1. History and first examples

During the last 200 years mathematics has enjoyed a return of geometric techniques. For one who has not studied the subject at a higher level, “geometry” may be a word reminiscent of theorems about triangles and circles dating back to ancient Greece, possibly the topic of some middle and high school classes. Of course, there have been numerous developments in the subject dating from ancient times to the 1800s, yet what makes the latter period so interesting, both historically and mathematically, is the systematic and rigorous development of topology. Many describe topology as the study of “rubber geometry”, or more informatively, of objects which allow deformation [4, 3, 6]. This intuitive point of view has become a favourite subject for many popular accounts since the 1950s. The surge of flabby geometry is a reflection of contemporary developments, which was not the original intent of the theory. We will attempt to recover the logical sequence of constructions which lead to the modern definition of a topological space by Kazimierz Kuratowski dating back to 1922 [1, 2].

Our story starts in the early days of calculus during the 17th century. Sir Isaac Newton and Gottfried Leibniz were the two main scholars working on this new, exciting field. They developed ideas which have nowadays become an inseparable part of mathematics – derivatives, integrals, differential equations, and many others. Despite the brilliant foresight of these great men, from a modern point of view their work saw little if no rigor as evidenced by the scarcity of definitions which were often not general enough. For example, they talked about converging sequences, series, and taking limits without really having concrete explanations of what these meant. Alongside with advances in physics, calculus experienced a boom. The foundational problems had to wait until the early 19th century to be resolved. We will follow closely on the development of continuity as a case study of how numerous other ideas evolved. Augustin Louis Cauchy was the first to formulate our modern “epsilon-delta” definition of continuity [6, 7].

Before delving into the details, let us step back for a second and try to understand what we are trying to formalize. We consider a map \( f : \mathbb{R} \to \mathbb{R} \), or in other words a function from the real numbers to the real numbers. In fanciful mathematical language, this is the good old graph we plot on the \( xy \)-plane. In this context, we can vaguely formulate continuity as the ability to draw the graph with a pencil without having to lift it from the plane. This concept is illustrated in Figure 1 below.

![continuity and discontinuity](image)

This is an extremely naïve definition as it leaves numerous questions unresolved. There are functions which do not jump at isolated points as the one above, but infinitely often in a dense manner. These cannot be graphed in such an idyllic fashion, yet they allow some continuous points. Even more, the mere mention of a “pencil” makes the flaw extreme. Cauchy’s “epsilon-delta” definition fixes these issues. According to it, the function \( f \) is continuous at some point \( x \in \mathbb{R} \) if

\[
\forall \varepsilon > 0, \exists \delta > 0, \forall y, |y - x| < \delta \Rightarrow |f(y) - f(x)| < \varepsilon.
\]
In plain English, the above expression says: for all \( \varepsilon > 0 \), there exists some \( \delta > 0 \), such that if \( y \) is no further than \( \delta \) from \( x \), then \( f(y) \) is no further than \( \varepsilon \) from \( f(x) \). It is now clear the notorious nickname of this definition refers to two of the variables used.

2. Metric spaces

Let us take a closer look at the above definition. All of \( x, y, f(x), f(y) \) are real values, so to measure distances we used the absolute value function in the expressions \( |y-x| \) and \( |f(y) - f(x)| \). Everything we said so far has been adapted to the specific case of real numbers. The real line is possibly one of the simplest geometric objects, so despite the fact that we have a working definition of continuity, it is not a very useful one. Most calculus classes cover both single and multiple variables. Hence, the next object of interest is a function \( f: \mathbb{R}^m \rightarrow \mathbb{R}^n \), which carries \( m \)-dimensional space to \( n \)-dimensional space. We can adjust our previous definition using the Euclidean distances \( \|y-x\|_m \) and \( \|f(y) - f(x)\|_n \). It is important to note that the domain and codomain (colloquially referred to as input and output – in this case \( \mathbb{R}^m, \mathbb{R}^n \)) are different sets with possibly different notions of distance. This is to our advantage since we can talk about continuity of a broader range of functions. On the other hand, we are still using subtraction in \( \mathbb{R}^m \) and \( \mathbb{R}^n \). There are many geometric objects of interest which do not allow algebraic operations on them.

In order to dispose ourselves from the subtraction, we will start by considering an arbitrary set \( X \). We reduced the definition of continuity in and out of \( X \) to providing a means of measuring distance between the points of \( X \) (in set theoretic terms, its elements). A metric on \( X \) is a function \( d: X \times X \rightarrow \mathbb{R} \), or in plainer language, \( d \) takes as parameters two points in the set \( X \) and returns a real number. The metric \( d \) emulates our understanding of distance – given two points, it “measures” the distance between them. There are several conditions we need to impose on \( d \) in order to ensure it is well-behaved.

