MATH 147: Analysis 1
Supplementary Topics

Instructor: Brian E. Forrest & Laurent.W. Marcoux

Fall 2020

Table of Contents

Math 147 - September 2020
Cardinality
L.W. Marcoux

1. Introduction.

1.1. Suppose that someone (whom we shall call "Einstein") brought you a grocery bag filled with bananas and apples, and asked you whether you have more bananas than apples, or more apples than bananas, or an equal amount of each. Very likely, you'd just laugh and start eating the bananas, because they are a complete fruit, and most doctors agree that they are good for you.

But what if that someone *paid* you (handsomely) to answer the question? Suddenly your interest in the problem would grow.

At this stage, there is a good chance that you would count the number of apples, count the number of bananas, and you would then compare the two numbers you got.

Let's think about this for a second.

1.2. Did you see what you did there? Nobody asked you how many apples and bananas there were. Einstein only asked you whether you have more bananas than apples, or more apples than bananas, or an equal amount of each. You went beyond the call of duty and answered a question no one asked. Of course, it did allow you to answer Einstein's question. But in this case, your solution was feasible in part because an average grocery bag will not hold more than ... oh, maybe ... 300 bananas or 300 apples.

1.3. Now suppose that you go to a soccer stadium, and on the pitch you have an enormous pile of shoes. Let's agree that the pile is more than enormous, it is in fact *ginormous.* What Einstein wants you to tell him is whether the pile includes more left shoes, or right shoes. But there are so many of each, that each time you try counting them (still an excellent initial strategy, by the way), you lose count.

Einstein is offering you a large sum of money to answer the question: *are there more left shoes than right shoes in that pile, more right shoes than left shoes, or an equal number of each?*

Now is the time to ask yourself: what would Madame Curie do?

1.4. Here's an idea. You head up to the stadium announcer's booth, and over the loudspeaker you issue instructions to your friends on the pitch as follows:

> *each time you find a left shoe and a right shoe, pair them up and put them aside.*

What is going to happen next, pray tell? Hopefully you will agree that one of the following three scenarios will play out.

- There were more left shoes in the pile when we started than right shoes, and so after all of the pairs have been put aside, you will have a non-empty pile of left shoes left over.

- There were more right shoes in the pile when we started than left shoes, and so after all of the pairs have been put aside, you will have a non-empty pile of right shoes left over.
- There were the same number of each, and after all of the pairs of shoes have been put aside, nothing is left over.

Let's think about this for another second.

1.5. Did you see what you did there this time? In the first instance, you found a *surjection* (i.e. an *onto* map) from the set of left shoes onto the set of right shoes. The existence of this surjection means that you had at least as many left shoes as right shoes. In the second instance, you found an *injection* (i.e. a *one-to-one* map) from the set of left shoes to the set of right shoes – i.e. you had at least as many right shoes as left shoes. Finally, in the third instance, you found a *bijection* (i.e. a *one-to-one and onto* map) between both sets of shoes.

Without ever having counted how many left shoes nor right shoes there were in the pile, you can nevertheless compare the size of these two sets of shoes using this pairing method, and you can decide that you must have had "*the same amount*" of left shoes and right shoes if the pairing was perfect - i.e. if it was described by a bijection. Clearly we could not have found a bijection had there not been an equal number of left and right shoes.

Marie Curie would be proud of you.

1.6. Let's return to our bananas and apples.

- When you counted the bananas and found that you had m of them, you actually established a bijection between the set of bananas and the set $\{1, 2, \ldots, m\} \subseteq \mathbb{N}$.
- When you counted the apples and found that you had n of them, you actually established a bijection between the set of bananas and the set $\{1, 2, \ldots, n\} \subseteq \mathbb{N}$.

What happens when $m > n$? In that case, you can find a surjection $\alpha : \{1, 2, \ldots, m\}$ onto $\{1, 2, \ldots, n\}$ by setting $\alpha(k) = k$, $1 \leq k \leq n$ and $\alpha(k) = 1$ if $n < k \leq m$.

What happens when $m < n$? In that case, you can find an injection $\beta : \{1, 2, \ldots, m\}$ onto $\{1, 2, \ldots, n\}$ by setting $\beta(k) = k$, $1 \leq k \leq m$.

What happens when $m = n$? In that case, you can find a bijection $\gamma : \{1, 2, \ldots, m\}$ onto $\{1, 2, \ldots, n\}$ by setting $\gamma(k) = k$, $1 \leq k \leq m = n$.

In other words, your "*counting*" really just boils down to finding a surjection, an injection or a bijection from the set $\{1, 2, \ldots, m\}$ to the set $\{1, 2, \ldots, n\}$. But by composing these maps with the bijections above, this just boils down to finding a surjection, an injection or a bijection from the set of bananas to the set of apples. In the previous case with the pile of shoes, we eliminated the middle-man; that is, we didn't bother counting, or equivalently, we didn't bother to find the bijections between the set of left (resp. right) shoes and a set $\{1, 2, \ldots, m\}$ (resp. $\{1, 2, \ldots, n\}$) for an appropriate choice of m and n.

1.7. The counting system works great with finite sets, such as sets of shoes and sets of bananas. It doesn't work so great with infinite sets, however, such as infinite sets of monkeys, each with a typewriter, attempting to reproduce Shakespeare. As we shall see below - *you can't count the set of real numbers.* That is, there is no bijection $\alpha : \mathbb{N} \to \mathbb{R}$.

So what if Einstein asks you whether there are more integers than even integers, or are there the same amount of each?

Well, if this should happen, definitely the second thing you should ask yourself is: given that both sets are infinite (in the sense defined below), what does Einstein mean by "*more*", or "*the same amount*"? Having said this, clearly the first thing you should be asking yourself is: "why doesn't Einstein just leave me alone?"

1.8. The idea is to use the pairing method that we previously used for shoes to *define* what we mean when we say that a set A is *bigger than*, *smaller than*, or *the same size* as a set B.

More specifically, in complete analogy to the situation above with the shoes and the bananas/apples, we say the following.

1.9. Definition. *Let A and B be non-empty sets.*

(a) *We say that the* ***cardinality of*** *A is* ***less than or equal to the cardinality of*** *B, and we write $|A| \le |B|$, if there exists an injection $\alpha : A \to B$.*
(b) *We say that the* ***cardinality of*** *A is* ***greater than or equal to the cardinality of*** *B, and we write $|A| \ge |B|$, if there exists a surjection $\beta : A \to B$.*
(c) *We say that the* ***cardinality of*** *A is* ***equal to the cardinality of*** *B, and we write $|A| = |B|$, if there exists a bijection $\gamma : A \to B$.*

1.10. One extremely important thing to note about this definition is the presence of the expression "if there exists" in each item (a), (b), and (c) above. The definition of $|A| = |B|$ in (c), for example, does not require that every map $\delta : A \to B$ be a bijection. Finding a map $\delta : A \to B$ that is a surjection but not a bijection does not in general mean that we can't find a different map which is a bijection. (Actually it does if A and B are finite, but the whole point of this essay on cardinality is to learn to compare infinite sets.)

Let's have a look at this madness *in action.*

1.11. Example. Let $A = \mathbb{Z}$ denote the set of integers, and let

$$B := 2\mathbb{Z} = \{\dots, -4, -2, 0, 2, 4, \dots\}$$

denote the set of *even* integers.

The map

$$\begin{array}{rccc} \alpha : & A & \to & B \\ & a & \mapsto & \begin{cases} a & \text{if } a \text{ is even} \\ 0 & \text{if } a \text{ is odd} \end{cases} \end{array}$$

defines a surjection from A onto B. This map is not a bijection, since $\alpha(1) = \alpha(3) = 0$.

The map

$$\begin{array}{rccc} \beta : & A & \to & B \\ & a & \mapsto & \begin{cases} a+1 & \text{if } a \text{ is odd} \\ 0 & \text{if } a \text{ is even} \end{cases} \end{array}$$

also defines a surjection from A onto B. This map is not a bijection, since $\beta(2) = \beta(4) = 0$.

The existence of α (or of β) tells us that the cardinality of A is greater than or equal to that of B; that is, $|\mathbb{Z}| \geq |2\mathbb{Z}|$.

Notice that the map

$$\begin{array}{rccc} \varrho : & B & \to & A \\ & b & \mapsto & b \end{array}$$

is an injection from $B = 2\mathbb{Z}$ into (but not onto) $A = \mathbb{Z}$. By our Definition above, $|2\mathbb{Z}| = |B| \leq |A| = |\mathbb{Z}|$. At least our notation is consistent, and suggests that A is (intuitively speaking) *bigger than – or at least no smaller than –* the set B.

This is where infinite sets stop behaving like finite sets. The map ϱ says that we can pair up each even number in $B = 2\mathbb{Z}$ with its copy in $A = \mathbb{Z}$, and of course that means that nothing in B was paired up with an odd number in A using ϱ, so there are a lot of "left-overs" in A that didn't get paired up with anything. But our definition of two sets having the same cardinality only required us to find *one* bijection.

Consider the map

$$\begin{array}{rccc} \tau : & A & \to & B \\ & a & \mapsto & -2a. \end{array}$$

This particular map τ *is* a bijection, and thus – by definition – A and B have the same cardinality. We should think of this as meaning that they have the same "size". Of course, the bijection τ between A and B need not be unique – in general it isn't. For example, we could have use

$$\begin{array}{rccc} \xi : & A & \to & B \\ & a & \mapsto & 2a + 163,334,224 \end{array}$$

just as well. In both cases, it is akin to matching one left shoe to one right shoe with no shoes leftoever.

1.12. Example. As a second example, let $A = \mathbb{N} = \{1, 2, 3, \ldots, \}$, and $B = \{2, 3, 4, \ldots, \}$. The map

$$\begin{array}{rccc} \tau : & A & \to & B \\ & a & \mapsto & a + 1. \end{array}$$

defines a bijection from A onto B. We say that A and B have the same cardinality!

Again, the map

$$\begin{array}{rccc} \beta : & B & \to & A \\ & a & \mapsto & a \end{array}$$

defines an injection which is not surjective. Thus $|B| \leq |A|$. But the fact that *there exists* a bijection τ as above gives us the stronger notion that $|B| = |A|$.

Those of you without a NetFlix or NHL subscription and hence nothing better to do may have seen this example described as the *Infinite Hotel.* A hotel with rooms numbered $\{1, 2, 3, \ldots, \}$ is full, and a new person comes looking for a room, so the manager just shifts the person in room a to room $a+1$ and creates a vacancy, as well as infinitely many unhappy guests.

1.13. Since cardinality and comparing the sizes of non-empty sets boils down to find injections, surjections and bijections, the following exercise becomes especially useful.

Exercise. Let X be a non-empty set, and denote by $\mathbb{P}(X)$ the power set of X. Define a relation $\sim$ on $\mathbb{P}(X)$ by setting $A \sim B$ if $|A| = |B|$.

Prove that $\sim$ is an **equivalence relation** on $\mathbb{P}(X)$. That is,

(a) Prove that for all $A \in \mathbb{P}(X)$, $A \sim A$.
(b) Prove that for all $A, B \in \mathbb{P}(X)$, $A \sim B$ implies that $B \sim A$.
(c) Prove that if A, B and $C \in \mathbb{P}(X)$, $A \sim B$ and $B \sim C$, then $A \sim C$.

Let's enjoy a bit of terminology, shall we?

1.14. Definition. *A set A is said to be:*

(i) ***finite*** *if either $A = \varnothing$, and so A has cardinality 0, written $|A| = 0$, or there exists $n \in \mathbb{N}$ and a bijection $f : \{1, 2, ..., n\} \to A$, in which case we write $|A| = n$;*
(ii) ***denumerable*** *if there exists a bijection $g : \mathbb{N} \to A$. We write $|A| = \aleph_0$;*
(iii) ***countable*** *if A is either finite or denumerable;*
(iv) ***uncountable*** *if A is not countable.*

1.15. Examples.

(a) It is obvious that $\mathbb{N}$ is denumerable (hence countable), since the identity map $g : \mathbb{N} \to \mathbb{N}$ is clearly a bijection.
(b) In Example 1.12, we showed that $B := \{2, 3, 4, \ldots\}$ is denumerable as well.
(c) Consider the set $\mathbb{Z}$ of integers:

$$\mathbb{Z} = \{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \ldots\}.$$

We claim that $\mathbb{Z}$ is denumerable, hence countable, as well. To do that, we shall establish a bijection $\beta : \mathbb{N} \to \mathbb{Z}$ as follows:

n	1	2	3	4	5	6	7	...
$\beta(n)$	0	-1	1	-2	2	-3	3	...

This picture is great for understanding what the function β does. But to *use* it in a proof, it's better to have an explicit description of it. For any $n \in \mathbb{N}$, we define

$$\beta(n) = \begin{cases} \dfrac{n-1}{2} & \text{if } n \text{ is odd} \\[2ex] -\dfrac{n}{2} & \text{if } n \text{ is even .} \end{cases}$$

We leave it as an exercise for you to show that β is indeed a bijection.

We have shown that $\aleph_0 := |\mathbb{N}| = |\{2, 3, 4, \ldots\}| = |\mathbb{Z}|$. Now let's try something a bit more interesting. Note that if X is any denumerable set, then there exists a bijection $\beta : \mathbb{N} \to X$. In this case, we tend to write x_n instead of $\beta(n)$ for each $n \geq 1$, and so this allows us to write

$$X = \{x_1, x_2, x_3, \ldots\}.$$

1.16. Proposition. *Suppose that $\{A_n\}_{n=1}^{\infty}$ is a (countable!!!) collection of denumerable sets. Then $A = \cup_{n\geq 1} A_n$ is denumerable.*
Proof. We shall prove this in the case where $A_n \cap A_m = \varnothing$ if $n \neq m$. The general case follows easily from this (**exercise**). Since A_n is denumerable, we can write $A_n = \{a_{n1}, a_{n2}, a_{n3}, ...\}$ for each $n \geq 1$. Construct a new sequence by following the arrows below. (We have only included the first 13 arrows; there are infinitely many!)

$$\begin{tikzcd}
a_{11} \arrow[r] & a_{12} \arrow[dl] & a_{13} \arrow[r] & a_{14} \arrow[dl] & a_{15} & \cdots \\
a_{21} \arrow[d] & a_{22} \arrow[ur] & a_{23} \arrow[dl] & a_{24} & a_{25} & \\
a_{31} \arrow[ur] & a_{32} \arrow[dl] & a_{33} \arrow[ur] & a_{34} & a_{35} & \\
a_{41} \arrow[d] & a_{42} \arrow[ur] & a_{43} & a_{44} & a_{45} & \\
a_{51} \arrow[ur] & a_{52} & a_{53} & a_{54} & a_{55} & \\
\vdots & & & & &
\end{tikzcd}$$

Then $A = \{a_{11}, a_{12}, a_{21}, a_{31}, a_{22}, a_{13}, a_{14}, a_{23}, a_{32}, \ldots\}$. Thus A is countable.

$\square$

1.17. Corollary. *The set $\mathbb{Q}$ of rational numbers is countable.*
Proof. Indeed, for each $n \geq 1$, let $A_n = \{0/n, 1/n, -1/n, 2/n, -2/n, 3/n, -3/n, ...\}$. Then $\mathbb{Q} = \cup_{n\geq 1} A_n$. By (b) above, $\mathbb{Q}$ is countable.

$\square$

1.18. Proposition 1.16 is typically considered to be the key to being able to prove the more general statement:

The countable union of countable sets is countable.

We shall leave that as an exercise for you to consider.

Recall that between any two distinct rational numbers there are infinitely many irrational numbers, and that between any two distinct irrational numbers one can find infinitely many rational numbers. One might be tempted to believe that the set $\mathbb{I}$ of irrational numbers and the set $\mathbb{Q}$ of rational numbers are equipotent. That would be a mistake.

Our goal now is to show that the interval $(0, 1) \subseteq \mathbb{R}$ is uncountable. This gives us our first glimpse into the notion that not all infinite sets were born equal. That is, in the sense of cardinality, there are different notions of "infinity".

1.19. Theorem. *The set* $(0,1)$ *of real numbers between* 0 *and* 1 *is uncountable.*
Proof.
Using Cantor's diagonal process and a proof by contradiction, we shall prove that $(0,1)$ is uncountable.

If $(0,1)$ were countable, then we could find a bijection $\alpha : \mathbb{N} \to (0,1)$. Let $x_n := \alpha(n)$, $n \geq 1$. This allows us to write $(0,1) = \{x_n\}_{n=1}^{\infty}$, where each x_n is expressed in decimal form as

$$
\begin{aligned}
x_1 &= 0.x_{11}\ x_{12}\ x_{13}\ x_{14}... \\
x_2 &= 0.x_{21}\ x_{22}\ x_{23}\ x_{24}... \\
x_3 &= 0.x_{31}\ x_{32}\ x_{33}\ x_{34}... \\
x_4 &= 0.x_{41}\ x_{42}\ x_{43}\ x_{44}... \\
&\vdots
\end{aligned}
$$

and of course, each x_{ij} is an integer between 0 and 9.

Let $y = 0.y_1\ y_2\ y_3\ y_4...$ where $y_n = 7$ if $x_{nn} \in \{0,1,2,3,4,5\}$ and $y_n = 3$ if $x_{nn} \in \{6,7,8,9\}$. Then $y \neq x_n$, since $y_n \neq x_{nn}$ for all $n \geq 1$.

In other words, $y \in (0,1)$, but $y \notin \{x_n\}_{n=1}^{\infty}$, a contradiction of the hypothesis that α was surjective!

Thus no such bijection α exists, and so $(0,1)$ is uncountable.

$\square$

We leave the following Corollary as an exercise!

1.20. Corollary. *The set* $\mathbb{R}$ *of real numbers is uncountable.*

1.21. Let us denote $|\mathbb{R}|$ by c (for *continuum*). Since $\mathbb{N} \subseteq \mathbb{R}$, $\aleph_0 \leq c$. In fact, since $\mathbb{R}$ is uncountable, $\aleph_0 < c$. That is, there are at least two infinite cardinals. In fact, there are many, many more.

1.22. Theorem. *For any set* X, $|X| < |\mathcal{P}(X)|$, *where* $\mathcal{P}(X) = \{Y : Y \subseteq X\}$ *is the power set of* X.
Proof. First note that the map

$$
\begin{array}{rccc}
f: & X & \to & \mathcal{P}(X) \\
 & x & \mapsto & \{x\}
\end{array}
$$

is injective, and so $|X| \leq |\mathcal{P}(X)|$.

Next, suppose that $g : X \to \mathcal{P}(X)$ is any surjective map. Given $x \in X$, either $x \in g(x)$, or $x \notin g(x)$. Let $T = \{x \in X : x \notin g(x)\}$. Since g is surjective, there exists $z \in X$ so that $g(z) = T$.

