The Completion of a Metric Space

Brent Nelson

Let (E, d) be a metric space, which we will reference throughout. The purpose of these notes is to guide you through the construction of the "completion" of (E, d). That is, we will construct a new metric space, $(\overline{E}, \overline{d})$, which is complete and contains our original space E in some way (to be made precise later).

1 Initial Construction

This construction will rely heavily on sequences of elements from the metric space (E, d). So, for notational convenience we will let $\underline{a}$ denote the sequence whose nth entry is $a_n \in E$ (that is, $\underline{a} = (a_n)_{n\in\mathbb{N}}$), $\underline{b}$ will denote the sequence whose nth entry is $b_n \in E$, etc.

Consider the following set:

$$\mathcal{C}[E] := \{\underline{a} \colon (a_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence in } (E, d)\}.$$

You should think of $\mathcal{C}[E]$ as a new space where each point $\underline{a} \in \mathcal{C}[E]$ is a Cauchy sequence $\underline{a} = (a_n)_{n\in\mathbb{N}}$ from (E, d). Note that for each $x \in E$, we can define $\underline{x} \in \mathcal{C}[E]$ by letting $x_n = x$ for all $n \in \mathbb{N}$. This gives us a way to think of $\mathcal{C}[E]$ as containing our original space E, which we will make more precise in Section 3. For $\underline{a}, \underline{b} \in \mathcal{C}[E]$ we define

$$D(\underline{a}, \underline{b}) := \lim_{n\to\infty} d(a_n, b_n),$$

(you will show that this limit exists in Exercise 1.1).

This might seem like an unnecessarily complicated thing to do, but there is a good intuitive reason for why to consider $\mathcal{C}[E]$ (see Section 5). The hope with this initial construction is that $(\mathcal{C}[E], D)$ is a complete metric space, but, as will be seen in part (v) of Exercise 1.2, D fails to even be a metric. Hence, we will have to make some adjustments to this initial construction, which we shall undertake in the following sections.

Exercises

1.1 For any $\underline{a}, \underline{b} \in \mathcal{C}[E]$, show that the sequence $\{d(a_n, b_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence of real numbers and hence converges. [Hint: use the reverse triangle inequality]

1.2 Show that D is almost a metric in the following sense. For any $\underline{a}, \underline{b}, \underline{c} \in \mathcal{C}[E]$ we have

- (i) $D(\underline{a}, \underline{b}) \in [0, +\infty)$;
- (ii) $D(\underline{a}, \underline{a}) = 0$;
- (iii) $D(\underline{a}, \underline{b}) = D(\underline{b}, \underline{a})$;
- (iv) $D(\underline{a}, \underline{b}) \leq D(\underline{a}, \underline{c}) + D(\underline{c}, \underline{b})$; and
- (v) $D(\underline{a}, \underline{b}) = 0$ does **not** imply $\underline{a} = \underline{b}$.

2 Equivalence Relations

As we saw with Exercise 1.2, D is a nearly a metric except for the fact that distinct sequences $\underline{a}$ and $\underline{b}$ could have zero distance between them according to D. We can fix this issue by replacing $\mathcal{C}[E]$ with a set where such sequences $\underline{a}$ and $\underline{b}$ are thought of as the "same" point. Towards this end we consider the following definition.

Definition 2.1. Let X be a set. An **equivalence relation on** X is a binary relation $\sim$ satisfying for all $x, y, z \in X$:

(i) $x \sim x$ (**Reflexivity**);

(ii) $x \sim y$ if and only if $y \sim x$ (**Symmetry**); and

(iii) $x \sim y$ and $y \sim z$ implies $x \sim z$ (**Transitivity**).

For $x, y \in X$ satisfying $x \sim y$, we say x **is equivalent to** y.

Usually, an equivalence relation is just a formal way of saying that two elements share a common quality. Here is a real world example:

Example 2.2. If $\mathcal{H}$ is the set of all humans who ever lived, then we can put a binary relation on $\mathcal{H}$ by defining human $x \sim$ human y to mean human x was born in the same year as human y. It easy to check that this is in fact an equivalence relation: human x is always born in the same year as themselves, so reflexivity is satisfied; if human x is born in the same year as human y then human y is also born in the same year as human x, so symmetry is satisfied; and if human x was born in the same year as human y and human y was born in the same year as human z, then clearly human x was born in the same year as human z, so transitivity is satisfied. Under this equivalence relation, Terence Tao is equivalent to Curtis James Jackson III, since both were born in 1975, and we would write "Terence Tao $\sim$ Curtis James Jackson III." Similarly, we could write "Brent Nelson $\sim$ Daniel Radcliffe" or say "Brent Nelson is equivalent to Taylor Swift."