Let \( X \) be a set and \( d: X \times X \rightarrow \mathbb{R} \) a map. The tuple \((X, d)\) is called a metric space if the metric satisfies:

(a) **non-negativity** (for all \( x, y \in X \), \( d(x, y) \geq 0 \)),
(b) **non-degeneracy** (for all \( x, y \in X \), \( d(x, y) = 0 \) if and only if \( x = y \)),
(c) **symmetry** (for all \( x, y \in X \), \( d(x, y) = d(y, x) \)), and
(d) **the triangular inequality** (for all \( x, y, z \in X \), \( d(x, z) \leq d(x, y) + d(y, z) \)).

These are all properties that we normally assume when considering distance measurements. Part (a) imposes that no two points can be at a negative distance from each other, (b) that two points are distance zero from each other if and only if they are the same point, and (c) that the distance from \( x \) to \( y \) is the same as the distance from \( y \) to \( x \). The most involved condition is the triangular inequality (d), which requires going from \( x \) to \( z \) directly is closer than going from \( x \) to \( y \) and then to \( z \).

At this point we can state Cauchy’s definition in full generality. Let \((X, d_X)\) and \((Y, d_Y)\) be metric spaces. A map \( f \) is called continuous at \( x \in X \), if

\[
\forall \varepsilon > 0, \exists \delta > 0, \forall y \in X, d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \varepsilon.
\]

If \( f \) is continuous at all points \( x \in X \), then we call it continuous. The definition of a metric space and the associated form of continuity are due to Maurice Fréchet who introduced them in 1906 [1, 2]. Despite the additional level of abstraction, it is clear that the above is nothing more than a restatement of Cauchy’s original definition with the exception of a few cosmetic changes. Furthermore, the conditions (a)-(d) imposed on a metric space are a very natural choice. It is important to note that metric spaces encompass numerous objects of interest to geometry, analysis, and other areas of mathematics. They also cover much more than we initially expected. For example, under certain conditions the set of continuous functions between two metric spaces can itself be turned into a metric space, which is one of the fundamental examples studied in functional analysis.

3. Topological spaces

In our first attempt to formalize continuity, we discussed the possibility of drawing the graph of a function without lifting the pencil. Our formal investigation has yielded Fréchet’s formulation of a metric space as a means of encoding geometric data. In order to reach a functioning definition, we were forced to introduce the distance metric, yet originally we had no intent of incorporating such an object. We described topology as the study of rubber geometry. As one deforms an object, different parts could be contracted or expanded, modifying the distance between points. These remarks suggest there should be a more intrinsic formulation of continuity.
and a means of incorporating geometric data in general, which does not refer to a distance function. The object
which we are alluding to is called a topological space.

There are several issues to address before we would be able to provide a concrete explanation. Let us consider
a metric space \((X, d)\), which we will often refer to simply as \(X\). Any subset \(Y \subset X\) induces a metric space
\((Y,d|_Y)\). Formally, the metric \(d\) on \(X\) restricts to \(d|_Y: Y \times Y \to \mathbb{R}\) which acts as a metric on \(Y\). From
this point of view, any subset of points could possibly act as a subspace of \(X\). In practice however, not all
subsets yield spaces which are significantly interesting. There are two special types of subsets which we will be
focusing on – respectively called open and closed. If \(x \in X\) is a point and \(r\) a positive real number, then the
ball of radius \(r\) or \(r\)-ball around \(x\), denoted \(B_X(x,r)\), is the set of all points \(y \in X\) no further than \(r\) from
\(x\), that is \(B_X(x,r) = \{y \in X \mid d(x,y) < r\}\). Such balls are often referred to as neighbourhoods of \(x\). Note
that the subscript at \(B\) indicates the metric space we are working in. We call a subset \(U \subset X\) open, if for all
\(x \in U\), there exists some radius \(r > 0\) such that \(B_X(x,r) \subset U\) (we often refer to such sets simply as opens).
This means every point of the open \(U\) contains a neighbourhood fully contained within \(U\). Provided with this
definition of metric openness we call a set \(C \subset X\) closed if its complement \(X \setminus C\) is open. To give an example,
consider the real numbers \(\mathbb{R}\) with the standard metric given by \(d(x,y) = |x - y|\). For every \(x \in \mathbb{R}\) and \(r > 0\), the
\(r\)-ball around \(x\) is nothing more than the interval \((x-r, x+r)\). For every two real numbers \(a < b\) the interval
\((a,b) \subset X\) is open, since for every \(c \in (a,b)\) the ball \(B_\mathbb{R}(c, \min\{b-c, c-a\})\) is contained in \((a,b)\). That is
in fact the largest ball around \(c\) contained in \((a,b)\). Similarly, the interval \([a,b]\) is closed since its complement
\((-\infty, a) \cup (b, \infty)\) is open. These properties are often attached to intervals without justification of the underlying
topological concepts. Now that we have a better understanding the real line, let us consider a higher dimensional
example. Consider the space \(\mathbb{R}^n\) with the metric induced by the Euclidean distance \(d(x,y) = \|x-y\|_n\). Opens in
\(\mathbb{R}^n\) can be colloquially described as regions which do not contain their boundary, much like opens intervals
do not contain their endpoints. Analogously, regions that contain their boundary are closed. Once again, we are
describing open and closed sets without any specific reference to the metric we are using; yet, the rigorous path
formulate these concepts in the case of metric spaces employs a distance function. These observations suggest
there must be a different solution to the problem we are looking at.