If $z \in T$, then $z \in g(z)$, so $z \notin T$, a contradiction.

If $z \notin T$, then $z \notin g(z)$, so $z \in T$, again a contradiction.

Thus g can not be surjective, and *a fortiori* it can not be bijective, so $|X| < |\mathcal{P}(X)|$.

$\square$

1.23. It follows that $\aleph_0 = |\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))| < \cdots$, and that there exist infinitely many infinite cardinal numbers. Where does $c = |\mathbb{R}|$ fit? We have seen that $\aleph_0 < c$, but does there exist an infinite cardinal λ such that $\aleph_0 < \lambda < c$? Writing $\aleph_1$ for the first cardinal bigger than $\aleph_0$, the question becomes: is $c = \aleph_1$?

The conjecture that there does *not* exist such a λ is known as the *Continuum Hypothesis* and is due to Georg Cantor. In 1938, Kurt Gödel proved that the Continuum Hypothesis does not contradict the usual axioms of set theory. In 1963, Paul Cohen prove that the negation of the Continuum Hypothesis is also consistent with the usual axioms of set theory.

Whereas the Axiom of Choice is freely used (but cited) by the majority of mathematicians, it is not standard to assume the Continuum Hypothesis nor its negation. In the few instances where it is used, it *must* be explicitly stated that one is using it. If it is possible to prove something without assuming the Continuum Hypothesis, then it is generally considered best to prove it without using it.

2. Exercises

Question 1. Prove that the countable union of countable sets is countable.

Question 2. Prove that every infinite set contains a denumerable set.

Question 3.

(a) Let A and B be non-empty sets, and suppose that there exists an injection $\alpha : A \to B$. Prove that there exists a surjection $\beta : B \to A$.

(b) Let A and B be non-empty sets, and suppose that there exists a surjection $\beta : B \to A$. Prove that there exists an injection $\alpha : A \to B$.

Question 4. Prove that the set $\mathbb{I}$ of irrational real numbers is uncountable.

Question 5. Let A and B be non-empty sets. Would it make sense to say that $|A| < |B|$ if we could find an injection $\alpha : A \to B$ which is not injective?

Question 6. This one may be a bit or even a *lot* more challenging.

A number $\gamma \in \mathbb{R}$ is said to be **algebraic** if there exists a non-zero polynomial

$$p(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_m x^m$$

with coefficients in $\mathbb{Z}$ (i.e. $p_k \in \mathbb{Z}$, $0 \leq k \leq m$) such that $p(\gamma) = 0$.

For example, $\sqrt{2} \in \mathbb{R}$ is algebraic, since we can choose $p(x) = 2 - x^2$. This obviously has coefficients in $\mathbb{Z}$, is non-zero, and $p(\sqrt{2}) = 0$.

It is much, much harder to show that π is not algebraic. A real number which is *not* algebraic is said to be **transcendental**. Thus π is transcendental, and it can be shown that $e = 2.718...$ is also transcendental.

Let $\Lambda := \{\gamma \in \mathbb{R} : \gamma \text{ is algebraic}\}$. Prove that Λ is denumerable. Conclude that the set of transcendental numbers is uncountable.

Culture: In 1900, the mathematician David Hilbert produced a list of 23 open mathematical problems which he felt were amongst the most important facing mathematicians at that time. The seventh of these problems was the following (*Irrationalität und Transcendenz bestimmter Zahlen.*)

Let $\alpha \in \Lambda \setminus \{0, 1\}$ and $\beta \in \Lambda$ be irrational. Is α^β necessarily transcendental?

As a specific example, one may consider $\alpha = 2$ and $\beta = \sqrt{2}$; is $2^{\sqrt{2}}$ transcendental?

Notes.

- In fact, Hilbert even allowed for the numbers α and β to be complex. (The definition of an algebraic complex number is the same as for a real number.) Thus his question also applies to
$$e^\pi = (-1)^{-i} = i^{-2i}.$$
- Interesting note:
$$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2.$$

As it turns out, Hilbert's problem was solved independently by Aleksandr Gelfond and Theodor Schneider, and is known as the

The Gelfond-Schneider Theorem. If α and β are algebraic numbers with $\alpha \neq 0, 1$ and β irrational, then α^β is transcendental.

Interestingly enough (and if this doesn't make you the most popular person at your next party – whichever decade that might occur in – then I don't know what will), Schneider learnt about this problem as a graduate student by attending a seminar of *Carl Siegel's* at Frankfurt University. He eventually became Siegel's PhD student, and Siegel gave him a number of possible problems upon which he could work, but Schneider chose this one. Although Siegel had mentioned in the seminar that this problem was open, he didn't mention that it was Hilbert's Seventh Problem until Schneider produced his proof.

The most accessible proof of this result that I know of is

Hille, E., *Gelfond's solution of Hilbert's Seventh Problem*, Amer. Math. Monthly **49** (1942), 654-661.

It still requires a background in algebra and complex analysis.

Cardinality II
L.W. Marcoux
September 18, 2020

1. Cardinal Arithmetic

1.1. In this section we shall briefly examine sums, products and powers of cardinal numbers. Finite numbers do not provide the best intuition, since we don't expect numbers other than 0 and 1 to satisfy $\lambda^2 = \lambda$, for example. This equality will be satisfied by infinite cardinals, as we shall soon see. We begin with an extremely useful result which is the usual tool for proving that two sets have the same cardinality. Although the result looks obvious, its proof is surprisingly non-obvious.

1.2. Theorem. *Let A and B be sets. If $|A| \le |B|$ and $|B| \le |A|$, then $|A| = |B|$.*
Proof.
This is currently a "bonus question" on one of the assignments.

$\square$

Using Theorem 1.2, we can prove the following:

1.3. Theorem. $c = |\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.
Proof. By Theorem 1.2, it suffices to prove that $c \le |\mathcal{P}(\mathbb{N})|$ and $|\mathcal{P}(\mathbb{N})| \le c$.
To see that $|\mathcal{P}(\mathbb{N})| \le c$, define

$$\begin{array}{rccc} f: & \mathcal{P}(\mathbb{N}) & \to & \mathbb{R} \\ & A & \mapsto & 0.a_1\ a_2\ a_3\ \ldots \end{array}$$

where $a_n = 0$ if $n \notin A$ and $a_n = 1$ if $n \in A$. It is not hard to verify that f is injective.
To see that $c \le |\mathcal{P}(\mathbb{N})|$, first note that $|\mathbb{N}| = |\mathbb{Q}|$ and hence (**exercise**) $|\mathcal{P}(\mathbb{N})| = |\mathcal{P}(\mathbb{Q})|$. Next, let

$$\begin{array}{rccc} g: & \mathbb{R} & \to & \mathcal{P}(\mathbb{Q}) \\ & x & \mapsto & \{y \in \mathbb{Q} : y < x\}. \end{array}$$

If $x_1 < x_2$ in $\mathbb{R}$, then there exists $q \in \mathbb{Q}$ so that $x_1 < q < x_2$, and so $q \notin g(x_1)$ but $q \in g(x_2)$, showing that $g(x_1) \ne g(x_2)$ and so g is injective.

$\square$

1.4. Definition. *Let α, β be cardinal numbers.*
*The **sum** $\alpha + \beta$ of α and β is defined to be the cardinal $|A \cup B|$, where A and B are* disjoint *sets such that $|A| = \alpha$ and $|B| = \beta$.*

*The **product** $\alpha\ \beta$ of α and β is the cardinal number $|A \times B|$, where A and B are sets with $|A| = \alpha$ and $|B| = \beta$.*

*The **power** β^α is defined as $|B^A|$, where A and B are sets with $|A| = \alpha$ and $|B| = \beta$.*

1.5. In "**ordinal arithmetic**", one defines $0 = \varnothing$, $1 = \{0\} = \{\varnothing\}$, $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$, $3 = \{0, 1, 2\}$, etc. In a mild abuse of notation, the same notation is used to denote the corresponding cardinal.

The proof of Theorem 1.3 shows that if A is any set and B is a subset of A, then B corresponds to a unique function $f_B : A \to \{0, 1\}$ given by $f_B(a) = 0$ if $a \notin B$, and $f_B(a) = 1$ if $a \in B$. (This is often called the **characteristic function** or the **indicator function** of B in A.)

The map $B \mapsto f_B$ is a bijection between $\mathcal{P}(A)$ and $\{0, 1\}^A = 2^A$. Thus $|\mathcal{P}(A)| = |2^A|$, and in particular, $|2^{\mathbb{N}}| = |\mathcal{P}(\mathbb{N})| = c$. But, from the definition of cardinal powers, this says that $c = |2^{\mathbb{N}}| = |2|^{|\mathbb{N}|} = 2^{\aleph_0}$.

1.6. Lemma.

(i) *If A is an infinite set, then A contains a denumerable subset.*
(ii) *If A is an infinite set and B is a finite set, then $|A| + |B| = |A|$.*

Proof.

(i) Since $A \neq \varnothing$, there exists $x_1 \in A$. Then $A \setminus \{x_1\} \neq \varnothing$, otherwise A would be finite.

In general, for $n \geq 1$, having chosen $\{x_1, x_2, ..., x_n\} \subseteq A$, we know that $A \setminus \{x_1, x_2, ..., x_n\} \neq \varnothing$, so we can find $x_{n+1} \in A \setminus \{x_1, x_2, ..., x_n\}$. (Doing this for all $n \geq 1$ requires the Axiom of Choice - or at least a weak version of it.)

The function $f : \mathbb{N} \to A$ defined via $f(n) = x_n$ is an injection, and it is a bijection between $\mathbb{N}$ and $B = \operatorname{ran} f = \{x_n\}_{n=1}^{\infty}$. Thus $|B| = \aleph_0$ and B is a denumerable subset of A.

(ii) Let A be an infinite set, and let $D \subseteq A$ be a denumerable subset of A. Suppose that $B = \{b_1, b_2, ..., b_n\}$. (We may suppose that $B \cap A = \varnothing$ (*why?*)). Define a map

$$\begin{array}{rccc} f: & A \cup B & \to & A \\ & z & \mapsto & z \text{ if } z \in A \setminus D \\ & b_i & \mapsto & d_i \text{ if } 1 \leq i \leq n, \\ & d_k & \mapsto & d_{k+n}, \text{ for all } k \geq 1. \end{array}$$

Then f is an bijection of $A \cup B$ onto A, and so $|A| + |B| = |A|$.

□

1.7. Theorem. *Let α, β, and γ be cardinal numbers. Then*

(i) *$\alpha + \beta$ is well-defined. That is, if $|A| = |C|$, $|B| = |D|$ and $A \cap B = \varnothing = C \cap D$, then $|A \cup B| = |C \cup D|$.*
(ii) *$\alpha + \beta = \beta + \alpha$ and $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.*
(iii) *If β is infinite and $\alpha \leq \beta$, then $\alpha + \beta = \beta$.*

Proof.

(i) **Exercise.**
(ii) **Exercise.**
(iii) The case where $\alpha < \aleph_0$ is Lemma 1.6.

First let us show that $\beta + \beta = \beta$. Choose a set B with $|B| = \beta$. Then $B \times 2 = (B \times \{0\}) \cup (B \times \{1\})$ is the union of two disjoint sets equipotent with B, so it suffices to show that $|B \times 2| = |B| = \beta$.

Let $\mathcal{F} = \{(X, f) : X \subseteq B \text{ and } f : X \to X \times 2 \text{ is a bijection}\}$, partially ordered by $(X_1, f_1) \leq (X_2, f_2)$ if $X_1 \subseteq X_2$ and $f_2|_{X_1} = f_1$.

If $X \subseteq B$ is denumerable, then $|X \times 2| = |X| = \aleph_0$ by Example **??**, and hence $\mathcal{F} \neq \varnothing$.

Suppose that $\mathcal{C} = \{(X_\alpha, f_\alpha)\}_{\alpha\in\Lambda}$ is a chain in $\mathcal{F}$.

Let $X = \cup_{\alpha\in\Lambda} X_\alpha$. For $x \in X$, choose $\alpha \in \Lambda$ such that $x \in X_\alpha$. Define $f(x) = f_\alpha(x)$. Then f is well-defined (*why?*). Moreover, (X, f) is an upper bound for $\mathcal{C}$; i.e. $f : X \to X \times 2$ is a bijection (**exercise**).

By Zorn's Lemma, $\mathcal{F}$ has a maximal element (Y, g). We claim that $B \setminus Y$ is finite. Otherwise, choose a denumerable set $Z \subseteq B \setminus Y$. Since $|Z| = |Z \times 2| = \aleph_0$, there exists a bijection $h : Z \to Z \times 2$. Define a bijection

$$\begin{array}{rccl} h: & Y \cup Z & \to & (Y \cup Z) \times 2 \\ & w & \mapsto & \begin{cases} g(w) & \text{if } w \in Y, \\ h(w) & \text{if } w \in Z. \end{cases} \end{array}$$

Then $(Y \cup Z, h) > (Y, g)$, contradicting the maximality of (Y, g).

This shows that $B \setminus Y$ is finite. Hence $|Y| = |B| = \beta$, and so $\beta = |Y| = |Y \times 2| = |Y| + |Y| = \beta + \beta$.

Finally, in general we have $\beta \leq \alpha + \beta \leq \beta + \beta = \beta$, so that by Theorem 1.2, $\alpha + \beta = \beta$.

$\square$

1.8. Theorem. *Let $\alpha, \beta, \gamma, \delta$ be cardinal numbers. Then*

(i) *$\alpha \cdot \beta$ is well-defined. That is, if $|A| = |C|$, $|B| = |D|$, then $|A \times B| = |C \times D|$.*
(ii) *$\alpha \cdot \beta = \beta \cdot \alpha$; $\alpha(\beta \cdot \gamma) = (\alpha \cdot \beta)\gamma$; and $\alpha(\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$.*
(iii) *$0 \cdot \alpha = 0$.*
(iv) *If $\alpha \leq \beta$ and $\gamma \leq \delta$, then $\alpha \cdot \gamma \leq \beta \cdot \delta$.*
(v) *$\alpha \cdot \alpha = \alpha$ if α is infinite.*

Proof.

(i) **Exercise.**
(ii) **Exercise.**
(iii) **Exercise.**
(iv) **Exercise.**
(v) Suppose that $|A| = \alpha$. Let $\mathcal{F} = \{(X, f) : X \subseteq A, f : X \to X \times X \text{ is a bijection}\}$, partially ordered by $(X_1, f_1) \leq (X_2, f_2)$ if $X_1 \subseteq X_2$ and $f_2|_{X_1} = f_1$.

Since A is infinite, it contains a denumerable set X. Now by Example **??**, if $X \subseteq A$ is denumerable, then there exists a function f so that $(X, f) \in \mathcal{F}$, and thus $\mathcal{F} \neq \varnothing$. By Zorn's Lemma (as before), there exists a maximal element (Y, g) in $\mathcal{F}$.

Then $|Y| \cdot |Y| = |Y|$, so it suffices to show that $|Y| = \alpha$.

Assume that $|Y| < \alpha$. Since $\alpha = |Y| + |A \setminus Y|$, it follows that $|A \setminus Y| = \alpha$, and so $|Y| < |A \setminus Y|$. Thus $A \setminus Y$ has a subset Z with $|Z| = |Y|$. Then $Y \times Z, Z \times Y$ and $Z \times Z$ are disjoint, infinite sets with cardinality $|Y|$, and so

$$\begin{aligned} |(Y \times Z) \cup (Z \times Y) \cup (Z \times Z)| &= |Y \times Z| + |Z \times Y| + |Z \times Z| \\ &= (|Y| \cdot |Y|) + (|Y| \cdot |Y|) + (|Y| \cdot |Y|) \\ &= |Y| + |Y| + |Y| \\ &= |Y| \\ &= |Z|. \end{aligned}$$

Thus there exists a bijection $h : Z \to (Y \times Z) \cup (Z \times Y) \cup (Z \times Z)$.

Define the map

$$\begin{array}{rccl} m: & Y \cup Z & \to & (Y \cup Z) \times (Y \cup Z) \\ & x & \mapsto & \begin{cases} g(x) & \text{if } x \in Y \\ h(x) & \text{if } x \in Z \end{cases} . \end{array}$$

Then m is a bijection and so $(Y \cup Z, m) \in \mathcal{F}$ with $(Y, g) < (Y \cup Z, m)$, contradicting the maximality of (Y, g) in $\mathcal{F}$.

This contradiction shows that $|Y| = \alpha$, and we are done, as $g : Y \to Y \times Y$ is the bijection which implies that $\alpha = \alpha \cdot \alpha$.

□

1.9. Remark. The perspicacious reader (and just about no one else) will have picked up a very subtle use of the Axiom of Choice in the above proof. You will recall that we said that with $Y \subseteq A$, $\alpha = |A| = |Y| + |A \setminus Y|$ and $|Y| < \alpha$ implies that $|A \setminus Y| = \alpha$.

In fact, how do we know this? Theorem 1.7 assures us that if $|Y| \leq |A \setminus Y|$ or if $|A \setminus Y| \leq |Y|$, then

$$\alpha = |A| = |Y| + |A \setminus Y| = \max(|Y|, |A \setminus Y|),$$

and thus $|A \setminus Y| = \alpha$. But how do we know that $|Y|$ and $|A \setminus Y|$ are comparable? In fact, for that we must go back to Remark **??**, and the statement that given any two sets V and W, either there exists an injection from V into W (i.e. $|V| \leq |W|$), or there exists an injection from W into V (i.e. $|W| \leq |V|$). In other words, any two cardinal numbers are comparable! As pointed out in that Remark, however, this assumption is equivalent to the Axiom of Choice.

We thank Y. Zhang for pointing this out.

1.10. Theorem. *Let α, β and γ be cardinal numbers. Then*

(i) *α^β is well-defined. That is, if A_1, A_2, B_1, B_2 are sets with $|A_1| = \alpha = |A_2|$ and $|B_1| = \beta = |B_2|$, then $|A_1^{B_1}| = |A_2^{B_2}|$.*
(ii) *$(\alpha^\beta)^\gamma = \alpha^{(\beta\gamma)}$.*
(iii) *$\alpha^{(\beta+\gamma)} = \alpha^\beta \, \alpha^\gamma$.*

Proof.

(i) **Exercise.**
(ii) Let A, B and C be sets with $|A| = \alpha$, $|B| = \beta$ and $|C| = \gamma$. We must show that $|(A^B)^C| = |A^{B \times C}|$.