Sets can often have many equivalence relations on them. With $\mathcal{H}$ as in the previous example, we could define another equivalence relation by saying human x is equivalent to human y if and only if they were born on the same day of the year. However, not all binary relations are equivalence relations. For the same set $\mathcal{H}$, if $x \sim y$ is defined to mean human x is a parent of human y, then this relation will fail to be reflexive, symmetric, and transitive. If $x \sim y$ is defined to mean human x is a sibling of human y, then this will be symmetric and transitive, but not reflexive. Here are some more mathematical examples:

Example 2.3. Fix a number $n \in \mathbb{N}$. Then for $x, y \in \mathbb{Z}$ write $x \sim y$ if and only if $x - y$ is divisible by n. Since $0 = x - x$ is always divisible by n, this relation is reflexive. Since $x - y$ is divisible by n if and only if $y - x = -(x - y)$ is divisible by n, this relation is symmetric. Finally, if $x - y$ is divisible by n and $y - z$ is divisible by n, then $x - z = (x - y) + (y - z)$ is also divisible by n, and so the relation is transitive. Hence $\sim$ is an equivalence relation.

Example 2.4. For $x, y \in \mathbb{R}$, write $x \sim y$ if and only if $x - y \in \mathbb{Z}$. This is easily checked to be an equivalence relation.

When we have an equivalence relation on a set, we can group all of the equivalent elements together in so-called "equivalence classes."

Definition 2.5. Given an equivalence relation $\sim$ on a set X, for $x \in X$ the **equivalence class of** x is defined as the set

$$[x] := \{y \in X : y \sim x\}.$$

Any element $y \in [x]$ is called a **representative of the equivalence class** $[x]$.

Because of the symmetry and transitivity properties of an equivalence relation, for $x, y \in X$ we can either have $[x] = [y]$ (in the case that $x \sim y$) or $[x] \cap [y] = \emptyset$. Also, by reflexivity we always have $x \in [x]$. Consequently, the equivalence classes of a set X give a way to express X as a disjoint union: let $X/\sim$ denote the collection of equivalence classes

$$X/\sim = \{[x] : x \in X\},$$

then

$$X = \bigsqcup_{[x] \in X/\sim} [x].$$

Let's revisit our previous examples:

Example 2.6. When $\mathcal{H}$ is the set of all humans who ever lived and human $x \sim$ human y means human x was born in the same year as human y, then the equivalence class [Terence Tao] is the set of all humans born in 1975.

Example 2.7. Fix $n \in \mathbb{N}$, and for $x, y \in \mathbb{Z}$ write $x \sim y$ if and only if $x - y$ is divisible by n. Then

$$\mathbb{Z} = [0] \sqcup [1] \sqcup \cdots \sqcup [n-1],$$

where for $j = 0, 1, \ldots, n-1$, $[j]$ is the set of numbers whose remainder is j when divided by n.

Example 2.8. For $x, y \in \mathbb{R}$, write $x \sim y$ if and only if $x - y \in \mathbb{Z}$. Then if x has a decimal expansion $x = a_0.a_1a_2\cdots$, then $[x] = \{n + 0.a_1a_2\cdots : n \in \mathbb{Z}\}$; that is, the set of numbers with the same decimal part as x.