Let \(X\) be an arbitrary set. We will consider a collection of subsets \(\mathcal{T}\) of \(X\) and declare these open. This is a
very drastic approach to defining openness. Disposing ourselves of the metric structure, we are left with nothing else
but to “define away” the problem. If the opens in \(\mathcal{T}\) are to behave analogously to the case of metric spaces,
we need to impose several conditions:

(a) both the empty set and the entire space are open \((\emptyset, X \in \mathcal{T})\);
(b) finite intersections of opens are open (if \(U_1, \ldots, U_n \in \mathcal{T}\), then \(\bigcap_{i=1}^n U_n \in \mathcal{T}\));
(c) arbitrary unions of opens are open (if \(U_\alpha \in \mathcal{T}\) for all \(\alpha \in A\), then \(\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}\)).

A pair \((X, \mathcal{T})\) which satisfies the above three properties is called a topological space. We often omit the collection
of opens \(\mathcal{T}\) and refer to a topological space simply by its set of points, in our case \(X\). Before continuing, let us
take a closer look at the definition above. Part (a) is very simple, but (b) and (c) could be very confusing. In
particular, it is not clear why we restricted intersections to be finite, yet allowed unions to be arbitrary (possibly
infinite). The simple answer is this mimics the familiar case of metric spaces and the definition of openness there.
The union of an arbitrary collection of opens in a metric space is again open, and any finite intersection of
opens is also open. The finiteness condition is imposed, since it is possible to construct an infinite collection of
open subsets in a metric space whose intersection is not open. If for any integer \(n \geq 1\) we let \(U_n\) denote the
interval \((-1/n, 1/n)\) in \(\mathbb{R}\), then \(\bigcap_{n=1}^\infty U_n = \{0\}\). The single point set \(\{0\}\) is definitely not open, which explains
the rationale behind the definition of a topological space. It is important to understand that topological spaces are
not removed from metric spaces. Every metric space \((X,d)\) induces a topology \(\mathcal{T}\) on \(X\) (given by the open sets
in the metric sense), which turns \((X, \mathcal{T})\) into a topological space. The converse however is not true. Not every
topological space \((X, \mathcal{T})\) is induced by some metric \(d\) on \(X\). If such a metric exists, we call the topological space
metrizable. Studying conditions which ensure metrizability is a major topic of interest in point set topology.

Introducing rigorously topological spaces was a significant detour from our discussion of continuity. It is
only logical to proceed by explaining when a map between two topological spaces is continuous. Once again our
definition will be inspired by analogies with the metric case. In particular, continuity of maps between metric
spaces was defined explicitly using the distance function. Consider a continuous map \(f: X \to Y\) between two
metric spaces. We can reformulate the continuity of \(f\) by saying that

\[ \forall x \in X, \forall \varepsilon > 0, \exists \delta > 0, \ f(B_X(x,\delta)) \subset B_Y(f(x),\varepsilon). \]
This follows directly from the definition of a ball and some simple logic. Since all balls are open, it seems this reformulation is on the right track. It is a slightly more difficult exercise to show that \( f \) is continuous if and only if for every open \( U \subset Y \), its preimage \( f^{-1}(U) \subset X \) is open too (the preimage \( f^{-1}(U) \) is the set of all points \( x \in X \) such that \( f(x) \) lies in \( U \)). This follows from the definition of metric openness and the above statement. Note that this way of expressing continuity has finally disposed us of the distance function and only refers to open sets. It is also considerably simpler to comprehend in comparison with the earlier postulate that incorporated three quantifiers. We can now state the final version of continuity we have been aiming for. Consider two topological spaces \((X, T_X), (Y, T_Y)\), and a map \( f : X \to Y \). We say that \( f \) is continuous if for every open \( U \subset Y \) (that is \( U \in T_Y \)), its preimage \( f^{-1}(U) \) is open in \( X \) (meaning \( f^{-1}(U) \in T_X \)).

It is important to make several remarks regarding the approach we took. Our discussion started by treating bare sets, and we proceeded to define two types of objects (metric and topological spaces) which add structure on top of a set. Many constructions in mathematics follow this trend. Topological spaces are only the basis for numerous further constructions that add more and more structure to them. To mention a few examples – manifolds, sheaves, and schemes are all objects which build on the basis of topological spaces. The modular approach to definitions often reduces the simplicity of exposition and allows scholars to easily build on earlier work. It is now apparent what the advantages of working with topological spaces are. Firstly, concepts such as continuity have simpler reformulations, which are often easier to deal with. Secondly, topological spaces could be used to model a very wide array of situations, hence their broad applications to all branches of mathematics.

References