

Math 101
Solution and Commentary
for the First Group Project

Stefan Karpinski

December 10, 1999

Rather than addressing each point of the project in order, I will first give a statement and proof of the Contraction Mapping Theorem using the principle of nested closed intervals. Ensuing commentary should address all points of the the assignment.

Theorem. *Every contraction of a complete metric space has a unique fixed point.*

This is a relatively easy generalization of the theorem you were asked to prove, which was simply the special case where our metric space is taken to be a closed subset of \mathbb{R} with the standard Euclidean metric inherited from \mathbb{R} . However neither of these two assumptions is essential to the proof, so it is simpler and cleaner to work in a more general context. The principle of nested closed intervals presents the only dilemma: what are nested closed intervals in a general metric space? Or more to the point, what does it mean for a general metric space to be complete? To answer this we simply generalize the concept of a closed interval to the concept of a closed ball

$$B(x; r) = \{y \in X : d(x, y) \leq r\}.$$

Given a sequence of points $\{x_n\}$ and a sequence of radii (positive real numbers) $\{r_n\}$ indexed over \mathbb{N} , we say that they are nested if for $m > n$ we have $B(x_m; r_m) \subseteq B(x_n; r_n)$. From homework we know that this is equivalent to requiring that $\{x_n\}$ and $\{r_n\}$ satisfy

$$d(x_n, x_m) \leq r_n - r_m \quad \text{whenever } m > n,$$

which in particular tells us that $r_m \leq r_n$ for $m > n$. With this generalization of nested closed intervals, we may define completeness as follows.

Definition. *A metric space (X, d) is said to be complete if for every pair of sequences $\{x_n\}, \{r_n\}$ satisfying nestedness we have*

$$\bigcap_{n \in \mathbb{N}} B(x_n; r_n) \neq \emptyset.$$

With this established, let's proceed with the proof.

Proof. Suppose $f: X \rightarrow X$ is our contraction mapping. We want to use the completeness of our metric space in some way to produce a fixed point of the contraction. To this end, pick any $x_0 \in X$ and define recursively $x_n = f(x_{n-1})$. This gives us our sequence of points $\{x_n\}$ but before we can use the principle of nested closed intervals (or closed balls as we have here) we need to define our sequence of radii $\{r_n\}$ so that we satisfy nestedness. This is the hardest part of the proof—the rest is just applying definitions cleverly. Notice that not every sequence $\{x_n\}$ has a sequence of radii $\{r_n\}$ such that nestedness is satisfied. To produce such a sequence of radii we need to use the way that $\{x_n\}$ is defined and the hypothesis that f is a contraction. Satisfying nestedness amounts to establishing an upper bound on $d(x_n, x_m)$ for $m > n$ which depends only on n and which decreases (sufficiently fast) as n increases. Let us do precisely that. Assume $m > n$.

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n+1}) + \cdots + d(x_{m-1}, x_m) \\ &\leq C^n d(x_0, x_1) + \cdots + C^m d(x_0, x_1) \\ &= C^n(1 + C + \cdots + C^{m-n})d(x_0, x_1) \\ &\leq C^n(1 + C + C^2 + \cdots)d(x_0, x_1) \\ &= C^n(1 - C)^{-1}d(x_0, x_1). \end{aligned}$$

Call this last value k_n . This bound certainly doesn't leap to mind immediately—it requires some sneaky tricks and several strong facts and results. Notable among them are (in order of usage) the triangle inequality, contractionness of f , non-negativity of C (tells us the infinite sequence is bigger than its finite initial segment), and finally the fact that $|C| < 1$ lets us evaluate the infinite sequence $1 + C + C^2 + \cdots = (1 - C)^{-1}$. Finally we define

$$\begin{aligned} r_n &= k_n + k_{n+1} + \cdots \\ &= (C^n + C^{n+1} + \cdots)(1 - C)^{-1}d(x_0, x_1) \\ &= C^n(1 - C)^{-2}d(x_0, x_1) \end{aligned}$$

so that we now have for $m > n$,

$$\begin{aligned} r_n - r_m &= k_n + k_{n+1} + \cdots + k_{m-1} \\ &\geq k_n \\ &\geq d(x_n, x_m) \end{aligned}$$

which means precisely that we have $\{r_n\}$ satisfying nestedness. The upshot of all of this trickery and sneaky approximation is that we now have a pair of sequences $\{x_n\}$ and $\{r_n\}$ satisfying nestedness, so we may use the completeness of our metric space to deduce that there is some point $x \in \bigcap_n B(x_n; r_n)$. (Moreover, since $\lim_{n \rightarrow \infty} r_n = 0$, this point x is unique but we don't need this fact—and