Recall our definition of $\mathcal{C}[E]$ from Section 1. Given $\underline{a}, \underline{b} \in \mathcal{C}[E]$ we define the following binary relation: $\underline{a} \sim \underline{b}$ if and only if $\lim_{n\to\infty} d(a_n, b_n) = 0$. That is, $\underline{a} \sim \underline{b}$ iff $\forall \epsilon > 0$ $\exists N \in \mathbb{N}$ such that $\forall n \geq N$ we have $d(a_n, b_n) < \epsilon$. In Exercise 2.1, you will show that this defines an equivalence relation. Consequently, we can consider equivalence classes under this equivalence relation:

$$[\underline{a}] = \{\underline{b} \in \mathcal{C}[E] \colon \underline{b} \sim \underline{a}\}.$$

We define $\overline{E}$ to be the set of equivalence classes in $\mathcal{C}[E]$ with respect to this equivalence relation: $\overline{E} := \mathcal{C}[E]/\sim$. There is a lot to unpack here: the points in $\overline{E}$ are equivalence classes $[\underline{a}]$; each equivalence class $[\underline{a}]$ is itself a subset of elements $\underline{b}$ from $\mathcal{C}[E]$; and each $\underline{b}$ is a Cauchy sequence of elements from E. The way that you should conceptualize $\overline{E}$ is to think of it as the set $\mathcal{C}[E]$ where we do not distinguish between $\underline{a}$ and $\underline{b}$ if they satisfy $\underline{a} \sim \underline{b}$. In this way, an element $[\underline{a}] \in \overline{E}$ can be identified with the representative $\underline{a}$, since any other representative $\underline{b} \in [\underline{a}]$ is (informally) "the same as" $\underline{a}$.

For $[\underline{a}], [\underline{b}] \in \overline{E}$ we wish to define

$$\overline{d}([\underline{a}], [\underline{b}]) := D(\underline{a}, \underline{b}),$$

but *a priori* it is not clear that this is *well-defined*; that is, since the left-hand side is defined in terms of equivalence classes, the right-hand side should not depend on the choice of representative of the equivalence class $[\underline{a}]$ or $[\underline{b}]$. Thus, one must check that if $\underline{c} \in [\underline{a}]$ and $\underline{d} \in [\underline{b}]$ then

$$D(\underline{a}, \underline{b}) = D(\underline{c}, \underline{d}).$$

This will follow from Exercise 2.2. Using this and Exercise 1.2, it follows that $(\overline{E}, \overline{d})$ is in fact a metric space. We will show in the later sections that this is actually a complete metric space and that it "contains" the original metric space (E, d) in a meaningful way.

Exercises

2.1 Show that the binary relation $\sim$ on $\mathcal{C}[E]$ defined above is an equivalence relation.

2.2 Show that $\underline{a} \sim \underline{b}$ if and only if $D(\underline{a}, \underline{b}) = 0$. Use this to verify that if $\underline{a} \sim \underline{c}$ and $\underline{b} \sim \underline{d}$, then $D(\underline{a}, \underline{b}) = D(\underline{c}, \underline{d})$.

3 Isometries

Definition 3.1. Given two metric spaces (E_1, d_1) and (E_2, d_2) and a map $f \colon E_1 \to E_2$, we say that f is an **isometry** if

$$d_2(f(x), f(y)) = d_1(x, y) \qquad \forall x, y \in E_1.$$

Note that any isometry is automatically one-to-one: if $f(x) = f(y)$ then $d_1(x, y) = d_2(f(x), f(y)) = 0$ which implies $x = y$. An isometry $f \colon E_1 \to E_2$ allows one to realize a copy of the metric space (E_1, d_1) inside of the metric space (E_2, d_2) in the form of $(f(E_1), d_2)$. This copy contains all of the original information about (E_1, d_1), in particular the distances between various points.

Example 3.2. Consider the map $f\colon \mathbb{R} \to \mathbb{R}^2$ defined by $f(x) = (x, 0)$. This map sends the real numbers to the horizontal axis in $\mathbb{R}^2$. It is easily checked that this is an isometry when $\mathbb{R}$ is given the usual metric and $\mathbb{R}^2$ is given either the 2-dimensional Euclidean metric, the d_1 metric, or the d_∞ metric (see Homework 3). Here we can think of the $f(\mathbb{R})$ as a copy of $\mathbb{R}$ living inside of $\mathbb{R}^2$.

Example 3.3. Consider the map $f\colon \mathbb{R} \to \mathbb{R}^2$ defined by $f(x) = (\frac{x}{\sqrt{2}}, \frac{x}{\sqrt{2}})$. This maps the real numbers to the line $y = x$ in $\mathbb{R}^2$. This is an isometry when $\mathbb{R}$ is given the usual metric and $\mathbb{R}^2$ is given the 2-dimensional Euclidean metric, but not when $\mathbb{R}^2$ is given the d_1 or d_∞ metrics.