Now if $f \in A^{B \times C}$, then for each $c \in C$, the function f_c given by $f_c(b) := f(b, c)$ defines an element of A^B. Define $\varphi_f : C \to A^B$ by $\varphi_f(c) = f_c$. Then the correspondence

$$\begin{array}{rccc} \Phi: & A^{B \times C} & \to & (A^B)^C \\ & f & \mapsto & \varphi_f \end{array}$$

is a bijection. Indeed, if $\Phi(f) = \Phi(g)$ for $f, g \in A^{B \times C}$, then $\varphi_f = \varphi_g$, and so for every $c \in C$, $f_c = \varphi_f(c) = \varphi_g(c) = g_c$. But $f_c = g_c$ for all $c \in C$ implies that $f(b, c) = f_c(b) = g_c(b) = g(b, c)$ for all $b \in B$ and for all $c \in C$, so that $f = g$. This shows that Φ is injective.

Given $\tau \in (A^B)^C$, we see that $\tau(c) \in A^B$ for all $c \in C$, and so we define $f : B \times C \to A$ via $f(b, c) = [\tau(c)](b)$. It is clear that $f \in A^{B \times C}$ and that $\Phi(f) = \tau$, so that Φ is onto. Finally, since Φ is a bijection, $|A^{B \times C}| = |(A^B)^C|$, completing the proof.

(iii) Now suppose that $B \cap C = \varnothing$. We must show that $|A^{B \cup C}| = |A^B \times A^C|$. But every $f : B \cup C \to A$ is defined by its restrictions to B and C, so we are done.

$\square$

1.11. Example.

(a) $\aleph_0 + \aleph_0 = \aleph_0$.

(b) $\aleph_0 + c = c$.

(c) $c \cdot c = c$.

(d) $c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{(\aleph_0 \cdot \aleph_0)} = 2^{\aleph_0} = c$.

The Axiom of Choice
L.W. Marcoux
September 14, 2020

1. The Zermelo-Fraenkel Axioms.

1.1. I will begin this discussion with a disclaimer: I am not a set theorist. Frustratingly, I don't even play one on TV. That having been said, let's learn some *stuff*$^{\text{TM}}$.

1.2. Set theory is the study of specific collections, called... wait for it... *sets* of objects that are called *elements* of that set. *Pure set theory* deals with sets whose elements are again sets. Experts in set theory claim that the theory of *hereditarily-finite* sets – i.e. those finite sets whose elements are also finite sets, whose elements in turn are also finite sets, whose elements ... etc., is formally equivalent to arithmetic, which we shall neither define nor attempt to explain.

The form of set theory which interests us here arose out of attempts by mathematicians of the late 19th and early 20th centuries to try to understand infinite sets, and originated with the work of the German mathematician **Georg Cantor**. He had published a number of articles in number theory between 1867[1] and 1871[2], but began work on what was to become set theory by 1874.

In 1878, Cantor formulated was is now referred to as the *Continuum Hypothesis*(CH). Suppose that X is a subset of the real line. The Continuum Hypothesis asserts that either there exists a bijection between X and the set $\mathbb{N}$ of natural numbers, or there exists a bijection between X and the set $\mathbb{R}$ of real numbers.

Many well-known and leading mathematicians of his day attempted to prove this statement, including Cantor himself, as well as **David Hilbert**, who included it as the first of the twenty-three unsolved problems he presented at the Second International Congress of Mathematics in Paris, in 1900. Any attempt to prove this would require one to understand not only sets of real numbers, but sets in general.

Like mangoes, mathematical theories need time to ripen. Early attempts to define sets led to inconsistencies and paradoxes. Originally, it was thought that any property P could be used to define a set, namely the set of all sets (or even other objects) which satisfy property P. The mathematician-philosopher **Bertrand Russell** produced the following worrisome example:

Russell's Paradox. Let P be the property of sets: $X \notin X$. That is, a set X satisfies property P if and only if X does not belong to X.

Let
$$\mathfrak{R} := \{X \text{ a set} : X \text{ satisfies property } P\} = \{X \text{ a set} : X \notin X\}.$$
The question becomes: does $\mathfrak{R}$ satisfy property P? If so, the $\mathfrak{R} \notin \mathfrak{R}$, and so $\mathfrak{R} \in \mathfrak{R}$. If not, then $\mathfrak{R} \in \mathfrak{R}$, so $\mathfrak{R} \notin \mathfrak{R}$.

This is not a good state of affairs, and the German mathematician **Ernst Zermelo**, who was also aware of this paradox, thought that set theory should be axiomatised in

[1] the year of Canada's birth

[2] the year that the Government of Canada promised B.C. a railway which they delivered in 1885, by the way

order to be rigorous and preclude paradoxes such as the one above. In 1908, Zermelo produced a first axiomitisation of set theory, although he was unable at that time to prove that his axiomatic system was consistent. There was also a second problem, insofar as some logician/set theorists were concerned, namely: Zermelo avoided Russell's Paradox by means of an axiom which he referred to as the *Separation Axiom*. This, however, he stated using *second order logic*, which is considered less desirable (in the sense that it requires stronger hypotheses) than the version offered by the Norwegian **Thoralf Skolem** and the German **Abraham Fraenkel**, which relied only on so-called *first-order logic*.

It occurs that the Hungarian mathematician **John von Neumann** also added to the list of so-called *Zermelo-Fraenkel Axioms* we shall enumerate below by introducing the *Axiom of Foundation*. We are not sufficiently versed in the history of this area to be able to explain why Skolem and von Neumann's names don't appear on the list of Axioms, although von Neumann's name appears in so many other mathematical contexts that even if he were alive today, he would not have the right to complain.

1.3. The Zermelo-Fraenkel Axioms (ZF).

- The Null Set Axiom.
  There exists a set $\varnothing$, called the *empty set*, which has no elements.
- The Axiom of Extension.
  Two sets A and B are equal if and only if they have the same elements.
- The Axiom of Regularity, aka the Axiom of Foundation.
  Every non-empty set A contains an element B such that no element of A belongs to B.
- The Axiom of Specification.
  If A is a set and P is a property, then there exists a set
  $$B := \{b \in A : b \text{ satisfies property } P\}.$$
- The Axiom of Pairing.
  Given two sets A and B, there exists a set $\{A, B\}$ whose only elements are A and B. (Combined with the Null Set Axiom, we deduce that $\{A\}$ is also a set.)
- The Axiom of Union.
  If A is a set, then there exists a set $\cup A$, called the *union of* A, defined by
  $$\cup A = \{b : b \in B \text{ for some } B \in A\}.$$
- The Axiom of Power Sets.
  Given a set A, there exists a set $\mathcal{P}(A) = \{B : B \subseteq A\}$ whose elements are all of the subsets of A.
- The Axiom of infinity. There exists an infinite set. More specifically, there exists a set $\mathfrak{Z}$ such that $\varnothing \in \mathfrak{Z}$ and $A \in \mathfrak{Z}$ implies that $\cup\{A, \{A\}\} \in \mathfrak{Z}$.

1.4. Just a quick comment about the Axiom of Specification: why doesn't it lead right back to Russell's Paradox? Because the Axiom of Specification assumes that you *begin* with a set, and you look at the members of *that set* which satisfy a property P. In Russell's Paradox, you don't have a starting set – instead you *manufacture* a set out of thin air through a property P. This is the kind of object we now refer to as a *class*, which is an object more general than a set.

1.5. There is one Axiom we have yet to specify, which – given the title of this note – we might want to do sooner rather than later.

- THE AXIOM OF CHOICE. Let Λ be a non-empty set, and let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a collection of non-empty sets. Then there exists a function $f : \Lambda \to \cup_{\lambda\in\Lambda} A_\lambda$ such that $f(\gamma) \in A_\gamma$ for each $\gamma \in \Lambda$.

In layman's terms, if you have a (non-empty) collection of (non-empty) sets, then you can choose an element from each set. In technical terms, we call such a function f a **choice function**.

We write (ZFC) to indicate (ZF) plus the Axiom of Choice.

1.6. Remark. There are a couple of remarks we should make here.

(a) First: what the heck is going on? This is so obviously true that it isn't even worth mentioning! How could it not be true?

(b) Doesn't it follow from the Axioms above? If not, what are they good for?

1.7. I sympathise with those who think it's obviously true. I really do. For one thing, the Axiom of Choice becomes an issue only if we have *a lot* of sets – in fact, only when we are dealing with infinitely many such sets. (Whether those sets are finite or infinite is not the issue – the issue is how many sets we have.)

The formal reason why this is so can be summarised in the following way: (ZFC) is a "*theory in first-order logic*", and our ability to "*choose*" an element from a (single) non-empty set is an application of a rule from that theory that carries the groovy moniker *existential instantiation*. We do not claim to be an expert in the theory of first-order logic, but the upshot that a hard-working, good-looking mathematician but non-logician like me can safely take away from this statement is that given a non-empty set, (ZF) allows you to pick an element of that set. If you are interested in the details of first-order logic, then be aware that treatments now exist to cure this, but also that some people exhibit these symptoms for years and still manage to lead productive lives, both inside and outside of mathematics.

Another thing to note is that existential instantiation is not a *constructive* argument. That is, we are not given a *method* of choosing an element from a non-empty set A – all we know is that it is possible. If we relabel A as A_λ, then saying that produces an element $a \in A$ is the same as saying that there exists a function $f : \{\lambda\} \to A_\lambda$ that satisfies $f(\lambda) = a$. That is, we have our choice function.

Having done this once, first-order logic allows us to repeat this process finitely often. That is, given *finitely many* non-empty sets $A_1, A_2, \ldots, A_n$, we can find a function $f : \{1, 2, \ldots, n\} \to \cup_{k=1}^n A_k$ such that $f(k) \in A_k$, $1 \le k \le n$.

What first-order logic and the Zermelo-Fraenkel Axioms do not do, however, is to carry out this procedure infinitely often, willy-nilly. We shall try to make the expression "*willy-nilly*" more precise below.

Willy-Nilly, or not Willy-Nilly. That is the question.

1.8. While it is important for one to try to be original, it is also important for one to acknowledge the accomplishments of those who came before one. In particular, the same Bertrand Russell referenced above once said:

> *To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed.*

It is interesting to try to figure out what exactly Russell meant by this, and what it says about his sartorial predilections.

The difference to which Bertrand Russell is alluding is related to the non-constructiveness of existential instantiation. Russell's underlying hypothesis while making this comment is that there is no way to distinguish between two socks in a pair. That is, he is assuming that the two socks in a given pair are for all intents and purposes identical. Thus, to pick a sock from a pair requires the rules of first-order logic to run (ZF) theory. This allows us to pick a sock from each pair for *finitely many pairs*, but there is no way to do this for *infinitely many pairs at once*. Thus in order to pick a sock from amongst each of infinitely many pairs of socks, you need something more than (ZF) theory, and the Axiom of Choice grants you your wish.

A second underlying assumption of Russell's is that each pair of shoes includes a left shoe and a right shoe. This changes everything. We no longer require existential instantiation to select a shoe from a pair – instead, we just specify *the left* shoe. (We could have said the right.) Thus, to "pick a shoe from each of infinitely many pairs of shoes", we simply specify that we shall always pick the left shoe, thereby circumventing existential instantiation altogether. Having a method to pick an element of the non-empty sets we are dealing with allows us to avoid having recourse to the Axiom of Choice.

1.9. Let us consider how such a thing might arise in practice in a mathematical setting. Let $\{A_t : t \in \mathbb{R}\}$ be a collection of non-empty subsets $\varnothing \neq A_t \subseteq \mathbb{N}$, one such set for every real number $t \in \mathbb{R}$. We would like to choose one element from each set. Can we do this in (ZF), or do we require the Axiom of Choice?

If we had an explicit *method* to choose the element $a_t \in A_t$, then we could indeed avoid the Axiom of Choice. But how can we specify an element of A_t without knowing exactly what A_t is? The natural numbers have a very special and very useful property: they are *(well-) ordered.*

Before defining a well-order we remind the reader that a **relation** ρ on a non-empty set X is a subset of the set $X \times X = \{(x, y) : x, y \in X\}$. Often, we write $x \, \rho \, y$ to mean the ordered pair $(x, y) \in \rho$. This is especially true when dealing, for example, with the usual relation $\leq$ for real numbers: no one writes $(x, y) \in \leq$; we all write $x \leq y$. Incidentally, the notation $\leq$ is used not only for the relation "less than or equal to" for real numbers; it frequently appears to indicate a specific kind of a relation known as a *partial order* on an arbitrary set, which we now define.

1.10. Definition. *A relation $\leq$ on a non-empty set X is said to be a* ***partial order*** *if, given x, y, and $z \in X$,*

(a) *$x \leq x$;*
(b) *if $x \leq y$ and $y \leq x$, then $x = y$; and*
(c) *if $x \leq y$ and $y \leq z$, then $x \leq z$.*

The prototypical example of a partial order is the partial order of inclusion on the power set $\mathcal{P}(A)$ of a non-empty set A. That is, set $\mathcal{P}(A) = \{B : B \subseteq A\}$ and for $B, C \in \mathcal{P}(A)$, we set $B \subseteq C$ to mean that every member of B is a member of C. It is easy to see that $\subseteq$ is a partial order on $\mathcal{P}(A)$.

The word *partial* refers to the fact that given two elements B and C in $\mathcal{P}(A)$, they might not be comparable. For example, if $A = \{x, y\}$, $B = \{x\}$ and $C = \{y\}$, then it is not the case that $B \subseteq C$, nor is it the case that $C \subseteq B$. Only *some* subsets of $\mathcal{P}(A)$ are comparable.

In dealing with the natural numbers, we observe that they come equipped with a partial order which we typically denote by $\leq$. In this setting, however, *any* two natural numbers are comparable. If (X, ρ) is a partial ordered set, and if $x, y \in X$ implies that either $x\rho y$ or $y\rho x$, then we say that ρ is a **total order** on X.

Thus $(\mathbb{N}, \leq)$ is a **totally ordered set**.

1.11. It gets even better (and yes, you should seriously ask yourself whether you deserve it). The most useful property of $\mathbb{N}$ for our current purposes is that it possesses one more *very* striking property:

given any non-empty subset $\varnothing \neq H \subseteq \mathbb{N}$, it admits a **minimum element**;

that is, there exists an element $m \in H$ such that $m \leq h$ for all $h \in H$.

If (X, ρ) is a partially ordered set with the property that every non-empty subset Y of X admits a minimum element, then we say that (X, ρ) is **well-ordered**.

The observation above is that $(\mathbb{N}, \leq)$ is well-ordered.

1.12. What is the relevance of this to the Axiom of Choice? Returning to the example in paragraph 1.9, we were given a collection $\{A_t : t \in \mathbb{R}\}$ of non-empty subsets of $\mathbb{N}$. Since we have just seen that $(\mathbb{N}, \leq)$ is well-ordered, we can specify a choice function

$$\begin{array}{rccl} f: & \mathbb{R} & \to & \cup_{t\in\mathbb{R}} A_t \\ & t & \mapsto & \min A_t. \end{array}$$

That is, we have a *rule* for specifying which element of A_t we are choosing – we are choosing the minimum element of A_t which we know exists (and is unique). We did not need to know in advance what A_t was, and we did not need existential instantiation to define $f(t)$.

Thus we don't need the Axiom of Choice to pick an element from each subset A_t simultaneously.

1.13. What if we had a collection $\{B_t : t \in \mathbb{R}\}$ of non-empty subsets of $\mathbb{R}$? That is, $\varnothing \neq B_t \subseteq \mathbb{R}$ for each $t \in \mathbb{R}$? Do we need the Axiom of Choice to choose an element from each set?

It is easy to check that $(\mathbb{R}, \leq)$ is a totally ordered set, as was $(\mathbb{N}, \leq)$. On the other hand, the open interval $(0, 1) = \{x \in \mathbb{R} : 0 < x < 1\}$ certainly doesn't have a minimum element, so that $(\mathbb{R}, \leq)$ is not a well-ordered set.

How would we specify an element of B_t without first knowing what B_t is? Unless we first know something about the set B_t, there *is no way* of specifying which element of B_t we are choosing. It follows that we *do need* the Axiom of Choice to do this.

(Of course, every subset of $\mathbb{N}$ is a subset of $\mathbb{R}$, so if we had originally chosen $B_t = A_t$ for all t, then we could have just used the "minimum" element method above. The point here is that if you have no information about B_t other than the fact that it is non-empty, then you can't use that argument here.)

1.14. We have not forgotten the second question raised in Remark 1.6, namely: doesn't the Axiom of Choice follow from the Axioms of (ZF)? In fact, we have already partially addressed this – (ZF) and first-order logic do *not* allow us to choose an element from each of infinitely-many sets at a time, in large part because of the non-constructive nature of existential instantiation. (For those who are wondering, yes, your humble author does love that phrase.)

There remains the question: does it contradict any of the (ZF) axioms? The simple answer is "No", although the proof that it does not is anything but simple. That (ZFC) is *consistent* (i.e. doesn't contain any inherent contradictions) was first demonstrated by the Austro-Hungarian mathematician **Kurt Gödel** in 1938. Does that mean that we can't do mathematics without the Axiom of Choice? Interestingly enough, in 1962, the American mathematician **Paul Cohen** demonstrated that one can assume all of the Axioms of Zermelo-Fraenkel Theory, and also assume that the Axiom of Choice is FALSE(!!!), and *still* arrive at a consistent mathematical system. For this and for his work on the Continuum Hypothesis (he showed that (CH) is independent of (ZFC)), he was awarded the Fields medal, which is generally considered to be the mathematical equivalent of the Nobel Prize. Given how prestigious and merited the Nobel Peace prize invariably is, a prize in mathematics is... (we leave it as an exercise for the reader to complete this sentence).

2. An apology for the Axiom of Choice.

2.1. As important as it is to choose one's socks wisely, if the only thing that every resulted from the Axiom of Choice was our ability to choose a sock from each pair in an infinite collections of pairs of socks, then we would not be talking about it today.

The issue is that this seemingly innocuous assumption has major implications. In particular, it is known that to imply that every vector space (over any field) admits a vector-space basis. Given how useful vector space bases are when studying linear algebra, it would be beyond tragic to lose them. Of course, one might argue that one should instead drop the Axiom of Choice and just assume that every vector space admits a basis. But the joke is then upon the one who argued this: as it so happens, if we assume that every vector space admits a basis, then one can also prove that the Axiom of Choice holds.

In other words, the Axiom of Choice is equivalent to the statement that every vector space admits a basis. One of the statements is true if and only if the second is true.

2.2. The number of statements that are equivalent to the Axiom of Choice makes is huge. Entire books are written on this subject - including the book *Equivalents of the Axiom of Choice, II* by Herman Rubin and Jean Rubin, which includes 250 statements, each equivalent to the Axiom of Choice. Below we shall mention three such equivalent statements which are among the most important.