the uniqueness of x in the above intersection is not to be confused with the uniqueness of the fixed point.) For arbitrary n we have

$$\begin{aligned} d(x, f(x)) &\leq d(x, f(x_n)) + d(f(x), f(x_n)) \\ &\leq d(x, x_{n+1}) + Cd(x, x_n) \\ &\leq r_{n+1} + Cr_n, \end{aligned}$$

but we can make this last term arbitrarily small by picking n large enough, so we must have $d(x, f(x)) = 0$ which means $x = f(x)$. Score! Moreover, this fixed point is unique: suppose $x = f(x)$ and $x' = f(x')$; then $d(x, x') = d(f(x), f(x')) \leq Cd(x, x')$, where $0 \leq C < 1$, which yields a contradiction unless $d(x, x') = 0$, in which case $x = x'$. \square

With this general form of the CMT we can immediately address parts 4 and 7 of the assignment: CMT does not apply to \mathbb{Q} because \mathbb{Q} is not complete and connectedness is completely irrelevant to the CMT. It happens that \mathbb{Q} is not connected, but that's just a coincidence—CMT applies to all sorts of disconnected sets such as $[0, 1] \cup [2, 3]$, or even something more complicated like $\bigcup_{n \in \mathbb{Z}} [2n, 2n + 1]$ and anything else you can cook up, as long as it's a complete metric space. As to a function on $A = \{x \in \mathbb{Q} : x \geq 1\}$ which is a contraction but has no fixed point in A , there are of course infinitely many such functions but one example would be the function

$$f(x) = 1 + \frac{1}{1+x}.$$

A much more challenging question is to produce a contraction on all of \mathbb{Q} which has no fixed point. (I'm not even certain that one exists but I'm pretty sure that one must. Think about it and let me know if you come up with anything.)

We can also address the last question of part 3. In the more restricted version of the theorem that you were asked to prove, we need A to be closed because if (X, d) is a complete metric space and A is a subspace of X , then A is complete if and only if A is closed. The simplest example of a contraction on $(0, 1)$ without a fixed point is just $f(x) = \frac{1}{2}x$.

Now for part 5. While we know that given any particular distinct $x, y \in A$ if $d(f(x), f(y)) < d(x, y)$, we can find some C such that $0 \leq C < 1$ and $d(f(x), f(y)) \leq Cd(x, y)$, we cannot guarantee that this will be the same C for all x, y and moreover, that we can put an upper bound on such C which is strictly less than 1. However, if our set A is compact then we can actually find some maximal C strictly less than 1. This is since continuous functions attain their maxima on compact sets. (What function am I actually applying this fact to? Hint it's not just f and it involves the metric (twice) and division.) What that means is that when, for example, $A = [0, 1] \cup [2, 3]$, all "pseudo-contractions" (functions $f: A \rightarrow A$ that satisfy the "pseudo-contraction" condition $d(f(x), f(y)) < d(x, y)$ —this is not a real term, I just made it up) actually do have to have a fixed point since we can find some contraction constant that

works for all pairs x, y because A is compact. What I'm saying is that in compact spaces, pseudo-contractions are real contractions. But if we choose A to be something non-compact such as all of \mathbb{R} , then we can find functions which satisfy $d(f(x), f(y)) < d(x, y)$ but have no fixed points. So for non-compact spaces, there are pseudo-contractions which are not real contractions. To find such a function on \mathbb{R} all we need to do is find a function whose graph never touches the line $x = y$ (so it has no fixed points) and such that the absolute value of the slope between any two points on the graph of the function is always less than 1 (this is equivalent to it satisfying the pseudo-contraction condition—why?). One such function is the half-hyperbola $f(x) = -(1+x^2)^{1/2}$. You should verify that this function satisfies the pseudo-contraction condition $d(f(x), f(y)) < d(x, y)$.

And finally, part 6. If f is a contraction allowing contraction constant C then it is immediate that $g = f^2$ is also a contraction allowing the still smaller contraction constant C^2 . On the other hand, $g = f^2$ may be a contraction without f being a contraction. For example, define $f: \mathbb{R} \rightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} -x, & x < 0, \\ 0, & x \geq 0. \end{cases}$$

Then $g = f^2 = 0$, which is clearly a contraction while f is blatantly not. Of course, since $f^2 = 0$, we have $f^n = 0$ for all higher n as well, so in particular, f^n is always a contraction. If f is a contraction with fixed point x then f^n will always be a contraction and will moreover have the same fixed point. (More generally, if f is any function then the fixed points of f^m will always include the fixed points of f^n when $m > n$.)

That seems to cover all of the material of the second group project.