Recall the definition of $(\overline{E}, \overline{d})$ from the previous section. Consider the map $\iota\colon E \to \overline{E}$ defined by

$$\iota(x) := [(x)_{n\in\mathbb{N}}];$$

that is, $\iota(x)$ is the equivalence class of the (constant) Cauchy sequence $(x)_{n\in\mathbb{N}}$ where each coordinate in the sequence is x. In Exercise 3.1 you will show that this is an isometry, so that $\iota(E)$ is a copy of E living inside of $\overline{E}$. It is in this sense that $\overline{E}$ contains the original metric space E. Furthermore, in Exercise 3.2 it will be shown that $\iota(E)$ is dense in $\overline{E}$, and so the closure of $\iota(E)$ is $\overline{E}$. This is precisely why the notation '$\overline{E}$' is the same as our notation for the closure of a set.

Exercises

3.1 Show that $\iota\colon E \to \overline{E}$ is an isometry.

3.2 Show $\iota(E)$ is dense in $\overline{E}$.

4 Completeness

Finally, we endeavor to show that $(\overline{E}, \overline{d})$ is complete: that every Cauchy sequence converges. This is the content of Exercise 4.1, but first let us unpack what it means to have a Cauchy sequence in $(\overline{E}, \overline{d})$. Each element in $\overline{E}$ is of the form $[\underline{a}]$. So a sequence of elements will look like $\left([\underline{a}^{(k)}]\right)_{k\in\mathbb{N}}$, where $[\underline{a}^{(k)}]$ for each $k \in \mathbb{N}$ is the equivalence class of $\underline{a}^{(k)} \in \mathcal{C}[E]$ (which is incidentally a Cauchy sequence of elements from E). Suppose $\left([\underline{a}^{(k)}]\right)_{k\in\mathbb{N}}$ is a Cauchy sequence in $(\overline{E}, \overline{d})$; that is, for any $\epsilon > 0$ there exists $K \in \mathbb{N}$ such that for all $k, l \geq K$ we have

$$\overline{d}([\underline{a}^{(k)}], [\underline{a}^{(l)}]) < \epsilon.$$

To show $(\overline{E}, \overline{d})$ is complete, we must show the sequence $([\underline{a}^{(k)}])_{k\in\mathbb{N}}$ converges to some $[\underline{a}] \in \overline{E}$ with respect to $\overline{d}$. For each $k, l \geq K$, $\underline{a}^{(k)}, \underline{a}^{(l)} \in \mathcal{C}[E]$ are representatives of the equivalence classes $[\underline{a}^{(k)}]$ and $[\underline{a}^{(l)}]$, respectively. So, by definition of $\overline{d}$, for all $k, l \geq K$ we have

$$D(\underline{a}^{(k)}, \underline{a}^{(l)}) = \overline{d}([\underline{a}^{(k)}], [\underline{a}^{(l)}]) < \epsilon.$$

Now, $\underline{a}^{(k)}, \underline{a}^{(l)} \in \mathcal{C}[E]$ are both Cauchy sequences of elements in E:

$$\underline{a}^{(k)} = (a_n^{(k)})_{n\in\mathbb{N}}$$
$$\underline{a}^{(l)} = (a_n^{(l)})_{n\in\mathbb{N}},$$

where $a_n^{(k)}, a_n^{(l)} \in E$ for each $n \in \mathbb{N}$. So, by definition of D, for all $k, l \geq K$ we have

$$\lim_{n\in\mathbb{N}} d(a_n^{(k)}, a_n^{(l)}) = D(\underline{a}^{(k)}, \underline{a}^{(l)}) < \epsilon.$$

In summary, if $([\underline{a}^{(k)}])_{k\in\mathbb{N}}$ is a Cauchy sequence, then for any $\epsilon > 0$ there exists $K \in \mathbb{N}$ such that for all $k, l \geq K$ we have

$$\lim_{n\to\infty} d(a_n^{(k)}, a_n^{(l)}) < \epsilon.$$

Similarly, what it means for $([\underline{a}^{(k)}])_{k\in\mathbb{N}}$ to converge to some $[\underline{a}] \in \overline{E}$ (where $\underline{a} = (a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in E) is that for any $\epsilon > 0$ there exists $K \in \mathbb{N}$ such that for all $k \geq K$ we have

$$\lim_{n\to\infty} d(a_n^{(k)}, a_n) < \epsilon.$$

Thus, your task in Exercise 4.1 is to determine what $\underline{a}$ should be (namely, what a_n should be for each $n \in \mathbb{N}$) and show that $([\underline{a}^{(k)}])_{k\in\mathbb{N}}$ converges to it in the above sense.