2.3. The Axiom of Choice II. This version is a slight modification of the Axiom of Choice, and you might want to think about how you would prove that it is equivalent to the Axiom of Choice:

> Let $\{A_\lambda\}_{\lambda\in\Lambda}$ be a non-empty collection of non-empty *disjoint* sets. That is, $\lambda_1 \neq \lambda_2 \in \Lambda$ implies that $A_{\lambda_1} \cap A_{\lambda_2} = \varnothing$.
>
> Then there exists a function $f : \Lambda \to \cup_{\lambda\in\Lambda} A_\lambda$ such that $f(\lambda) \in A_\lambda$ for all $\lambda \in \Lambda$.

2.4. The Well-Ordering Principle.

Let $\varnothing \neq X$ be a non-empty set. Then there exists a relation ρ on X such that (X, ρ) is well-ordered.

If $(X, \leq)$ is a partially ordered set, then a subset $C \subseteq X$ is said to be a **chain** in X if $c, d \in C$ implies that either $c \leq d$ or $d \leq c$. In other words, C is totally ordered using the relation inherited from X.

2.5. Zorn's Lemma.

Let $(X, \leq)$ be a partially ordered set. Suppose that for each chain C in X, there exists an element $\mu_C \in X$ such that $c \leq \mu_C$ for all $c \in C$. (We say that C is **bounded above** in X.)

Then $(X, \leq)$ admits a **maximal element** m. That is, there exists $m \in X$ such that if $x \in X$ and $m \leq x$, then $x = m$.

2.6. A note about maximal elements of partially ordered sets. Our decision to use "*maximal*" instead of "*maximum*" is not just fancy. They usually mean different things.

Let $\mathcal{A} = \{\varnothing, \{x\}, \{y\}, \{z\}, \{x,y\}, \{x,z\}, \{y,z\}\}$ be the set of *proper* subsets of $X = \{x, y, z\}$, and partially order $\mathcal{A}$ by inclusion. Thus $\{x\} \leq \{x,y\}$ because $\{x\} \subseteq \{x,y\}$.

Observe that $\{x,y\}$ is a maximal element of $\mathcal{A}$. *Nothing is bigger than* $\{x,y\}$ in $\mathcal{A}$. But $\{z\} \not\subseteq \{x,y\}$, and so $\{x,y\}$ is not a maximum element of $\mathcal{A}$. Note that $\{x,z\}$ is also a maximal elements. Maximal elements need not be unique. (Are maximum elements unique when they exist? Think about it!)

The issue is that a maximal element doesn't have to be comparable to every element in the set, while a maximum element does. In other words, a maximal element only has to be bigger than or equal to those things that you can actually compare it to, while a maximum element has to be bigger than or equal to everything in the set! In this example, $\varnothing$ would be a mimimum, and thus also a minimal element of $(\mathcal{A}, \leq)$.

That the Well-Ordering Principle implies the Axiom of Choice is definitely within reach of the interested reader. We leave it as an exercise.

2.7. The following quote is worth keeping in mind:

The Axiom of Choice is obviously true, the Well-Ordering Principle obviously false, and who can tell about Zorn's lemma?

Jerry Bona

It's funny people like Professor Bona who put the world of mathematics into perspective and allow us to laugh at ourselves for days at a time by revealing some heretofore hidden lacuna in our intuition. It is also worth keeping in mind that the Well-Ordering Principle is *not* constructive. It does not suggest any method of actually finding a well-order on a given set X; it simply says that one exists.

The question of how you might try to well-order the complex numbers is certainly worth thinking about.

2.8. Although your humble author has certainly not had a quote that compares in notoriety to that mentioned above, it strikes your humble author that the Well-Ordering Principle is not as "obviously false" as a somewhat different consequence of the Axiom of Choice, namely the Banach-Tarski Paradox. This we shall address in the next section.

3. The prosecution rests its case.

3.1. Given how "obviously true" the Axiom of Choice is, and how perfectly intuitive it may seem, it would seem outrageous not to include it amongst the axioms of set theory.

Consider then the following story, which by all accounts is true (see J. Mycielski, Notices of the AMS **53**, no. 2, page 209): **Alfred Tarski**, a Polish logician, proved that the Axiom of Choice was equivalent to the following statement.

> *Given any infinite set X, there is a bijection from X to $X \times X := \{(x, y) : x, y \in X\}$.*

He then submitted this result to the *Comptes Rendus de l'Académie des Sciences*, where it was refereed by two famous French mathematicians, **Maurice René Fréchet**, and **Henri Lebesgue**. Both referees rejected Tarski's paper. Fréchet wrote that the equivalence of two well-known truths is not a new result. Lebesgue claimed that both statements were false, and so their equivalence was of no interest.

Tarski is said to have never submitted another paper to *Comptes Rendus*. (Feel free to revise the sentence you completed at the end of Section 1.)

3.2. Perhaps the most damning evidence one can provide against the Axiom of Choice is the following result which the Axiom of Choice implies, and was alluded to above.

3.3. The Banach-Tarski Paradox. *Let $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}$, so that S denotes the (solid) ball in $\mathbb{R}^3$ centred at the origin. It is possible to partition the ball S into finitely many disjoint pieces in such a way, that by translating and rotating them, they can be rearranged into two disjoint identical balls, each having the same volume as S.*

Note that stretching, bending, and twisting pieces is not allowed. Once you have determined the pieces, you can only translate and rotate them.

If you are very thirsty, you might want to try this out on a bag of oranges at home. Note that the inordinate amount of time you spend cleaning up your mess, as well as your abject failure in accomplishing this task, might result from the fact that most knives on the market today produce so-called "measurable" cuts of your oranges. The same applies to grapefruit, and need we say (?), most other food products.

The Banach-Tarski Paradox is, to coin a phrase, "paradoxical". It is infuriating, in large measure due to the fact that it is so highly counterintuitive. On the other hand, once you accept the Axiom of Choice, this result follows as a consequence, and you have no choice but to accept it and move on with your life.

4. So what to do?

4.1. Zermelo-Fraenkel theory plus Choice (ZFC) is popular because it works. It works well. The Banach-Tarski Paradox and the existence of non-measurable sets aside, it produces so many great results that the vast majority of mathematicians use it without hesitation. Having said that – the Axiom of Choice is considered *special*, and unlike the other (ZF) axioms. In general, when it is not required to prove a result, a proof that avoids it is considered better than a proof that uses it. Also, it is often considered "good form" to indicate when it is being used, although some uses of the Axiom of Choice have become so commonplace that mathematicians expect the reader to realise on their own that the Axiom of Choice has been used in the argument.

There is a group of mathematicians who feel that to prove that a mathematical object exists, one must construct it. They are referred to as *constructivists* (at least this is how they are referred to in polite company). Your humble author has not met a constructivist mathematician himself, and it occurs to your humble author as he writes this that if we were to use constructivist logic, then we could not conclude that one exists, could we? [How do I know that the articles that refer to them aren't all made up?] But this author is not a constructivist, and so is willing to accept that they do exist.

4.2. We conclude by indicating that the Continuum Hypothesis, which was also shown by Gödel and by Cohen to be independent of (ZFC), is not universally accepted at all. Any argument that uses the Continuum Hypothesis should mention where it is being used, and should be at all avoided if there is any way of obtaining the same result without it. Good results have been proven that require it, and are usually stated as: *assuming the continuum hypothesis, we show that ...* .

4.3. Quiz. Can you avoid the Axiom of choice to choose an element from

(a) a finite set?
(b) each member of an infinite set of sets, each of which has only one element?
(c) each member of a finite set of sets if each of the members is infinite?
(d) each member of an infinite set of sets of rationals?
(e) each member of an infinite set of finite sets of reals?
(f) each member of a denumerable set of sets if each of the members is denumerable?
(g) each member of an infinite set of sets of reals?
(h) each member of an infinite set of sets of integers?

Equivalent formulations of the Axiom of Choice
L.W. Marcoux
September 18, 2020

1. Three wise formulations

In this note we shall provide a proof of the equivalence of the Axiom of Choice, Zorn's Lemma and the Well-Ordering Principle. We remind the reader that a **poset** is a partially ordered set, as defined in a previous note on the Axiom of Choice. We begin with the definition of an **initial segment**, which will be required in the proof.

We also remind the reader that the statement that, given a collection $\{X_\lambda\}_{\lambda\in\Lambda}$ of non-empty sets, the existence of a **choice function** $f : \Lambda \to \cup_{\lambda\in\Lambda} X_\lambda$ such that $f(\lambda) \in X_\lambda$ for all $\lambda \in \Lambda$ is the statement that

$$\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing,$$

for the simple reason that

$$\prod_{\lambda\in\Lambda} X_\lambda := \{f : \Lambda \to \cup_{\lambda\in\Lambda} X_\lambda \text{ a function such that } f(\lambda) \in X_\lambda \text{ for all } \lambda \in \Lambda\}.$$

1.1. Definition. *Let $(X, \leq)$ be a poset, $C \subseteq X$ be a chain in X and $d \in C$. We define*

$$P(C, d) = \{c \in C : c < d\}.$$

An ***initial segment*** *of C is a subset of the form $P(C, d)$ for some $d \in C$.*

1.2. Example.

(a) For each $r \in \mathbb{R}$, $(-\infty, r)$ is an initial segment of $(\mathbb{R}, \leq)$.
(b) For each $n \in \mathbb{N}$, $\{1, 2, ..., n\}$ is an initial segment of $\mathbb{N}$.

1.3. Theorem. *The following are equivalent:*

(i) *The Axiom of Choice (AC): given a non-empty collection $\{X_\lambda\}_{\lambda\in\Lambda}$ of non-empty sets, $\prod_{\lambda\in\Lambda} X_\lambda \neq \varnothing$.*
(ii) *Zorn's Lemma (ZL): Let $(Y, \leq)$ be a poset. Suppose that every chain $\mathcal{C} \subseteq Y$ has an upper bound. Then Y has a maximal element.*
(iii) *The Well-Ordering Principle (WO): Every non-empty set Z admits a well-ordering.*

Proof.

(i) implies (ii): This is the most delicate of the three implications. We shall argue by contradiction.

Suppose that $(X, \leq)$ is a poset such that every chain in X is bounded above, but that X no maximal elements. Given a chain $C \subseteq X$, we can find an upper bound u_C for C. Since u_C is not a maximal element, we can find $v_C \in X$ with $u_C < v_C$. We shall refer to such an element v_C as a **strict upper bound** for C.

By the Axiom of Choice, for each chain C in X, we can choose a strict upper bound $f(C)$. If $C = \varnothing$, we arbitrarily select $x_0 \in X$ and set $f(\varnothing) = x_0$.

We shall say that a subset $A \subseteq X$ satisfies **property L** if

(I) The partial order $\leq$ on X when restricted to A is a well-ordering of A, and
(II) for all $x \in A$, $x = f(P(A, x))$.

• **Claim 1:** *if $A, B \subseteq X$ satisfy property L and $A \neq B$, then either A is an initial segment of B, or B is an initial segment of A.*

Without loss of generality, we may assume that $A \setminus B \neq \varnothing$. Let

$$x = \min \{a \in A : a \notin B\}.$$

Note that x exists because A is well-ordered. Then $P(A, x) \subseteq B$. We shall argue that $B = P(A, x)$. If not, then $B \setminus P(A, x) \neq \varnothing$, and using the well-orderedness of B,

$$y = \min \{b \in B : b \notin P(A, x)\}$$

exists. Thus $P(B, y) \subseteq P(A, x)$.

Let $z = \min (A \setminus P(B, y))$. Then $z \leq x = \min (A \setminus B)$.

- **Subclaim 1:** $P(A, z) = P(B, y)$.
  By definition, $P(A, z) \subseteq P(B, y)$.
  To obtain the reverse inclusion, we first argue that if $t \in P(B, y) = A \cap P(B, y)$, then $P(A, t) \cup \{t\} \subseteq P(B, y)$. By hypothesis, $t \in P(B, y)$, so suppose that $u \in P(A, t)$. Now $t \in P(B, y) \subseteq P(A, x)$, so $u < t < x$ implies that $u \in P(A, x)$. In other words, $P(A, t) \subseteq P(A, x) \subseteq B$. But then $u \in B$ and $u < t < y$ implies that $u \in P(B, y)$.
  We now have that if $s \in P(B, y)$, then $P(A, s) \cup \{s\} \subseteq P(B, y) \subseteq P(A, x) \subseteq A$. This forces $s < z := \min (A \setminus P(B, y))$, so that $s \in P(A, z)$.
  Together, we find that $P(B, y) \subseteq P(A, z) \subseteq P(B, y)$, which proves the subclaim.

Returning to the proof of the claim, we now have that $z = f(P(A, z)) = f(P(B, y)) = y$. But $y \in B$, so $y \neq x$. Hence $z < x$. Thus $y = z \in P(A, x)$, contradicting the definition of y. We deduce that $P(A, x) = B$, and hence that B is an initial segment of A, thereby proving our claim.

Suppose that $A \subseteq X$ has property L, and let $x \in A$. It follows from the above argument that given $y < x$, either $y \in A$ or y does not belong to any set B with property L.

Let $V = \cup \{A \subseteq X : A \text{ has property } L\}$.

• **Claim 2:** *We claim that if $w = f(V)$, then $V \cup \{w\}$ has property L.*

Suppose that we can show this. Then $V \cup \{w\} \subseteq V$, so $w \in V$, a contradiction. This will complete the proof.

• **Subclaim 2a:** *First we show that V itself has property L.* We must show that V is well-ordered, and that for all $x \in V$, $x = f(P(V, x))$.

(a) V is well-ordered.
  Let $\varnothing \neq B \subseteq V$. Then there exists $A_0 \subseteq X$ so that A_0 has property L and $B \cap A_0 \neq \varnothing$. Since A_0 is well-ordered and $\varnothing \neq B \cap A_0 \subseteq A_0$, $m := \min(B \cap A_0)$ exists. We claim that $m = \min(B)$.
  Suppose that $y \in B$. Then there exists $A_1 \subseteq X$ so that A_1 has property L and $y \in A_1$. Now, both A_0 and A_1 have property L:

◇ if $A_0 = A_1$, then $m = \min(B \cap A_1)$, so $m \leq y$.

◇ if $A_0 \neq A_1$, then either

- A_0 is an initial segment of A_1, so $A_0 = P(A_1, d)$ for some $d \in A_1$. Then

$$m = \min(B \cap A_0) = \min(B \cap A_1),$$

since $r \in A_1 \setminus A_0$ implies that $m < d \leq r$. Hence $m \leq y$;, or

- A_1 is an initial segment of A_0, say $A_1 = P(A_0, d) \subsetneq A_0$ for some $d \in A_0$. Then

$$m = \min(B \cap A_0) \leq \min(B \cap A_1).$$

Hence $m \leq y$.

In both cases we see that $m \leq y$. Since $y \in B$ was arbitrary, $m = \min(B)$. Thus, any non-empty subset B of V has a minimum element, and so V is well-ordered.

(b) Let $x \in V$. Then there exists $A_2 \subseteq X$ with property L so that $x \in A_2$. Then $x = P(A_2, x)$. Suppose that $y \in V$ and $y < x$. Then there exists $A_3 \subseteq X$ with property L so that $y \in A_3$. Since A_2 and A_3 both have property L, either

- $A_2 = A_3$, and so $y \in A_2$; or
- $A_2 = P(A_3, d)$ for some $d \in A_3$. Since $x \in A_2$, $P(A_2, x) = P(A_3, x)$ and therefore $y \in A_2$; or
- $A_3 = P(A_2, d)$ for some $d \in A_2$. Then $y \in A_3$ implies that $y \in A_2$.

In any of these three cases, $y \in A_2$. Hence $P(V, x) \subseteq P(A_2, x)$. Since $A_2 \subseteq V$, we have that $P(A_2, x) \subseteq P(V, x)$, whence $P(A_2, x) = P(V, x)$. But then

$$x = f(P(A_2, x)) = f(P(V, x)).$$

By (a) and (b), V has property L.

We now return to the proof of Claim 2. That is, we prove that if $w = f(V)$, then $V \cup \{w\}$ has property L.

(I) $V \cup \{w\}$ is well-ordered.
We know that V is well-ordered by part (a) above. Suppose that $\varnothing \neq B \subseteq V \cup \{w\}$. If $B \cap V \neq \varnothing$, then by (a) above, $m := \min(B \cap V)$ exists. Clearly $m \in V$ implies $m \leq f(V) = w$, so $m = \min(B \cap (V \cup \{w\}))$.
If $\varnothing \neq B \subseteq V \cup \{w\}$ and $B \cap V = \varnothing$, then $B = \{w\}$, and so $w = \min(B)$ exists. Hence $V \cup \{w\}$ is well-ordered.

(II) Let $x = V \cup \{w\}$. If $x \in V$, then $x = f(P(V, x))$ by part (a). If $x = w$, then

$$P(V \cup \{w\}, x) = P(V \cup \{w\}, w) = V,$$

so $x = w = f(V) = f(P(V \cup \{w\}, x))$.

By (I) and (II), $V \cup \{w\}$ has property L. As we saw in the statement following Claim 2, this completes the proof that the Axiom of Choice implies Zorn's Lemma. Now let us never speak of this again.

(ii) implies (iii): Let $X \neq \varnothing$ be a set. It is clear that every finite subset $F \subseteq X$ can be well-ordered. Let $\mathcal{A}$ denote the collection of pairs $(Y, \leq_Y)$, where $Y \subseteq X$ and $\leq_Y$ is a well-ordering of Y. For $(A, \leq_A)$, $(B, \leq_B) \in \mathcal{A}$, observe that A is an **initial segment** of B if the following two conditions are met:

- $A \subseteq B$ and $a_1 \leq_A a_2$ implies that $a_1 \leq_B a_2$;
- if $b \in B \setminus A$, then $a \leq_B b$ for all $a \in A$.

Let us partially order $\mathcal{A}$ by setting $(A, \leq_A) \leq (B, \leq_B)$ if A is an initial segment of B. Let $\mathcal{C} = \{C_\lambda\}_{\lambda \in \Lambda}$ be a chain in $\mathcal{A}$.

Then (**exercise**): $\cup_{\lambda \in \Lambda} C_\lambda$ is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, $\mathcal{A}$ admits a maximal element, say $(M, \leq_M)$. We claim that $M = X$. Suppose otherwise. Then we can choose $x_0 \in X \setminus M$ and set $M_0 = M \cup \{x_0\}$. define a partial order on M_0 via: $x \leq_{M_0} y$ if either (a) $x, y \in M$ and $x \leq_M y$, or (b) x is arbitrary and $y = x_0$. Then $(M_0, \leq_{M_0})$ is a well-ordered set and $(M, \leq_M) < (M_0, \leq_{M_0})$, a contradiction of the maximality of $(M, \leq_M)$. Thus $M = X$ and $\leq_M$ is a well- ordering of X.

(iii) implies (i): This is a homework exercise!!

