You could have invented topology

Luis A. Florit

Suppose you have a real function $f : \mathbb{R} \to \mathbb{R}$, and you want to determine if f *is continuous at* $x_0 \in \mathbb{R}$, that is, if:

$$\textit{As } x \textit{ approaches to } x_0, \textit{ the value } f(x) \textit{ approaches to } f(x_0). \tag{1}$$

We say that f *is continuous* if it is continuous at every point $x_0 \in \mathbb{R}$.

We have two problems with this *intuitive* definition, and both relate to formalization. The most obvious one is to formalize the idea that "a approaches to b". Here there is implicitly some kind of *distance* involved: we mean that the distance between a and b gets small while a moves and b stays still. Yet, what does "gets small" mean? The second and less obvious problem is what we are trying to determine when we say "as".

Let's call $d(a, b)$ the distance between a and b, even without explicitly saying what it is (and even without knowing what kind of objects a and b are!). So what (1) says is that

$$d(f(x), f(x_0)) \textit{ gets small as } d(x, x_0) \textit{ gets small.} \tag{2}$$

Of course, whatever $d(a, b)$ is, it should be a nonnegative real number, after all, it represents a distance. One practical way of formalizing that "$d(x, x_0)$ gets small" is to say that $d(x, x_0)$ is smaller than a positive number ϵ, that is, $0 \leq d(x, x_0) < \epsilon$, and then take ϵ arbitrarily small. Thus, in (2) we are saying that

$$d(f(x), f(x_0)) < \epsilon \textit{ whenever } d(x, x_0) \textit{ is small enough.} \tag{3}$$

This "small enough" amount depends on all the objects involved: on ϵ itself, the function f and the point x_0. In other words, given $\epsilon > 0$, there should be another (likely small) positive number $\delta = \delta(\epsilon, x_0, f)$ such that

$$d(f(x), f(x_0)) < \epsilon \quad \textit{if} \quad d(x, x_0) < \delta. \tag{4}$$

Since this property should be verified <u>for all</u> $\epsilon > 0$ arbitrary small, we incorporate this into (4) by saying that

$$\textit{for all } \epsilon > 0 \textit{ there is } \delta > 0 \textit{ such that } d(f(x), f(x_0)) < \epsilon \quad \textit{if} \quad d(x, x_0) < \delta. \tag{5}$$

Congratulations, you have (re)invented the formal notions of *limit* and *continuous function.* Notice that the fact that δ depends on ϵ, x_0 and f, although implicitly, is already logically included in the statement (5).

Now, the natural distance between two real numbers a and b is given by the absolute value of their difference, i.e., $d(a,b) = |a-b|$. We can now substitute this into (5) to conclude that:

$$\textit{for all } \epsilon > 0 \textit{ there is } \delta > 0 \textit{ such that } |f(x) - f(x_0)| < \epsilon \;\; \textit{if} \;\; |x - x_0| < \delta. \tag{6}$$

However, the important point here is that this explicit expression of the distance is irrelevant, as long as it trully represents a *distance*. Indeed, up to now we did everything without saying what it was, and by (5) we now know that

Each time we have a notion of distance, we can talk about continuous functions.

More importantly, distances can be defined in arbitrary sets, not only in $\mathbb{R}$. However, you would certainly agree that any distance function d must satisfy at least the following three very natural geometric properties to have the right to be called a distance:

$$d(a,b) = 0 \Leftrightarrow a = b, \;\; d(a,b) = d(b,a), \;\; \textit{and} \;\; d(a,b) + d(b,c) \geq d(a,c), \;\; \forall a,b,c. \tag{7}$$

Notice that, by taking $c = a$ in the third property, called *triangle inequality*, we get $d \geq 0$, a fact that we already used above. A set X endowed with function $d : X \times X \to \mathbb{R}$ satisfying (7) is called a *metric space*, and is denoted by (X, d), or simply by X. You can now easily extend the notion of continuity for maps of any metric space, $f : (X,d) \to (X,d)$, or even between two different metric spaces, $f : (X,d) \to (X', d')$.

Let's go a little further by writing (5) in a more set theoretical language. Given a point z in a metric space (X,d) and $\epsilon > 0$, we define the *metric ball of radius ϵ centered at z* as

$$B_\epsilon(z) := \{x \in X : d(x,z) < \epsilon\}. \tag{8}$$

Then, simply by rewriting (5) using (8) for $f : (X,d) \to (X,d)$ we obtain

$$\textit{for all } \epsilon > 0 \textit{ there is } \delta > 0 \textit{ such that } f(B_\delta(x_0)) \subset B_\epsilon(f(x_0)), \tag{9}$$

or, equivalently,

$$\textit{for all } \epsilon > 0 \textit{ there is } \delta > 0 \textit{ such that } B_\delta(x_0) \subset f^{-1}(B_\epsilon(f(x_0))). \tag{10}$$

Now, observe that any metric ball B has the following key property: for any $x \in B$, there is $\tilde{\epsilon} > 0$ (depending on x) such that $B_{\tilde{\epsilon}}(x) \subset B$. Indeed, if $B = B_\epsilon(z)$ and we take $\tilde{\epsilon} := \epsilon - d(z,x)$, then $\tilde{\epsilon} > 0$ and by the triangle inequality in (7) we have $B_{\tilde{\epsilon}}(x) \subset B_\epsilon(z)$. Any subset B of X with this property is called *open*. We can use this concept to rephrase (10) by saying that, for every open subset B containing $f(x_0)$, there should exist an open subset U containing x_0 such that $U \subset f^{-1}(B)$. Equivalently,

$$\textit{for every open subset } B \textit{ containing } f(x_0), \; f^{-1}(B) \textit{ is an open subset containing } x_0. \tag{11}$$

Since for a continuous function this should hold for every x_0 in X, we can simply eliminate x_0 and say that

$$f^{-1}(B) \text{ is open if } B \subset X \text{ is open.} \tag{12}$$

Here, we have to be a bit careful since f does not need surjective and hence $f^{-1}(B)$ could be empty. But obviously from the definition the empty set $\emptyset$ is open, so no problem here.

It is clear now from (11) that all we need in order to talk about continuous functions is the collection τ of all open subsets of X. In the case of a metric space, with the three properties of the distance function in (7) you can check that:

i) Both $X \in \tau$ and $\emptyset \in \tau$;

ii) Union of elements in τ belongs to τ;

iii) Intersections of finitely many elements in τ belongs to τ.

A collection τ of subsets of an arbitrary set Z that satisfies these three properties is called a *topology* for Z, and the pair (Z, τ) is called a *topological space*. By the previous observation, every metric space has a natural topology induced by its distance, for which (12) becomes

$$\forall B \in \tau,\ f^{-1}(B) \in \tau. \tag{13}$$

Without effort you can check for yourself that, for the naturally induced topology on a metric space, the two notions (5) and (13) are equivalent. Yet, for (13) no distance function is required, only the collection τ satisfying $(i)+(ii)+(iii)$. This is a much more general setting, and topological spaces are the most basic structure where we can talk about continuous functions.

Compare (1) and (13), and think deeply about the jump. ;o)

IMPA – Estrada Dona Castorina, 110
22460-320 — Rio de Janeiro — Brazil
luis@impa.br