A completely reasonable question to ask at this point is "Does it really need to be this complicated?" Well, in the case that E was already complete, the answer is actually "No!" As you will see in Exercise 4.2, when E is already complete, $\iota(E)$, the copy of it living inside $\overline{E}$, is actually all of $\overline{E}$. That is, $\overline{E}$ does not add anything that wasn't already in E, which is to be expected. However, when E is not complete $\iota(E)$ is really a strict subset $\overline{E}$ and so this whole rigmarole did actually produce something new and necessary. It is the goal of the next, and final, section to give a visual intuition for what we are really doing in this construction of $(\overline{E}, \overline{d})$.

Exercises

4.1 Show $(\overline{E}, \overline{d})$ is complete.

4.2 Show $\iota(E) = \overline{E}$ if and only if E is complete.

5 The Intuition

We will imagine our initial space E as a 2-dimensional amorphous blob, lying flat on the ground. If we assume E is not initially complete, then this blob will have lots of tiny pinpricks/holes in it which represent missing points (points that will eventually appear in $\overline{E}$). Indeed, the moral of our construction (specifically Exercise 4.2) is that a metric space really only fails to be complete if it is "missing" points that its Cauchy sequences *want* to converge to. To produce the completion of E, we have to find a way to close up these pinpricks, but our only resource is the space E itself.

So, make a copy of our blob and place it directly on top of the previous copy. Here we should imagine they have a certain thickness to them, like a sheet of paper, so that our second copy is literally resting on top of the first. Mathematically, we have constructed E^2, which we can think of as length two sequences of elements in E.

Make another copy of E and set it on top of the stack: this gives E^3. Iterating this procedure we'll obtain an infinitely tall stack of copies of our metric space, which corresponds to E^∞: the set of sequences $(x_n)_{n\in\mathbb{N}}$ with $x_n \in E$, for all $n \in \mathbb{N}$. Each sequence corresponds to choosing a point from each copy in the stack, roughly forming an vertical path up to the "top" of the stack. The Cauchy sequences $\mathcal{C}[E]$ are those paths that eventually start to straighten out (i.e. eventually do not jump around too much horizontally).

To visualize the equivalence relation on $\mathcal{C}[E]$, imagine we climb under the stack and are looking up through the bottom (initial) copy. Then two Cauchy sequences (vertical paths) are equivalent if they get closer and closer together as they near the top (which is infinitely high up). In particular, if we pick out a point on our bottom copy of E, say $x \in E$, and stare straight up through it, we are seeing the constant sequence $(x)_{n\in\mathbb{N}}$.

Now, if we attempt to look up through one of the pinpricks we started with, it will be very difficult to see that it actually remains a pinprick all the way to the top. In fact, due to the infinite height of the stack it will appear to close up (imagine looking down a very deep well, or up a very tall skylight). But this is precisely saying that, up to our equivalence relation, we have managed to plug all the pinpricks in E and therefore made it complete. Hence this infinite procedure was necessary, since any finite stack of copies would have still had open pinpricks.

Interestingly, this construction gives us another way to produce the real numbers $\mathbb{R}$ from the much tamer set of rationals $\mathbb{Q}$. Indeed, the completion of $(\mathbb{Q}, d)$, where $d(x, y) = |x - y|$, can be identified with $(\mathbb{R}, d)$. Note that for $x, y \in \mathbb{Q}$ we always have $d(x, y) \in \mathbb{Q}$, and also that the notion of a limit can be defined with only rational numbers (replace $\epsilon > 0$ in the definition with $\frac{1}{m}$ for any $m \in \mathbb{N} \subset Q$). Hence the completion of the rationals gives an independent construction of the real numbers. We could very well have begun our course with this construction, rather than assuming the existence of the reals as we did, but then we would have had to verify that $\mathbb{R}$ satisfies Properties I-VII from the beginning of our course. This is a non-trivial task, but does offer a more rigorous foundation for the real numbers.