$\square$

The Cantor set has Lebesgue measure zero

L.W. Marcoux

December 3, 2020

1. Lebesgue outer measure

1.1. Our goal in this section is to define a "measure of length" for as many subsets of $\mathbb{R}$ as possible. We would like our new notion to agree with our intuition in the cases we know; for example, it seems reasonable to ask that our generalized notion of "length" of a finite interval (a, b) should be $(b-a)$ when $a < b$ in $\mathbb{R}$. We shall therefore use this as our starting point, and we shall use this intuition to extend our notion of "length" to a greater variety of sets by approximation.

For $a \le b \in \mathbb{R}$, we define the **length** of (a, b) to be $b - a$, and we write

$$\ell((a,b)) := b - a.$$

We also set $\ell(\varnothing) = 0$, and $\ell((-\infty, b)) = \ell((a, \infty)) = \ell((-\infty, \infty)) = \infty$ for all $a, b \in \mathbb{R}$. In this way we have defined $\ell(I)$ whenever I is an open interval in $\mathbb{R}$.

1.2. Definition. *Let $E \subseteq \mathbb{R}$. A countable collection $\{I_n\}_{n=1}^\infty$ of open intervals is said to be a* ***cover of by open intervals*** *E if $E \subseteq \cup_{n=1}^\infty I_n$.*

For each subset E of $\mathbb{R}$, we define a quantity $m^*E \in [0, \infty] := [0, \infty) \cup \{\infty\}$ as follows:

$$m^*E := \inf\{\sum_{n=1}^\infty \ell(I_n) : \{I_n\}_{n=1}^\infty \text{ a cover of } E \text{ by open intervals}\}.$$

Recall that for any set X, we denote by $\mathfrak{P}(X) = \{Y : Y \subseteq X\}$ the **power set** of X.

1.3. Definition. *Let $\varnothing \neq X$ be a set. An* ***outer measure*** *μ on X is a function*

$$\mu : \mathfrak{P}(X) \to [0, \infty]$$

which satisfies

(a) *$\mu\varnothing = 0$;*
(b) *if $E \subseteq F \subseteq X$, then $\mu E \le \mu F$. We say that μ is* ***monotone increasing****; and*
(c) *if $F_n \subseteq X$ for all $n \ge 1$, then*

$$\mu(\cup_{n=1}^\infty F_n) \le \sum_{n=1}^\infty \mu(F_n).$$

It is worth noting that by virtue of (b), condition (c) is equivalent to condition:

(d) if $E, F_1, F_2, F_3, \ldots \subseteq X$ and if $E \subseteq \cup_{n=1}^\infty F_n$, then

$$\mu E \le \sum_{n=1}^\infty \mu F_n.$$

Condition (c) (or equivalently condition (d)) is generally referred to as the **countable subadditivity** or **σ-subadditivity** of μ.

1.4. Proposition. *The function m^* defined in Definition 1.2 is an outer measure on $\mathbb{R}$.*
Proof.

(a) Let $E = \varnothing$. With $I_n = \varnothing$, $n \ge 1$, it is clear that $\{I_n\}_{n=1}^\infty$ is a cover of E by open intervals, and so

$$0 \le m^*\varnothing \le \sum_{n=1}^\infty \ell(I_n) = \sum_{n=1}^\infty 0 = 0.$$

Thus $m^*\varnothing = 0$.

(b) Let $E \subseteq F \subseteq \mathbb{R}$.

If $\{I_n\}_{n=1}^\infty$ is a cover of F by open intervals, then it is also a cover of E by open intervals. It follows immediately from the definition that

$$m^*E \le m^*F.$$

(c) Suppose that $\{E_n\}_{n=1}^\infty$ is a countable collection of subsets of $\mathbb{R}$. We wish to prove that $m^*(\cup_{n=1}^\infty E_n) \le \sum_{n=1}^\infty m^*E_n$.

If $\sum_{n=1}^\infty m^*E_n = \infty$, then we have nothing to prove. Thus we consider the case where $\sum_{n=1}^\infty m^*E_n < \infty$. Set $E := \cup_{n=1}^\infty E_n$.

Let $\varepsilon > 0$ and for each $n \ge 1$, choose a cover $\{I_k^{(n)}\}_{k=1}^\infty$ of E_n by open intervals such that

$$\sum_{k=1}^\infty \ell(I_k^{(n)}) < m^*E_n + \frac{\varepsilon}{2^n}.$$

Then $\{I_k^{(n)}\}_{k,n=1}^\infty$ is a cover of E by open intervals, and so

$$\begin{aligned} m^*E &\le \sum_{n=1}^\infty \sum_{k=1}^\infty \ell(I_k^{(n)}) \\ &\le \sum_{n=1}^\infty (m^*E_n + \frac{\varepsilon}{2^n}) \\ &= \left(\sum_{n=1}^\infty m^*E_n\right) + \varepsilon. \end{aligned}$$

(Here we have used the fact that if $0 \le a_n \in \mathbb{R}$, $n \ge 1$, and if $\sigma : \mathbb{N} \to \mathbb{N}$ is any permutation, then $\sum_{n=1}^\infty a_n = \sum_{n=1}^\infty a_{\sigma(n)}$.)

Since $\varepsilon > 0$ was arbitrary,

$$m^*E \le \sum_{n=1}^\infty m^*E_n.$$

□

1.5. Corollary. *Let $E \subseteq \mathbb{R}$ be a countable set. Then $m^*E = 0$.*
Proof. Suppose that E is denumerable, say $E = \{x_n\}_{n=1}^\infty$.

Let $\varepsilon > 0$ and for each $n \ge 1$, set $I_n = (x_n - \frac{\varepsilon}{2^{n+1}}, x_n + \frac{\varepsilon}{2^{n+1}})$. Then $\{I_n\}_{n=1}^\infty$ is a cover of E by open intervals, and therefore

$$0 \le m^*E \le \sum_{n=1}^\infty \ell(I_n) = \sum_{n=1}^\infty \frac{\varepsilon}{2^n} = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we have $m^*E = 0$.

The case where E is finite is left as an exercise.

□

1.6. Corollary. *The outer measure of the rational numbers is* 0*; i.e.* $m^*\mathbb{Q} = 0$.

We have defined outer measure m^*E for *any* subset E of $\mathbb{R}$, and we have done this based upon an intuitive notion of what the length of an open interval (a,b) should be, namely $b-a$. At first glance, it seems obvious that $m^*(a,b) = \ell(b-a) = b-a$. But upon reflection, we see that this is *not* how $m^*(a,b)$ is defined. This leaves us with an interesting problem: how does our notion of measure $m^*(a,b)$ of an interval compare with this notion of length? On the one hand, it is clear that $m^*(a,b) \le \ell(b-a) = b-a$, since $I_1 := (a,b)$ and $I_n = \varnothing$, $n \ge 2$ yields a cover of (a,b) by open intervals. On the other hand, the notion of outer measure of (a,b) requires us to consider *all* covers of (a,b) by intervals, not only the obvious cover by the interval (a,b) itself. We now turn to this problem. It will prove useful to first consider the outer measure of closed, bounded intervals $[a,b]$, as these are "*compact*".

1.7. Ah, but here's the rub M-147. What do we mean when we say that a subset of $\mathbb{R}$ is *compact*? Typically, the notion of compactness that is introduced in Math 147 is that of **sequential compactness**: that is, $\varnothing \ne K \subseteq \mathbb{R}$ is said to be (**sequentially**) **compact** if, given any sequence $(x_n)_n$ in K, there exists a subsequence $(x_{n_k})_k$ of $(x_n)_n$ and a point $y \in K$ such that $\lim_k x_{n_k} = y$. One then proves that K is compact if and only if K is closed and bounded. This last result is often referred to in polite society as the **Heine-Borel Theorem**.

There is another, more "generalisable" notion of compactness which will be useful to us. It requires a notion of covers similar to the notion of covers by open intervals defined above, except that we no longer require the open sets to be intervals.

1.8. Definition. *Let* $K \subseteq \mathbb{R}$ *be a set. An* ***open cover*** *of* K *is a collection* $\{G_\lambda\}_{\lambda\in\Lambda}$ *of open sets such that*

$$K \subseteq \cup_{\lambda\in\Lambda} G_\lambda.$$

We say that an open cover $\{G_\lambda\}_{\lambda\in\Lambda}$ *of* K ***admits a finite subcover*** *if there exists a finite subset* $\{\lambda_1, \lambda_2, \dots, \lambda_N\} \subseteq \Lambda$ *for which*

$$K \subseteq \cup_{n=1}^N G_{\lambda_n}.$$

Finally, $K \subseteq \mathbb{R}$ *is said to be* ***(topologically) compact*** *if every open cover of* K *admits a finite subcover.*

1.9. It is worth observing that – unlike the definition of a *cover of* E *by open intervals* in Definition 1.2 – the definition of a *cover* appearing in the definition of *compactness* does not require the set Λ to be countable. It is also worth observing that this definition extends easily to topological spaces. That is, if (X,τ) is a topological space and if $K \subseteq X$, we may define an open cover of K exactly as above. It would take a while to prove that a subset of $\mathbb{R}$ is compact if and only if it is sequentially compact, and thus if and only if it is closed and bounded. But the argument showing that any closed interval $[a,b]$ with $a < b$ in $\mathbb{R}$ is compact is short and sweet enough to include here, especially since we shall require it afterwards!

1.10. Proposition. *Let* $a < b \in \mathbb{R}$*. Then* $[a,b] \subseteq \mathbb{R}$ *is compact.*

Proof. Let's do something totally unexpected. Let's begin with the definition, shall we? Let $\mathcal{G} := \{G_\lambda\}_{\lambda\in\Lambda}$ be an open cover of $[a,b]$. We must show that it admits a finite subcover. We argue by contradiction. Suppose, therefore, that $\mathcal{G}$ does not admit a finite subcover of $[a,b]$.

Set $a_1 := a$, $b_1 = b$. Now, by definition, $[a_1, b_1] \subseteq \cup_{\lambda\in\Lambda} G_\lambda$. Let $c_1 = \frac{a_1+b_1}{2}$ be the midpoint of $[a_1, b_1]$. Then $[a_1, b_1] = [a_1, c_1] \cup [c_1, b_1]$. Since $\mathcal{G}$ does not admit a finite subcover of $[a_1, b_1]$, either $[a_1, c_1]$ or $[c_1, b_1]$ cannot be covered by finitely many sets in $\mathcal{G}$. If $[a_1, c_1]$ cannot be covered by finitely many sets in $\mathcal{G}$, we set $a_2 = a_1$ and $b_2 = c_1$. Otherwise, let $a_2 = c_1$ and $b_2 = b_1$. Either way,

- $[a_2, b_2] \subseteq [a_1, b_1]$;
- $b_2 - a_2 = \frac{b_2 - a_2}{2}$; and
- $[a_2, b_2]$ cannot be covered by finitely many sets in $\mathcal{G}$, although $[a_2, b_2] \subseteq \cup_{\lambda\in\Lambda} G_\lambda$.

The idea is to now apply the above argument to the interval $[a_2, b_2]$ to obtain a new interval $[a_3, b_3]$ (which is either the left half or the right half of $[a_2, b_2]$) such that $\mathcal{G}$ covers $[a_3, b_3]$, but $\mathcal{G}$ does not admit a finite subcover of $[a_3, b_3]$.

More generally, we continue this process to obtain intervals

$$\cdots [a_n, b_n] \subseteq [a_{n-1}, b_{n-1}] \subseteq \cdots \subseteq [a_2, b_2] \subseteq [a_1, b_1]$$

such that $\mathcal{G}$ is a cover of $[a_n, b_n]$ but does not admit a finite subcover of $[a_n, b_n]$, and

$$b_n - a_n = \frac{b_{n-1} - a_{n-1}}{2} = \frac{b_1 - a_1}{2^{n-1}}.$$

By the Nested Intervals Theorem, there exists a real number x_0 such that

$$\cap_{n=1}^\infty [a_n, b_n] = \{x_0\}.$$

Now $x_0 \in [a_1, b_1] \subseteq \cup_{\lambda\in\Lambda} G_\lambda$, so there exists $\lambda_0 \in \Lambda$ such that $x_0 \in G_{\lambda_0}$. Since G_{λ_0} is open, there exists $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq G_{\lambda_0}$.

Next, fix $N \in \mathbb{N}$ sufficiently large such that $b_N - a_N < \delta$. Since $a_N \leq x_0 \leq b_N$ (because $x_0 \in [a_N, b_N]$), it follows that $x_0 - \delta < a_N < x_0 < b_N < x_0 + \delta$, and thus

$$[a_N, b_N] \subseteq (x_0 - \delta, x_0 + \delta) \subseteq G_{\lambda_0},$$

contradicting the fact that $\mathcal{G}$ does not admit a finite subcover of $[a_N, b_N]$.

This contradiction implies that $\mathcal{G}$ must admit a finite subcover of $[a, b]$. This is all we shall require below.

□

1.11. Proposition. *Let $a < b \in \mathbb{R}$. Then*

(a) $m^*[a, b] = b - a$, *and therefore*

(b) $m^*(a, b] = m^*[a, b) = m^*(a, b) = b - a$.

Proof.

(a) Let $\varepsilon > 0$ and note that with $I_1 = (a - \frac{\varepsilon}{2}, b + \frac{\varepsilon}{2})$ and $I_n = \varnothing$, $n \geq 2$, the collection $\{I_n\}_{n=1}^\infty$ is a cover of $[a, b]$ by open intervals and thus

$$m^*[a, b] \leq \sum_{n=1}^\infty \ell(I_n) = \ell(I_1) = (b - a) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, $m^*[a, b] \leq b - a$.

We now turn the question of obtaining a lower bound for $m^*[a, b]$.

Suppose that $\{I_n\}_{n=1}^\infty$ is an cover of $[a, b]$ by open intervals. (Note: without loss of generality, we may assume that $I_n \neq \varnothing$, $1 \leq n < \infty$.) We must show that $\sum_{n=1}^\infty \ell(I_n) \geq b - a$.

Since $[a,b]$ is compact, we can find a finite subcover $\{I_1, I_2, \dots, I_N\}$ of $[a,b]$. If $\ell(I_n) = \infty$ for some $1 \le n \le N$, then the inequality

$$\sum_{n=1}^{\infty} \ell(I_n) \ge \sum_{n=1}^{N} \ell(I_n) \ge b - a$$

trivially holds. Thus we shall assume that $\ell(I_n) < \infty$ for all $1 \le n \le N$, and we may then write $I_n = (a_n, b_n)$, $1 \le n \le N$.

Since $a \in [a,b] \subseteq \cup_{n=1}^{N} I_n$, we can find $1 \le n_1 \le N$ such that $a \in I_{n_1} = (a_{n_1}, b_{n_1})$. If $b_{n_1} > b$, we stop.

Otherwise, $a < b_{n_1} \le b$, so $b_{n_1} \in [a,b] \subseteq \cup_{n=1}^{N} I_n$, and we can find $n_2 \in \{1, 2, \dots, N\} \setminus \{n_1\}$ such that $b_{n_1} \in (a_{n_2}, b_{n_2})$. If $b_{n_2} > b$, we stop.

Otherwise, $a_{n_1} < a < b_{n_1} < b_{n_2} \le b$, so $b_{n_2} \in [a,b] \subseteq \cup_{n=1}^{N} I_n$, and we can find $n_3 \in \{1, 2, \dots, N\} \setminus \{n_1, n_2\}$ such that $b_{n_2} \in (a_{n_3}, b_{n_3})$. If $b_{n_3} > b$, we stop.

Eventually this process must end, since we have only $N < \infty$ intervals, and each stage n_k is chosen from among $\{1, 2, \dots, N\} \setminus \{n_1, n_2, \dots, n_{k-1}\}$. Suppose therefore that $1 \le M \le N$ is the minimal integer such that $b_{n_M} > b$. Then

$$[a,b] \subseteq \cup_{k=1}^{M} (a_{n_k}, b_{n_k}).$$

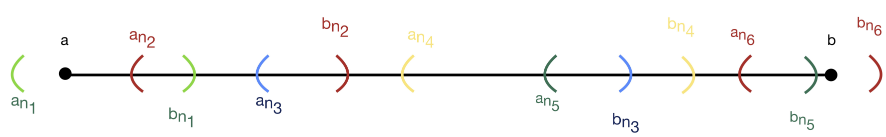

FIGURE 1. AN EXAMPLE WHERE $M = 6$.

Now,

$$\begin{aligned}
\sum_{n=1}^{\infty} \ell(I_n) &\ge \sum_{n=1}^{N} \ell(I_n) \\
&\ge \sum_{k=1}^{M} \ell((a_{n_k}, b_{n_k})) \\
&= (b_{n_1} - a_{n_1}) + (b_{n_2} - a_{n_2}) + \cdots + (b_{n_M} - a_{n_M}) \\
&= b_{n_M} + (b_{n_{M-1}} - a_{n_M}) + (b_{n_{M-2}} - a_{n_{M-1}}) + \cdots + (b_{n_1} - a_{n_2}) - a_{n_1} \\
&> b_{n_M} - a_{n_1} \\
&> b - a.
\end{aligned}$$

Since $\{I_n\}_{n=1}^{\infty}$ was an arbitrary cover of $[a,b]$, it follows that

$$m^*[a,b] \ge b - a.$$

Combining this with the reverse inequality above, we conclude that

$$m^*[a,b] = b - a.$$

(b) Consider the interval $(a,b]$.

For all $0 < \varepsilon < b - a$, we have that $[a + \varepsilon, b] \subseteq (a,b] \subseteq [a,b]$, and thus monotonicity of Lebesgue outer measure implies that

$$(b-a) - \varepsilon = m^*[a+\varepsilon, b] \le m^*(a,b] \le m^*[a,b] = b - a.$$

Since ε was arbitrary (subject to $0 < \varepsilon < b - a$), we conclude that $m^*(a,b] = b - a$.

The remaining cases are similar, and are left as an exercise for the reader.

□

1.12. Corollary. *Let $a, b \in \mathbb{R}$. Then*

$$m^*(-\infty, b) = m^*(-\infty, b] = m^*(a, \infty) = m^*[a, \infty) = m^*\mathbb{R} = \infty.$$

Proof. Consider the interval $(-\infty, b)$.

By monotonicity of Lebesgue outer measure, for each $n \geq 1$,

$$n = m^*[b - n, b) \leq m^*(-\infty, b),$$

and thus $m^*(-\infty, b) = \infty$.

The remaining cases are similar, and are left as an exercise for the reader.

□

Here is a completely ridiculous proof of something we already know.

1.13. Corollary. *Let $a < b \in \mathbb{R}$. Then $[a, b]$ is uncountable.*

Proof. This follows immediately from Proposition 1.11 and Corollary 1.5.

