \epsilon$. If you take ϵ to be pixel size and compute the corresponding $nmin$, then you can be sure that the points in $\mathcal{O}$ are, up to computer-visible resolution, points in A.

The harder thing to justify rigorously is that $\mathcal{O}$ fills up A in a representative fashion. In fact it will not if any of the p_is are set to be zero. Otherwise $\mathcal{O}$ should visit each T_js region about $p_j \cdot numits2$ times unless a highly improbable sequence of indices was selected by the random number generator (which is possible, of course). Ordinarily, then, we'd expect it to fill in each T_js part of the image of A pretty well. The experimenter can adjust the p_js if they feel the image is not representative.

3.3.3. A comparison of the two algorithms. The probabilistic algorithm is more efficient than the deterministic algorithm for getting a good picture of the attractor. It is beyond the scope of this course to prove that, but let's try and get a sense of why it is true. In everything that follows let $(X; T_1, ..., T_K)$ be an IFS with contraction factor $c = \max\{c_1, ..., c_K\}$.

EXERCISE 3.17. Suppose you are using the deterministic algorithm on the initial set $\{x_0\}$ for some $x_0 \in X$.

(1) What is the maximum number of elements in the set $\mathcal{T}^{numits}(\{x_0\})$?
(2) Let's make a really rough approximation by calling an "operation" the application of a single T_j to a single element of X. How many operations will the computer have to perform in order to compute $\mathcal{T}^{numits}(\{x_0\})$?

So you can see that *numits* has to be kept fairly small to make sure that your computer doesn't crash and can do all those operations in a timely fashion. However, if *numits* is small, $\mathcal{T}^{numits}(\{x_0\})$ may not be particularly close to A.

EXERCISE 3.18. Compute an upper bound on $d_H(\mathcal{T}^{numits}(\{x_0\}), A)$. (It is convenient to let M represent the maximum Euclidean distance between any two points in X.) Under what circumstances might this distance not be particularly small?

Now let's compare those results with what happens in the probabilistic algorithm.

EXERCISE 3.19. Now suppose that you are beginning with an arbitrary $x_0 \in X$ and using the probabilistic algorithm to compute the orbit out to $numits2$.

(1) Using the same rough definition of an operation as before, how many operations does it take to compute the orbit of x_0 up to $numits2$?
(2) If you cut off the first $nmin - 1$ elements of the orbit and consider the set $\mathcal{O} = \{x_{nmin}, x_{nmin+1}, x_{nmin+2}, ..., x_{numits2}\}$, give an upper bound on $\min\{\epsilon \text{ such that } \mathcal{O} \subset A_\epsilon\}$.
(3) Suppose $nmin$ is greater than the number of iterations $numits$ used in the deterministic algorithm, and suppose $numits2$ is about the same size as the number of points in $\mathcal{T}^{numits}(\{x_0\})$. Do you believe $\mathcal{T}^{numits}(\{x_0\})$ ought to be a better approximation of A than $\mathcal{O}$ is, or is $\mathcal{O}$ the better approximation of A? Explain.

3.4. The Collage Theorem

Suppose that you have an image L that you would like to store in as little space as possible. Or maybe you just want to be able to replicate L using an iterated function system for fun. The collage theorem, proved by Barnsley in 1985 and appearing on page 94 of [**Bar12**], gives you a way to measure how close the attractor of an iterated function system will be to your target image L.

THEOREM 3.20 (The Collage Theorem). *Let X be a compact subset of $\mathbb{R}, \mathbb{R}^d$, or $\mathbb{C}$ and let $L \in \mathcal{H}(X)$ be given. If $(X; T_1, ..., T_K)$ is an IFS with contraction factor c and attractor A, then*

$$d_H(L, A) \leq \frac{d_H(L, \mathcal{T}(L))}{1 - c}.$$

How is this theorem used? It certainly does not tell you how to get an IFS for which $d_H(L, \mathcal{T}(L))$ is small. You basically have to experiment with that, and there is literature out there containing strategies. You have full control over how many maps you want to use, their contraction factors, etc. But no matter what IFS you come up with from that process, the closeness of A to your target image L is bounded by how close L was to $\mathcal{T}(L)$.

EXERCISE 3.21. Will an IFS with a large contraction factor or a small contraction factor do a better job of approximating L? Does the number of T_is you use matter? Explain.

3.5. Exercises

EXERCISE 3.22. Suppose $\mathcal{H}(X)$ is a space of fractals with $A, B \in \mathcal{H}(X)$ and $\delta \geq 0$. Prove that

$$(A \cup B)_\delta = A_\delta \cup B_\delta \tag{3.2}$$

EXERCISE 3.23. Suppose A and B are elements of some space of fractals $\mathcal{H}(X)$ and let $\epsilon > 0$. Prove or give a counterexample to:

$$(A \cap B)_\epsilon = A_\epsilon \cap B_\epsilon$$

EXERCISE 3.24. Suppose $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ is a collage of the maps $T_i : X \to X$, $i = 1, 2, ..., K$. Prove that if $A \subset B$ for $A, B \in \mathcal{H}(X)$, then $\mathcal{T}(A) \subset \mathcal{T}(B)$.

EXERCISE 3.25. Give an example of a transformation $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ for which there are $A, B \in \mathcal{H}(X)$ with $A \subset B$, yet $\mathcal{T}(A) \not\subset \mathcal{T}(B)$.

EXERCISE 3.26. Make an IFS with three transformations in $\mathbb{R}^2$ that we haven't seen in class, and put it into the deterministic algorithm in mathematica. Please turn in your affine maps and your image of the attractor.

EXERCISE 3.27. Consider the spiral for Exercise 1.36. Here's a fun fact: the number of arms does not depend on the central transformation that has a large contraction factor. Instead it depends on the number of highly contractive transformations you put around the outside. Adapt the IFS for Exercise 1.36 to have three arms, and put your answer into the probabilistic mathematica code and see what the attractor looks like. Adjust until you are happy with your image.

EXERCISE 3.28. In this example you are going to play with probabilities when making the Sierpinski triangle. Compare and contrast the number of iterations necessary to produce a "good" view of the triangle when you use the following probabilities:

(1) $p_1 = .33, p_2 = .33, p_3 = .34$.

(2) $p_1 = .2, p_2 = .46, p_3 = .34$.
(3) $p_1 = .1, p_2 = .56, p_3 = .34$.

EXERCISE 3.29. Let $(X; T_1, ..., T_K)$ be an IFS and let $\{j_n\}_{n=1}^{numits2}$ be a sequence of indices from $1, 2, ..., K$ as would be selected for the probabilistic algorithm. Suppose a is an element of the attractor A of the IFS. Prove that the orbit of a in the algorithm is contained in A.

EXERCISE 3.30. Let $(X; T_1, ..., T_K)$ be an IFS with attractor A and let $\{j_n\}_{n=1}^{\infty}$ be a sequence of indices from $1, 2, ..., K$ as would be selected for the probabilistic algorithm (except infinite, theoretically). Suppose $x_0 \in X$ and compute its orbit $\{x_n\}_{n=1}^{\infty}$ as in that algorithm. Prove that $l(x_n, A) = \min\{d(x_n, a)$ such that $a \in A\}$ tends to 0 as $n \to \infty$.