□

2. Lebesgue measure and the Cantor set

2.1. Although we have defined Lebesgue outer measure m^* for an arbitrary subset $E \subseteq \mathbb{R}$, the function $m^* : \mathfrak{P}(\mathbb{R}) \to [0, \infty]$ behaves poorly in some ways. (More precisely, it fails to be "translation-invariant and σ-additive", whatever that means, and it just so happens to mean something good. Ok, *translation-invariant* means that if you translate a set to the left or to the right, it doesn't change its outer measure, and σ*-additive* means that if a set E is the disjoint union of countably many *disjoint* sets $\{E_n\}_n$, then the outer measure of E should just be the sum of the outer measures of the E_n's.) Most people agree that the problem is not with our definition of the function m^*, but with our unrealistic expectations that we can find the measure of an arbitrary subset of $\mathbb{R}$.

We begin with Carathéodory's definition of a Lebesgue measurable set, since it is the most practical definition to use. Those of us who follow the serpentine path of pure mathematics all the way to PM450 will see that Lebesgue measurable sets are "almost" countable intersections of open sets, in a way which we shall make precise at that point.

2.2. Definition. *A set E is $\mathbb{R}$ is said to be* ***Lebesgue measurable*** *if, for all $X \subseteq \mathbb{R}$,*

$$m^*X = m^*(X \cap E) + m^*(X \setminus E).$$

We denote by $\mathfrak{M}(\mathbb{R})$ the collection of all Lebesgue measurable sets.

2.3. Remarks. Since our attention here is almost exclusively focused upon Lebesgue measure, we shall allow ourselves to drop the adjective "Lebesgue" and refer only to "measurable sets".

Informally speaking, we see that a set $E \subseteq \mathbb{R}$ is measurable provided that it is a "*universal slicer*", in the sense that it "slices" every other set X into two *disjoint* sets, namely $X \cap E$ and $X \setminus E$, where Lebesgue outer measure *is additive*!

We also note that the inequality

$$m^* X \le m^*(X \cap E) + m^*(X \setminus E)$$

is free from the σ-subadditivity of Lebesgue outer measure. In checking to see whether a given set is measurable or not, it therefore suffices to verify that the reverse inequality holds for all sets $X \subseteq \mathbb{R}$.

Before we proceed to the examples, which shall obtain a result which allows us to show that the set $\mathfrak{M}(\mathbb{R})$ of Lebesgue measurable sets itself has an interesting structure.

2.4. Definition. *A collection $\Omega \subseteq \mathfrak{P}(\mathbb{R})$ is said to be an* ***algebra of sets*** *if*

(a) $\mathbb{R} \in \Omega$;
(b) $E \in \Omega$ *implies that* $E^c := \mathbb{R} \setminus E \in \Omega$; *and*
(c) *if* $N \ge 1$ *and* $E_1, E_2, \dots, E_N \in \Omega$, *then*

$$E := \cup_{n=1}^N E_n \in \Omega.$$

We say that Ω is a ***σ-algebra of sets*** *if Ω is an algebra of sets which satisfies the additional property:*

(d) *if* $F_n \in \Omega$ *for all* $n \ge 1$, *then*

$$F := \cup_{n=1}^\infty F_n \in \Omega.$$

Informally, we often say that Ω is a σ-algebra.

2.5. Theorem. *The collection $\mathfrak{M}(\mathbb{R})$ of Lebesgue measurable sets in $\mathbb{R}$ is a σ-algebra of sets.*

Proof.

(a) Let us first verify that $\mathbb{R} \in \mathfrak{M}(\mathbb{R})$.

If $X \subseteq \mathbb{R}$, then $X \cap \mathbb{R} = X$, while $X \cap \mathbb{R}^c = X \cap \varnothing = \varnothing$. Thus

$$m^*(X \cap \mathbb{R}) + m^*(X \cap \mathbb{R}^c) = m^* X + m^* \varnothing = m^* X + 0 = m^* X.$$

By definition, $\mathbb{R} \in \mathfrak{M}(\mathbb{R})$.

(b) The fact that $\mathfrak{M}(\mathbb{R})$ is closed under complementation is clear, since the definition of a measurable set is symmetric in E and $\mathbb{R} \setminus E$.

(c) Let $\{E_n\}_{n=1}^\infty \subseteq \mathfrak{M}(\mathbb{R})$. We must show that $E := \cup_{n=1}^\infty E_n \in \mathfrak{M}(\mathbb{R})$.

Step 1. For each $N \ge 1$, consider $H_N := \cup_{n=1}^N E_n$. We shall argue by induction that $H_N \in \mathfrak{M}(\mathbb{R})$ for all $N \ge 1$.

For $N = 1$, this is trivially true, as $H_1 = E_1 \in \mathfrak{M}(\mathbb{R})$ by hypothesis.

Next suppose that $1 \le M \in \mathbb{N}$ and that $H_M \in \mathfrak{M}(\mathbb{R})$. Let $X \subseteq \mathbb{R}$ be arbitrary. The induction hypothesis says that

$$m^* X = m^*(X \cap H_M) + m^*(X \setminus H_M).$$

But $E_{M+1} \in \mathfrak{M}(\mathbb{R})$, and so

$$m^*(X \setminus H_M) = m^*((X \setminus H_M) \cap E_{M+1}) + M^*((X \setminus H_M) \setminus E_{M+1}).$$

By the subadditivity of m^*,

$$\begin{aligned} m^* X &= m^*(X \cap H_M) + m^*(X \setminus H_M) \\ &= m^*(X \cap H_M) + m^*((X \setminus H_M) \cap E_{M+1})) + m^*((X \setminus H_M) \setminus E_{M+1}) \\ &\geq m^*(X \cap (H_M \cup E_{M+1})) + m^*(X \setminus (H_M \cup E_{M+1})) \\ &= m^*(X \cap H_{M+1}) + m^*(X \setminus H_{M+1}) \end{aligned}$$

Since the reverse inequality holds for any outer measure, we see that

$$m^* X = m^*(X \cap H_{M+1}) + m^*(X \setminus H_{M+1}),$$

and thus $H_{M+1} \in \mathfrak{M}(\mathbb{R})$.

By induction, $H_N \in \mathfrak{M}(\mathbb{R})$ for all $N \geq 1$.

Step 2. Next we shall write each H_N as a *disjoint* union of sets in $\mathfrak{M}(\mathbb{R})$.

Let $F_1 := H_1$, and for $n \geq 2$, set $F_n := H_n \setminus H_{n-1}$. Clearly $F_i \cap F_j = \varnothing$ for $1 \leq i \neq j$, and $H_N = \uplus_{n=1}^N F_n$ for all $N \geq 1$.

Let $N \geq 2$ be an integer and note that $H_{N-1}, H_N \in \mathfrak{M}(\mathbb{R})$ implies that $H_{N-1}^c, H_N^c \in \mathfrak{M}(\mathbb{R})$ by (b). By Step 1, $(H_N^c \cup H_{N-1}) \in \mathfrak{M}$, and using (b) once again,

$$F_N = H_N \setminus H_{N-1} = (H_N^c \cap H_{N-1})^c \in \mathfrak{M}(\mathbb{R}).$$

Step 3. Now we claim that if $X \subseteq \mathbb{R}$, then for each $N \geq 1$,

$$m^*(X \cap (\uplus_{n=1}^N F_n)) = \sum_{n=1}^N m^*(X \cap F_n).$$

The claim is trivially true when $N = 1$.

Suppose that $1 \leq M \in \mathbb{N}$ and that

$$m^*(X \cap (\uplus_{n=1}^M F_n)) = \sum_{n=1}^M m^*(X \cap F_n).$$

By the measurability of F_{M+1} and the fact that all F_j's are disjoint,

$$\begin{aligned} m^*(X \cap (\uplus_{n=1}^{M+1} F_n)) &= m^*((X \cap (\uplus_{n=1}^{M+1} F_n)) \setminus F_{M+1}) + \\ &\qquad m^*((X \cap (\uplus_{n=1}^{M+1} F_n)) \cap F_{M+1}) \\ &= m^*(X \cap (\uplus_{n=1}^{M} F_n)) + m^*(X \cap F_{M+1}) \\ &= \sum_{n=1}^M m^*(X \cap F_n) + m^*(X \cap F_{M+1}) \\ &\qquad \text{by the induction hypothesis} \\ &= \sum_{n=1}^{M+1} m^*(X \cap F_n). \end{aligned}$$

This completes the induction step and therefore proves our claim.

Step 4. Finally, observe that $E = \cup_{n=1}^\infty E_n = \cup_{n=1}^\infty H_n = \uplus_{n=1}^\infty F_n$. We shall use this to prove that $E \in \mathfrak{M}(\mathbb{R})$.

Let $X \subseteq \mathbb{R}$. For all $N \geq 1$, $H_N \in \mathfrak{M}(\mathbb{R})$ and so

$$\begin{aligned} m^* X &= m^*(X \cap H_N) + m^*(X \setminus H_N) \\ &= m^*(X \cap (\cup_{n=1}^N F_n)) + m^*(X \setminus H_N) \\ &\geq m^*(X \cap (\cup_{n=1}^N F_n)) + m^*(X \setminus E) && \text{as } (X \setminus E) \subseteq (X \setminus H_N) \\ &= \sum_{n=1}^N m^*(X \cap F_n) + m^*(X \setminus E) && \text{by STEP 3.} \end{aligned}$$

Taking limits, we see that

$$m^* X \geq \sum_{n=1}^\infty m^*(X \cap F_n) + m^*(X \setminus E).$$

Keeping in mind that m^* is σ-subadditive, we note that

$$\begin{aligned} m^*(X \cap E) &= m^*(X \cap (\cup_{n=1}^\infty F_n)) \\ &= m^*(\cup_{n=1}^\infty (X \cap F_n)) \\ &\leq \sum_{n=1}^\infty m^*(X \cap F_n). \end{aligned}$$

Combining these last two estimates, we conclude that

$$m^* X \geq m^*(X \cap E) + m^*(X \setminus E),$$

and therefore that $E \in \mathfrak{M}(\mathbb{R})$.

□

It's high time that we produce examples of Lebesgue measurable sets. Thanks to the previous Theorem, given a subset $\mathfrak{S} \subseteq \mathfrak{M}(\mathbb{R})$, the entire σ-algebra generated by $\mathfrak{S}$ (namely the smallest σ-algebra of subsets of $\mathbb{R}$ which contains $\mathfrak{S}$ – why should this exist?) is also contained in $\mathfrak{M}(\mathbb{R})$.

2.6. Proposition.

(a) *If $E \subseteq \mathbb{R}$ and $m^* E = 0$, then $E \in \mathfrak{M}(\mathbb{R})$.*
(b) *For all $b \in \mathbb{R}$, $E := (-\infty, b) \in \mathfrak{M}(\mathbb{R})$.*
(c) *Every open and every closed set is Lebesgue measurable.*

Proof.

(a) Let $E \subseteq \mathbb{R}$ be a set with $m^* E = 0$, and let $X \subseteq \mathbb{R}$. By monotonicity of outer measure,

$$m^*(X \cap E) \leq m^* E = 0,$$

and

$$m^*(X \cap E^c) \leq m^* X.$$

Thus

$$m^* X = 0 + m^* X \geq m^*(X \cap E) + m^*(X \setminus E).$$

As we have seen, this is the statement that $E \in \mathfrak{M}(\mathbb{R})$.

(b) Fix $b \in \mathbb{R}$ and set $E = (-\infty, b)$. Let $X \subseteq \mathbb{R}$ be arbitrary. We must show that

$$m^* X \geq m^*(X \cap E) + m^*(X \setminus E).$$

If $m^* X = \infty$, then there is nothing to prove, and so we assume that $m^* X < \infty$. Let $\varepsilon > 0$ and let $\{I_n\}_{n=1}^\infty$ be a cover of X by open intervals such that

$$\sum_{n=1}^\infty \ell(I_n) < m^* X + \varepsilon < \infty.$$

It follows that each interval I_n has finite length, and so we may write $I_n = (a_n, b_n)$, $n \geq 1$. (As always, there is no harm in assuming that each $I_n \neq \varnothing$, otherwise we simply remove I_n from the cover.)

Set $J_n = I_n \cap E = (a_n, b_n) \cap (-\infty, b)$. Clearly each J_n, $n \geq 1$ is an open interval, possibly empty.

Let $K_n = I_n \setminus E = (a_n, b_n) \cap [b, \infty)$. Then

$$K_n \in \{\varnothing, [b, b_n), (a_n, b_n)\},$$

depending upon the values of a_n and b_n. But then we can find $c_n < d_n$ in $\mathbb{R}$ such that

$$K_n \subseteq L_n := (c_n, d_n)$$

and $\ell(L_n) - m^* K_n < \frac{\varepsilon}{2^n}$, $n \geq 1$.

In particular, for each $n \geq 1$, $I_n \subseteq K_n \cup L_n$ and

$$(\ell(J_n) + \ell(L_n)) - \ell(I_n) < \frac{\varepsilon}{2^n}.$$

Now

$$\begin{aligned} m^* X &> \left(\sum_{n=1}^\infty \ell(I_n) \right) - \varepsilon \\ &> \left(\sum_{n=1}^\infty (\ell(J_n) + \ell(L_n) - \frac{\varepsilon}{2^n}) \right) - \varepsilon \\ &= \sum_{n=1}^\infty \ell(J_n) + \sum_{n=1}^\infty \ell(L_n) - \sum_{n=1}^\infty \frac{\varepsilon}{2^n} - \varepsilon. \end{aligned}$$

But $X \cap E \subseteq \cup_{n=1}^\infty J_n$ and $X \setminus E \subseteq \cup_{n=1}^\infty L_n$, and so

$$m^* X > m^*(X \cap E) + m^*(X \setminus E) - 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary,

$$m^* X \geq m^*(X \cap E) + m^*(X \setminus E),$$

proving that $E \in \mathfrak{M}(\mathbb{R})$.

(c) Fix $b \in \mathbb{R}$. We have just seen that $(-\infty, b) \in \mathfrak{M}(\mathbb{R})$. Since $\mathfrak{M}(\mathbb{R})$ is an algebra of sets, $E^c = [b, \infty) \in \mathfrak{M}(\mathbb{R})$ as well. But then $E_n := [b + \frac{1}{n}, \infty) \in \mathfrak{M}(\mathbb{R})$ for all $n \geq 1$, and since the latter is a σ-algebra,

$$(b, \infty) = \cup_{n=1}^\infty E_n \in \mathfrak{M}(\mathbb{R}) \qquad \text{for all } b \in \mathbb{R}.$$

If $a < b$, then $(a, b) = (-\infty, b) \cap (a, \infty) \in \mathfrak{M}(\mathbb{R})$. Since $\mathbb{R} \in \mathfrak{M}(\mathbb{R})$ by Theorem 2.5, we see that we have shown that every open interval lies in $\mathfrak{M}(\mathbb{R})$.

We saw in the Assignments that every open set $G \in \mathfrak{G}$ is a countable (disjoint) union of open intervals. Since $\mathfrak{M}$ is a σ-algebra, this means that every open set $G \in \mathfrak{M}$. But $\mathfrak{M}$ is also closed under complementation, and so every closed set lies in $\mathfrak{M}$ as well.

□

Now that we know that $\mathfrak{M}(\mathbb{R}) \neq \varnothing$, the following definition makes sense.

2.7. Definition. *Let m^* denote Lebesgue outer measure on $\mathbb{R}$. We define **Lebesgue measure** m on $\mathbb{R}$ to be the restriction of m^* to $\mathfrak{M}(\mathbb{R})$. That is, Lebesgue measure is the function*

$$\begin{array}{rccc} m: & \mathfrak{M}(\mathbb{R}) & \to & [0,\infty] \\ & E & \mapsto & m^*E. \end{array}$$

2.8. Example. The Cantor middle thirds set. Recall from Corollary 1.5 that if $E \subseteq \mathbb{R}$ is countable, then $m^*E = 0$. By Proposition 2.6, it follows that $E \in \mathfrak{M}(\mathbb{R})$, and thus $mE = m^*E = 0$. In other words, every countable set is Lebesgue measurable with Lebesgue measure zero.

We shall now construct an *uncountable set* C – in fact one whose cardinality is $\mathfrak{c}$, the cardinality of the real line $\mathbb{R}$ – whose measure mC is equal to zero. Since $m\mathbb{R} = \infty$, we see that the Lebesgue measure of a set is not so much a reflection of its cardinality, as much as a question of *how* the points in the set are distributed. Having said that, when the set in question is countable, the above argument shows that it is always "thinly distributed", in this analogy.

The **Cantor set** is typically obtained as the intersection of a countable family of sets, each iteratively constructed from the previous as follows:

We set $C_0 = [0,1]$, and for $n \geq 1$, we set $C_n = \frac{1}{3}C_{n-1} \cup (\frac{2}{3} + \frac{1}{3}C_{n-1})$. Thus

- $C_1 = [0,\frac{1}{3}] \cup [\frac{2}{3},1]$;
- $C_2 = [0,\frac{1}{9}] \cup [\frac{2}{9},\frac{1}{3}] \cup [\frac{6}{9},\frac{7}{9}] \cup [\frac{8}{9},1]$,
- $C_3 = [0,\frac{1}{27}] \cup [\frac{2}{27},\frac{3}{27}] \cup [\frac{6}{27},\frac{7}{27}] \cup [\frac{8}{27},\frac{9}{27}] \cup [\frac{18}{27},\frac{19}{27}] \cup [\frac{20}{27},\frac{21}{27}] \cup [\frac{24}{27},\frac{25}{27}] \cup [\frac{26}{27},1]$;
- $\vdots$

The figure below shows each of the sets C_n, $0 \leq n \leq 7$.

FIGURE 2. AN ILLUSTRATION FROM HTTP://MATHFORUM.ORG/MATHIMAGES/INDEX.PHP/CANTOR_SET.

The **Cantor middle thirds set** is defined as the intersection of all of these sets, i.e.

$$C := \cap_{n=0}^{\infty} C_n.$$

Alternatively, beginning with $C_0 = [0,1]$, one can think of obtaining C_1 from C_0 by removing the (open) "middle third" interval $(\frac{1}{3},\frac{2}{3})$, resulting in the two intervals which comprise C_1 above. To obtain C_2 from C_1, one removes the (open) "middle third" of each of the two intervals in C_1, and so on. This motivates the term *middle thirds* in the above nomenclature.

It should be clear from the construction above that

(a) $C_0 \supseteq C_1 \supseteq C_2 \supseteq C_3 \supseteq \cdots \supseteq C$. Furthermore, each set C_n is closed, $n \geq 0$ and $mC_0 = 1 < \infty$. It is left as an exercise for the reader to show that

$$m^*C = \lim_{n\to\infty} mC_n = \lim_{n\to\infty} \frac{2^n}{3^n} = 0.$$

(b) Being closed and bounded, C is compact – hence measurable with $mC = 0$. Also, C_0 is compact and the collection $\{C_n\}_{n=0}^\infty$ clearly has the Finite Intersection Property. Thus $C = \cap_{n=0}^\infty C_n \neq \varnothing$. (We shall in fact show that C is uncountable!)

Now let us approach things from a different angle. Given $x \in [0,1]$, consider the **ternary expansion** of x, namely

$$x = 0.x_1\, x_2\, x_3\, x_4 \ldots$$

where $x_k \in \{0,1,2\}$ for all $k \geq 1$. As with decimal expansions, the expression above is meant to express the fact that

$$x = \sum_{k=1}^\infty \frac{x_k}{3^k}.$$

Non-uniqueness of this expansion is a problem here, as it is with decimal expansions. For example,

$$\frac{1}{3} = 0.0222222\cdot = 0.1000000\cdots.$$

We leave it as an exercise for the reader to show that the expansion of $x \in [0,1)$ *is* unique *except when* there exists $N \geq 1$ such that

$$x = \frac{r}{3^N} \quad \text{for some} \quad 0 < r < 3^N, \text{ where } 3 \nmid r.$$

When this is the case, we have that

$$x = 0.x_1 x_2\, x_3\, x_4 \cdots x_N,$$

where $x_N \in \{1,2\}$.

If $x_N = 2$, we shall use that expression.

If $x_N = 1$, then

$$\begin{aligned} x &= 0.x_1\, x_2\, x_3 \cdots x_{N-2}\, x_{N-1}\, 1000\cdots \\ &= 0.x_1\, x_2\, x_3 \cdots x_{N-2}\, 02222\cdots, \end{aligned}$$

and we shall agree to adopt the *second* expression.

Finally, we shall use the convention that

$$1 = 0.2222\cdots.$$

With this convention, over $x \in [0,1]$ has a *unique* ternary expansion.

Now

- $x \in C_1$ if and only if $x_1 \neq 1$;
- $x \in C_2$ if and only if $x \in C_1 \cap C_2$, i.e. if and only if $x_1 \neq 1 \neq x_2$.
- $x \in C_3$ if and only if $x \in C_2 \cap C_3$, i.e., if and only if $1 \notin \{x_1, x_2, x_3\}$.

More generally, for $N \geq 1$, $x \in C_N$ if and only $1 \notin \{x_1, x_2, \ldots, x_N\}$.

From this it follows that $x \in C$ if and only if $x_n \neq 1$, $n \geq 1$. In other words,

$$C = \{x = 0.x_1\, x_2\, x_3\, x_4 \ldots : x_n \in \{0,2\} \text{ for all } n \geq 1\}.$$

We can therefore think of an element $x = 0.x_1x_2x_3\dots$ of C as a function $f_x : \mathbb{N} \to \{0,2\}$, where $x_n := f_x(n)$, $n \geq 1$. Hence

$$|C| = |\{0,2\}^{\mathbb{N}}| = |\{0,2\}|^{|\mathbb{N}|} = 2^{\aleph_0} = \mathfrak{c},$$

the cardinality of $\mathbb{R}$.

Thus C is an uncountable, measurable set whose Lebesgue measure is nonetheless equal to 0.

MATH 147 - SEPTEMBER 2020
NORMS AND METRICS
L.W. MARCOUX

1. INTRODUCTION.

1.1. This note will assume that you have seen the notion of a **vector space** before, although the definition of a vector space over the real numbers (and over the complex numbers) is given in the next section. Even if you haven't seen vector spaces before, this discussion will include examples such as $\mathbb{R}^2$ (the "$x, y-$plane") and $\mathbb{R}^3$ ("three space"), with which you are undoubtedly familiar.

1.2. One of the concepts that appears in the course notes for Math 147 are that of *limits of sequences*. This is *absolutely crucial* to an understanding of calculus, and what is more, it is crucial to an understanding of an area of mathematics known as *Analysis*. In this note, we shall try to (at least partially) distill the essence of this concept, and show how we might be able to generalise this notion. (As it so happens, we can go even further than what we do in this note, but one must learn to crawl before one can walk. Unless one is a giraffe. My understanding is that they don't ever crawl – they just walk. From day one. That kind of thing just blows me away. I'm not sure how much giraffes know about limits, norms and metric spaces, but walking? Walking they just seem to "get".)

Let us remind ourselves of the notion of a limit of a sequence of real numbers.

1.3. Definition. *A sequence $(x_n)_{n=1}^\infty$ of real numbers is said to **converge** to the real number x if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $n \geq N$ implies that*

$$|x_n - x| < \varepsilon.$$

1.4. What interests us right now is the very last part of this definition; this is, the part that goes:

$$|x_n - x| < \varepsilon.$$

It involves absolute values, and we've seen them before. We've come to know and love them the same way we our significant others, if not moreso. We certainly understand them better than we understand our significant others, but that's another matter altogether and best not addressed here.

What exactly is it about them that we understand so well?

1.5. The absolute value function is the function:

$$\begin{array}{rccc} |\cdot|: & \mathbb{R} & \to & [0, \infty) \\ & x & \mapsto & \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases} . \end{array}$$

Here are at least three things we know.

N1. $|x| \geq 0$ for all $x \in \mathbb{R}$, and $|x| = 0$ if and only if $x = 0$.
N2. $|\kappa x| = |\kappa|\,|x|$ for all $\kappa \in \mathbb{R}$ and all $x \in \mathbb{R}$.
N3. $|x + y| \leq |x| + |y|$ for all $x, y \in \mathbb{R}$.

Notice that in item (ii), we've introduced a new variable κ, as opposed to just calling that quantity y. What's that all about?

As it so happens, it's all about the following: $\mathbb{R}$ is a **vector space** over $\mathbb{R}$. That is, $\mathbb{R}$ is a set equipped with two operations, called *addition* (denoted by $+$) and *scalar multiplication* (which we shall denote by $\cdot$ or more often just by juxtaposition) satisfying all of the properties of a vector space (see the next section).

So - in this example we are viewing $\mathbb{R}$ both as a set of *vectors* (here is where we denote elements of $\mathbb{R}$ as x and y), and as *scalars* (and here is where we denote them as κ).

1.6. Let $\mathbb{R}^2 := \{(x, y) : x, y \in \mathbb{R}\}$. From Example 2.3, we see that $\mathbb{R}^2$ is a vector space over $\mathbb{R}$, where we set

$$(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2),$$

and

$$\kappa(x, y) := (\kappa x, \kappa y)$$

for all $\kappa \in \mathbb{R}$.

(a) If we set $\|(x, y)\|_1 := |x| + |y|$, we witness an extraordinary event, namely:
  - $\|(x, y)\|_1 \geq 0$ for all $(x, y) \in \mathbb{R}^2$ and $\|(x, y)\|_1 = 0$ if and only if $(x, y) = (0, 0)$.
  - $\|\kappa(x, y)\|_1 = \|(\kappa x, \kappa y)\|_1 = |\kappa x| + |\kappa y| = |\kappa|\ \|(x, y)\|_1$ for all $(x, y) \in \mathbb{R}^2$ and $\kappa \in \mathbb{R}$, and
  - $\|(x_1, y_1)+(x_2, y_2)\|_1 = |x_1+x_2|+|y_1+y_2| \leq |x_1|+|x_2|+|y_1|+|y_2| = \|(x_1, y_1)\|_1 + \|(x_2, y_2)\|_1$.

  We have just checked that $\|\cdot\|_1$ verifies the same three properties (N1), (N2) and (N3) as the absolute value function on $\mathbb{R}$.

(b) If we set $\|(x, y)\|_\infty := \max(|x|, |y|)$, we witness a similar and equally extraordinary event:
  - $\|(x, y)\|_\infty \geq 0$ for all $(x, y) \in \mathbb{R}^2$ and $\|(x, y)\|_\infty = 0$ if and only if $(x, y) = (0, 0)$.
  - $\|\kappa(x, y)\|_\infty = \|(\kappa x, \kappa y)\|_\infty = \max(|\kappa x|, |\kappa y|) = |\kappa|\ \|(x, y)\|_\infty$ for all $(x, y) \in \mathbb{R}^2$ and $\kappa \in \mathbb{R}$, and
  -

$$\begin{aligned}\|(x_1, y_1) + (x_2, y_2)\|_\infty &= \max(|x_1 + x_2|, |y_1 + y_2|) \\ &\leq \max(|x_1| + |x_2|, |y_1| + |y_2|) \\ &\leq \max(|x_1|, |y_1|) + \max(|x_2|, |y_2|) \\ &= \|(x_1, y_1)\|_\infty + \|(x_2, y_2)\|_\infty.\end{aligned}$$

  We have just checked that $\|\cdot\|_\infty$ verifies the same three properties (N1), (N2) and (N3) as the absolute value function on $\mathbb{R}$.

Note that checking that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ satisfy these three properties involved different calculations, but the end result was that both examples indeed satisfied these properties.

1.7. Example. It gets even crazier. We noted in Example 2.4 that

$$\mathcal{C}([0,1],\mathbb{R}) := \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}$$

is a vector space over $\mathbb{R}$.

Let us define, for $f \in \mathcal{C}([0,1],\mathbb{R})$,

$$\|f\|_\infty := \max\{|f(x)| : x \in [0,1]\}.$$

Hold on a minute - how do we know that that maximum exists? That is anything but obvious, but as it so happens, it is one of the things we shall be proving later in this course!!!

Assuming for the moment that the maximum indeed exists for each *continuous* function on $[0,1]$, observe that

- $\|f\|_\infty \geq 0$, and $\|f\|_\infty = 0$ if and only if $|f(x)| = 0$ for all $x \in [0,1]$, i.e. if and only if $f \equiv 0$.
- $\|\kappa f\|_\infty = \max\{|\kappa f(x)| : x \in [0,1]\} = |\kappa| \max\{|f(x)| : x \in [0,1]\} = |\kappa|\ \|f\|_\infty$ whenever $\kappa \in \mathbb{R}$, and
- $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$ for all $f, g \in \mathcal{C}([0,1],\mathbb{R})$. (We leave it to the reader to check this.)

Again - we have just observed that the function $\|f\|_\infty := \max\{|f(x)| : x \in [0,1]\}$ satisfies properties (N1), (N2) and (N3) of the absolute value function on $\mathbb{R}$.

1.8. The point of mathematics is not to make up definitions and look for examples. Rather, it is to look at a number of examples in which you are already interested, to observe some common feature amongst your examples, and to then isolate that common feature by giving it a name.

Once this is done, one tries to prove things about all objects which share that feature, and apply it to new examples that one hasn't yet considered.

Properties (N1), (N2) and (N3) of the absolute value are a case in point.

1.9. Definition. *Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. Let $\mathcal{V}$ be a vector space over $\mathbb{K}$. A **norm** on $\mathcal{V}$ is a function*

$$\begin{array}{rccc} \|\cdot\| : & \mathcal{V} & \to & \mathbb{R} \\ & x & \mapsto & \|x\| \end{array}$$

such that

N1. $\|x\| \geq 0$ *for all* $x \in \mathcal{V}$, *and* $\|x\| = 0$ *if and only if* $x = 0$.
N2. $\|\kappa x\| = |\kappa|\,\|x\|$ *for all* $\kappa \in \mathbb{K}$ *and all* $x \in \mathcal{V}$.
N3. $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in \mathcal{V}$.

*In a fit of boundless energy and creativity, we refer to the ordered pair $(\mathcal{V}, \|\cdot\|)$ as a **normed linear space**.*

In analogy with the absolute value, we think of the norm of a vector $x \in \mathcal{V}$ as its *magnitude* or *size*. Regardless of whether the vector space was a vector space over the real numbers or the complex numbers, we want the magnitude of the vector to be a non-negative real number, of course.

1.10. At the outset of this note, we were considering the quantity $|x_n - x|$ for certain real numbers x_n and x. A moment's thought tells us that $|x_n - x|$ just represents the *distance* between x_n and x. What kinds of properties do we associate to a distance, and especially the distance defined through this absolute value?

1.11. Note that if we denote the distance between real numbers x and y by $d(x,y) := |x-y|$, then

$$\begin{array}{rccc} d: & \mathbb{R}\times\mathbb{R} & \to & \mathbb{R} \\ & (x,y) & \mapsto & |x-y| \end{array}$$

is a function of two variables which satisfies:

M1. $d(x,y) \geq 0$ for all $x, y \in \mathbb{R}$, and $d(x,y) = 0$ if and only if $x = y$.
M2. $d(x,y) = |x-y| = |y-x| = d(y,x)$ for all $x, y \in \mathbb{R}$, and
M3. $d(x,y) = |x-y| \leq |x-z| + |z-y| = d(x,z) + d(z,y)$ for all $x, y, z \in \mathbb{R}$.

1.12. By now we are much older and wiser than we were two pages ago. Let's think about this for a second. Ok, that was a bit too quick. Let's try again, but this time maybe we'll think about it for more than just one second.

Let $(\mathcal{V}, \|\cdot\|)$ be a normed linear space. In analogy to the notation of the distance between two real numbers, let us set

$$d(x,y) := \|x-y\| \text{ for all } x, y \in \mathcal{V}.$$

Observe that (using Properties (N1), (N2) and (N3) of the norm):

(M1) $d(x,y) = \|x-y\| \geq 0$ for all $x, y \in \mathcal{V}$ and $d(x,y) = \|x-y\| = 0$ if and only if $x - y = 0$, i.e. if and only if $x = y$.
(M2) $d(x,y) = \|x-y\| = \|(-1)(y-x)\| = |(-1)|\ \|y-x\| = \|y-x\| = d(y,x)$ for all $x, y \in \mathcal{V}$, and
(M3) $d(x,y) = \|x-y\| = \|(x-z)+(z-y)\| \leq \|x-z\| + \|z-y\| = d(x,z) + d(z,y)$ for all $x, y, z \in \mathcal{V}$.

This is behaving exactly like the distance function does on the real line! Wicked. (I'm referring to the math, not to us.)

1.13. Suppose that $(\mathcal{V}, \|\cdot\|)$ is a normed linear space $\mathcal{V}$ over $\mathbb{K}$, and that $(x_n)_{n=1}^\infty$ is a sequence in $\mathcal{V}$, with $x \in \mathcal{V}$. I can't speak for how *you* would try to define convergence of $(x_n)_n$ to x, but given what I know about the real numbers, I know what I would do.

1.14. Definition. *Let $(\mathcal{V}, \|\cdot\|)$ be a normed linear space over $\mathbb{K}$, and let $(x_n)_{n=1}^\infty$ be a sequence in $\mathcal{V}$. We say that $(x_n)_{n=1}^\infty$ **converges to** $x \in (\mathcal{V}, \|\cdot\|)$ if for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $n \geq N$ implies that*

$$\|x_n - x\| < \varepsilon.$$

1.15. Depending upon how little sleep we need, we might even find the time to consider the following example:

Let $\varnothing \neq X$ be an arbitrary non-empty set. Consider the map

$$\begin{array}{rccc} \mu: & X \times X & \to & \mathbb{R} \\ & (x,y) & \mapsto & \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}. \end{array}$$

Note that

- $\mu(x,y) \geq 0$ for all $x, y \in X$ and $\mu(x,y) = 0$ if and only if $x = y$;
- $\mu(x,y) = \mu(y,x)$ for all $x, y \in X$; and

- $\mu(x, y) \le \mu(x, z) + \mu(z, y)$ for all $x, y, z \in X$.

(The third item is the only one that takes any thought.)

But hold on – aren't these pretty much the properties (M1), (M2) and (M3) we had above? We didn't even start with a vector space, let alone a normed linear space!!!

1.16. Definition. *A* ***metric space*** *is an ordered pair (X, d), where $\varnothing \ne X$ is a non-empty set, and $d : X \times X \to \mathbb{R}$ is a function which satisfies*

M1. $d(x, y) \ge 0$ *for all* $x, y \in X$ *and* $d(x, y) = 0$ *if and only if* $x = y$.
M2. $d(x, y) = d(y, x)$ *for all* $x, y \in X$, *and*
M3. *for all* $x, y, z \in X$,

$$d(x, y) \le d(x, z) + d(z, y).$$

Informally, we say that X is the metric space and (not so informally) we say that d is a metric on X. In paragraph 1.15, we argued that μ is a metric on X. This particular metric is called the **discrete metric** on X.

1.17. Let's go back to our original definition of limits one more time. It depended upon our having a notion of the distance between x_n and x. The fact that $\mathbb{R}$ was a normed linear space (with $|\cdot|$ acting as the norm!!!) was incidental. In other words, at this stage, we can general the notion of limits even further!

1.18. Definition. *Let (X, d) be a metric space. A sequence $(x_n)_{n=1}^{\infty}$ in X is said to* ***converge to the limit*** *$x \in X$ if for all $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $n \ge N$ implies that*

$$d(x_n, x) < \varepsilon.$$

1.19. What have we done? (By that I mean, *let's have a look at what we've done*, as opposed to "*Oh no, no, no, no, no!!! What have we done*?!?!?!")

In paragraph 1.12, we demonstrated that if $(\mathcal{V}, \|\cdot\|)$ is a normed linear space, then the map

$$d(x, y) := \|x - y\|$$

defines a metric on $\mathcal{V}$. We refer to this particular metric on $\mathcal{V}$ as the **metric induced by the norm**, mostly because it's a metric, and because the norm induces it. (Funny how logical we can sometimes be, wouldn't you agree?)

Of course, a normed linear space is a vector space, and a vector space is also a non-empty set. We could have equipped it with the discrete metric from paragraph 1.15. This would be a metric on $\mathcal{V}$, but unless $\mathcal{V} = \{0\}$, it could never be the metric induced by the norm (why not?)

When dealing with normed linear spaces, the natural metric to consider is the metric induced by the norm, and unless someone specifically states that they will not be using that metric, then the assumption is that if someone speaks of a metric on a normed linear space, then this is the metric to which they are referring.

We eventually showed that the notion of a limit can be extended beyond the real numbers and all the way up to metric spaces. Moreover, the definition of a limit in the metric space setting is simply a reformulation of the definition of a limit for sequences of real numbers. In other words, to correctly understand limits in metric spaces, it is absolutely necessary that we correctly understand what a limit of a sequence of real numbers is and how to prove things with these. So enough culture for now – let's get back to work.

2. Vector spaces

2.1. In this section we shall use the notation $\mathbb{K}$ to denote either $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$. After having chosen one of these two possibilities, we do not change our choice for the rest of the conversation.

2.2. Definition. *A* ***vector space*** $\mathcal{V}$ *over* $\mathbb{K}$ *is a non-empty set equipped with two operations called* ***addition*** *(which we denote by $+$) and* SCALAR MULTIPLICATION *(which we denote by $\cdot$ or by juxtaposition), such that the following conditions hold for all $x, y, z \in \mathcal{V}$ and all $\alpha, \beta \in \mathbb{K}$:*

V1. $x + y = y + x$.
V2. $x + (y + z) = (x + y) + z$.
V3. *There exists an element* $0 \in \mathcal{V}$ *such that* $x + 0 = x$.
V4. *There exists an element* $-x \in \mathcal{V}$ *such that* $(-x) + x = 0$.
V5. $0x = 0$.
V6. $1x = x$.
V7. $\alpha(x + y) = \alpha x + \alpha y$.
V8. $(\alpha + \beta)x = \alpha x + \beta x$.
V9. $(\alpha\beta)x = \alpha(\beta x)$.

Note that when we say that addition and scalar multiplication are *operations* on $\mathcal{V}$, we mean that $x + y \in \mathcal{V}$ and $\alpha x \in \mathcal{V}$ for all $x, y \in \mathcal{V}$ and $\alpha \in \mathbb{K}$.

2.3. Examples. We shall list a few examples of vector spaces over $\mathbb{R}$ and over $\mathbb{C}$, but the verification of the fact that these are indeed vector spaces is left as exercises.

(a) For any $n \geq 1$, $\mathbb{R}^n := \{(x_1, x_2, \dots, x_n) : x_k \in \mathbb{R}, 1 \leq k \leq n\}$ is a vector space over $\mathbb{R}$, where we set
$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
and
$$\alpha(x_1, x_2, \dots, x_n) := (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$$
for all $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

(b) For any $n \geq 1$, $\mathbb{C}^n = \{(x_1, x_2, \dots, x_n) : x_k \in \mathbb{C}, 1 \leq k \leq n\}$ is a vector space over $\mathbb{C}$, where we set
$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
and
$$\alpha(x_1, x_2, \dots, x_n) := (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$$
for all $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$.

(c) For any $n \geq 1$, $\mathbb{C}^n = \{(x_1, x_2, \dots, x_n) : x_k \in \mathbb{C}, 1 \leq k \leq n\}$ is a vector space over $\mathbb{R}$, where we set
$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) := (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
and
$$\alpha(x_1, x_2, \dots, x_n) := (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$$
for all $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{C}^n$ and $\alpha \in \mathbb{R}$.

2.4. Examples.

(a) Let $\mathcal{V} = \mathcal{C}([0,1], \mathbb{R}) := \{f : [0,1] \to \mathbb{R} : f \text{ is continuous}\}$. Then $\mathcal{V}$ is a vector space over $\mathbb{R}$. The most difficult part of the proof of this fact is that if f, g are continuous, real-valued functions on $[0,1]$, and if $\alpha \in \mathbb{R}$, then $f + g$ and αf are continuous functions.

(b) Similarly $\mathcal{V} = \mathcal{C}([0,1], \mathbb{C})$ is a vector space both over $\mathbb{R}$ and over $\mathbb{C}$.

(c) Let $c_0(\mathbb{K}) := \{(x_n)_{n=1}^\infty : x_n \in \mathbb{K}, \lim_{n\to\infty} x_n = 0\}$. Then $c_0(\mathbb{K})$ is a vector space using the operations

$$(x_n)_n + (y_n)_n := (x_n + y_n)_n,$$

and

$$\alpha(x_n)_n := (\alpha x_n)_n,$$

for all $(x_n)_n$, $(y_n)_n \in c_0(\mathbb{K})$ and $\alpha \in \mathbb{K}$.

(d) Let $\ell_\infty(\mathbb{K}) := \{(x_n)_{n=1}^\infty : x_n \in \mathbb{K}, \sup_{n\geq 1} |x_n| < \infty\}$. Then $\ell_\infty(\mathbb{K})$ is a vector space using the same addition and scalar multiplication as those used for $c_0(\mathbb{K})$.

We can define a norm $\|\cdot\|_\infty$ on $\ell_\infty(\mathbb{K})$ via:

$$\|(x_n)_{n=1}^\infty\|_\infty := \sup\{|x_n| : n \geq 1\}.$$

We leave it as an exercise for the reader to check that this is indeed a norm on $\ell_\infty(\mathbb{K})$.

(e) Let $\mathbb{M}_{m\times n}(\mathbb{K}) := \{X = [x_{i,j}]_{1\leq i\leq m, 1\leq j\leq n} : x_{i,j} \in \mathbb{K} \text{ for all } i, j\}$ denote the set of all matrices with m rows and n columns with entries in $\mathbb{K}$. Then this is a vector space over $\mathbb{K}$ if we define

$$[x_{i,j}] + [y_{i,j}] := [x_{i,j} + y_{i,j}],$$

and

$$\alpha[x_{i,j}] := [\alpha x_{i,j}].$$

3. Exercises

Question 1.

Prove that a sequence $(X_n := (x_{n,1}, x_{n,2}, \dots, x_{n,m}))_{n=1}^\infty$ in $\mathbb{R}^m$ converges to a vector $Y := (y_1, y_2, \dots, y_m) \in (\mathbb{R}^m, \|\cdot\|_\infty)$ if and only if

$$\lim_{n\to\infty} x_{n,k} = y_k, \quad 1 \le k \le m.$$

Question 2.

Formulate and prove an analogous statement if we replace $\|\cdot\|_\infty$ by $\|\cdot\|_1$ in Question 1.

Question 3.

Does a similar statement hold for $(\ell_\infty(\mathbb{R}), \|\cdot\|_\infty)$? That is, suppose that for each $n \ge 1$, we have an element

$$X_n := (x_{n,k})_{k=1}^\infty \in \ell_\infty(\mathbb{R}),$$

and suppose that $Y = (y_k)_{k=1}^\infty \in \ell_\infty(\mathbb{R})$.

Is it true that using the distance function $d(W, Z) := \|W - Z\|_\infty$ on $\ell_\infty(\mathbb{R})$ that we obtained as in Definition 1.14, we have that

$$\lim_{n\to\infty} X_n = Y \quad \text{if and only if} \quad \lim_{n\to\infty} x_{n,k} = y_k \quad \text{for all } k \ge 1?$$

Question 4. Prove that

$$\|(x_n)_{n=1}^\infty\|_\infty := \sup\{|x_n| : n \ge 1\}$$

defines a norm on the vector space $\ell_\infty(\mathbb{K})$ defined in Example 2.4 (d).

Question 5.

Let $\varnothing \ne X$ be a non-empty set, and let μ denote the discrete metric on X as defined in paragraph 1.15.

Which sequences in X converge? That is, if $(x_n)_{n=1}^\infty$ is a sequence in X and $\lim_{n\to\infty} x_n = x$ with respect to the discrete metric μ, then what can you say about $(x_n)_{n=1}^\infty$?

Question 6.

For those of you with a background in vector spaces and matrices:

Let

$$\begin{array}{rccc} \tau : & \mathbb{M}_n(\mathbb{C}) \times \mathbb{M}_n(\mathbb{C}) & \to & \mathbb{R} \\ & (X, Y) & \mapsto & \operatorname{rank}(X - Y). \end{array}$$

Prove that τ is a metric on $\mathbb{M}_n(\mathbb{C})$. What does it mean for a sequence $(T_k)_{k=1}^\infty$ in $\mathbb{M}_n(\mathbb{C})$ to converge to a matrix $T \in (\mathbb{M}_n(\mathbb{C}), \tau)$?

Are you impressed yet?

Math 147 - September 2020
Topology
L.W. Marcoux

1. Introduction.

1.1. Recall that if $f : \mathbb{R} \to \mathbb{R}$ is a function, then we say that f is **continuous at the point** $x_0 \in \mathbb{R}$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \varepsilon$.

Our goal is to try to rephrase this concept in such a way that we can generalise it. We have already seen how to do this in **metric spaces**. The concept we wish to examine today is that of **topological spaces**, which generalise metric spaces.

1.2. Definition. *Let $G \subseteq \mathbb{R}$ be a set. We say that G is* ***open*** *if, for each $x \in G$, there exists $\delta > 0$ such that $(x - \delta, x + \delta) \subseteq G$.*

We say that $F \subseteq \mathbb{R}$ is ***closed*** *if $G := \mathbb{R} \setminus F$ is open.*

1.3. Remarks. Before we get to examples, it is important to stress a couple of things.

(a) The $\delta > 0$ appearing in the definition of an open set G *depends upon x*. That is, the $\delta = \delta_x$ that works for some $x \in G$ is not necessarily the same as the δ_y that works for $y \in G$.

(b) Open and closed are not opposites! The vast majority of sets are neither open nor closed, and some (which ones?) are both!!

(c) We tend to use the letter "G" for open (from the German *geöffnet*) and "F" for closed (from the French *fermé*).

1.4. Examples.

(a) Let $G = (0, 1) \subseteq \mathbb{R}$, and fix $x_0 \in (0, 1)$. Let $\delta = \min(|x_0 - 0|, |x_0 - 1|)$. Then

$$(x_0 - \delta, x_0 + \delta) \subseteq (0, 1).$$

Why is this obvious? Note that $|x_0 - 0|$ is nothing more than the distance from x_0 to 0, while $|x_0 - 1|$ is the distance from x_0 to 1. If $|y - x_0| < \delta$, then it is closer to x_0 than 0 and closer to x_0 than 1 is, so it must lie in $(0, 1)$. But $|y - x_0| < \delta$ if and only if $y \in (x_0 - \delta, x_0 + \delta)$.

Since $x_0 \in G$ was arbitrary, $(0, 1)$ is open in $\mathbb{R}$.

Note that if $x_0 = \frac{1}{2}$, then we are using $\delta = \frac{1}{2}$, while if $x_0 = \frac{1}{8}$ or $x_0 = \frac{7}{8}$, then $\delta = \min(\frac{1}{8}, \frac{7}{8}) = \frac{1}{8}$. *It depends upon our choice of x_0*!!!

(b) By a similar argument, $G_1 := (1, \frac{3}{2}) \subseteq \mathbb{R}$ is open. Note also that $G \cup G_1 = (0, 1) \cup (1, \frac{3}{2})$ is open, even though it is not a single interval.

(c) $G := \varnothing \subseteq \mathbb{R}$ is open. Why?

(d) Let $G = (0, 1) \cup (2, 3) \cup (4, 5) \cup (6, 7) \cup \cdots \cup (2n, 2n + 1) \cup \cdots$. Then G is open in $\mathbb{R}$.

(e) Let $F = [0, 1]$. Then $G := \mathbb{R} \setminus F = (-\infty, 0) \cup (0, \infty)$ is open (check!). Thus F is closed.

(f) Let $F = \{\pi\}$. Then $G := \mathbb{R} \setminus F = (-\infty, \pi) \cup (\pi, \infty)$ is open (check!). Thus F is closed.

(g) Let $H = [0, 1)$. Then $x_0 = 0 \in H$, but for any $\delta > 0$, $(x_0 - \delta, x_0 + \delta)$ fails to be contained in H. Thus H is not open.

Consider $K := \mathbb{R} \setminus H = (-\infty, 0) \cup [1, \infty)$. Then $1 \in K$, but for any $\delta > 0$, $(1 - \delta, 1 + \delta)$ fails to be contained in K. Thus K is not open, so H is not closed.

That is, H is neither open nor closed.

(h) We have seen that $G = \varnothing$ is open in $\mathbb{R}$. Let $F = \mathbb{R} \setminus G = \mathbb{R}$. If $x_0 \in F$, then for *any* $\delta > 0$, $(x_0 - \delta, x_0 + \delta) \subseteq \mathbb{R} = F$, so F is open. But then G is closed.

That is, G is both open *and* closed. (People sometimes say that G is a *clopen* set.)

(i) Let $H = \mathbb{Q}$. Then $7 \in H$, but for any $\delta > 0$, then interval $(7 - \delta, 7 + \delta)$ contains an irrational number (i.e. $(7 - \delta, 7 + \delta)$ fails to be contained in $H = \mathbb{Q}$), so $\mathbb{Q}$ is not open. If $K = \mathbb{R} \setminus H = \mathbb{R} \setminus \mathbb{Q}$, then $\sqrt{2} \in H$, but again - for any $\delta > 0$, then interval $(\sqrt{2} - \delta, \sqrt{2} + \delta)$ contains a rational number (i.e. $(\sqrt{2} - \delta, \sqrt{2} + \delta)$ fails to be contained in K), so K is not open and thus $\mathbb{Q}$ is not closed.

1.5. Let $f : \mathbb{R} \to \mathbb{R}$ be a function, and suppose that $x_0 \in \mathbb{R}$. Suppose that f is continuous at x_0, and let $G \subseteq \mathbb{R}$ be an open set that contains $f(x_0)$. Then there exists $\varepsilon > 0$ such that $(f(x_0) - \varepsilon, f(x_0) + \varepsilon) \subseteq G$.

Let $H = f^{-1}(G) := \{x \in \mathbb{R} : f(x) \in G\}$. Note that $x_0 \in f^{-1}(G)$, since $f(x_0) \in G$ by our choice of G. Since f is continuous at x_0, there exists $\delta > 0$ such that $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \varepsilon$. In other words, $x \in (x_0 - \delta, x_0 + \delta)$ implies that $f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$ and so $f(x) \in G$; that is, $x \in f^{-1}(G) = H$.

Notice, by the way, that $M := (x_0 - \delta, x_0 + \delta)$ is an open set which contains x_0.

We've just argued that if G is an open set that contains $f(x_0)$, and if f is continuous at x_0, then there is an open set M such that $x_0 \in M \subseteq f^{-1}(G)$.

1.6. Let's turn this around. Again, suppose that $f : \mathbb{R} \to \mathbb{R}$ be a function, and suppose that $x_0 \in \mathbb{R}$. Suppose furthermore that whenever G is an open set that contains $f(x_0)$, there is an open set M such that $x_0 \in M \subseteq f^{-1}(G)$.

Notice that we have not introduced any specific notion of distance here: we are speaking entirely about "*open sets*".

Let $\varepsilon > 0$, and let $G := (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$. Then G is an open set in $\mathbb{R}$, so by hypothesis, there exists an open set M in $\mathbb{R}$ such that $x_0 \in M \subseteq f^{-1}(G)$.

Now, M is open and $x_0 \in M$, so there exists $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq M$. It follows that $(x_0 - \delta, x_0 + \delta) \subseteq f^{-1}(G)$. In other words, $|x - x_0| < \delta$ implies that $f(x) \in G = (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$, or better yet, $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \varepsilon$.

Since this was true for all $\varepsilon > 0$, we see that f is continuous at x_0.

CONCLUSION. *A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at $x_0 \in \mathbb{R}$ if and only if whenever $G \subseteq \mathbb{R}$ is an open set that contains $f(x_0)$, we can find an open set M satisfying $x_0 \in M \subseteq f^{-1}(G)$.*

1.7. Now let X and Y be non-empty sets. Suppose that we had no idea how to measure distance between elements of X or between elements of Y, but that we did have an idea of what it meant for a subset of X (resp. a subset of Y) to be open. We could use the characterisation above to *define* what it means for f to be continuous at a point $x_0 \in X$!!

This begs the question: can we just choose arbitrary sets to be open? Or does there have to be some sort of rule for choosing open sets?

The idea is to look at open sets in $\mathbb{R}$ for inspiration.

1.8. Exercise.

(a) Prove that $\varnothing$ and $\mathbb{R}$ are open sets in $\mathbb{R}$.
(b) Prove that if $\{G_\lambda : \lambda \in \Lambda\}$ is *any* collection of *open* subsets of $\mathbb{R}$, then

$$G := \cup_{\lambda\in\Lambda} G_\lambda$$

is open in $\mathbb{R}$.
(c) Let $N \in \mathbb{N}$ and suppose that $H_n \subseteq \mathbb{R}$ is open, $1 \le n \le N$. Prove that $H = \cap_{n=1}^N H_n$ is open.

1.9. Example. It is not difficult to show that for each $n \ge 1$, then interval $H_n := (-\frac{1}{n}, \frac{1}{n})$ is open in $\mathbb{R}$. On the other hand (yeah, that one),

$$H := \cap_{n=1}^\infty H_n = \cap_{n\ge 1}(-\frac{1}{n}, \frac{1}{n}) = \{0\}$$

is not open in $\mathbb{R}$, so we cannot extend part (c) above to infinite intersections of open sets.

1.10. Definition. *Let $\varnothing \ne X$ be a set. A collection τ of subsets of X is said to be a **topology** for X if:*

(a) *$\varnothing$, X are in τ.*
(b) *If $G_\lambda \in \tau$ for all λ belonging to some set Λ, then $G := \cup_{\lambda\in\Lambda} G_\lambda \in \tau$. (We say that τ is* closed under arbitrary unions*, although this notion of closed is unrelated to "topologically closed".)*
(c) *If $N \in \mathbb{N}$ and $G_1, G_2, \ldots, G_N \in \tau$, then $G := \cap_{n=1}^N G_n \in \tau$. (We say that τ is* closed under finite intersections*.)*

*If τ is a topology on X, then we refer to the elements of τ as **open sets**, and a set $F \subseteq X$ is said to be **closed** if $X \setminus F \in \tau$; i.e. if $X \setminus F$ is open.*

*We say that (X, τ) is a **topological space**.*

1.11. Example. Let $\tau = \{G \subseteq \mathbb{R} :$ for each $x_0 \in G$ there exists $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq G\}$.

Then τ is a topology on $\mathbb{R}$, referred to as the **standard topology** on $\mathbb{R}$. Note that $G \in \tau$ (and hence is τ-open) if and only if G is open in the sense described in Definition 1.2.

1.12. Example. Let $\tau_d = \mathcal{P}(\mathbb{R}) = \{Y : Y \subseteq \mathbb{R}\}$. Then τ_d is a topology on $\mathbb{R}$, known as the **discrete** topology on $\mathbb{R}$.

1.13. Example.

Let $X = \{\text{dog}, \text{kangaroo}, \text{telescope}\}$. Set

$$\tau := \{\varnothing, \{\text{dog}\}, \{\text{kangaroo}\}, \{\text{dog}, \text{kangaroo}\}, \{\text{dog}, \text{kangaroo}, \text{telescope}\}\}.$$

We invite the reader to check that this is a topology on X.

1.14. Example. Let $\varnothing \ne X$ be a set. Then $\tau := \{\varnothing, X\}$ is a topology on X, referred to as the **trivial topology** on X. Let's see if you can guess why.

1.15. Consider $\mathbb{R}$ equipped with the discrete topology τ_d. Let (X, τ) be the topological space from Example 1.14

Let $f : X \to \mathbb{R}$ be the function $f(\text{dog}) = \pi$, $f(\text{kangaroo}) = -2000$, $f(\text{telescope}) = e^2$.

Note that $x_0 := \text{telescope} \in X$ and $G := \{e^2\} \in \tau_d$ – after all, *any* subset of $\mathbb{R}$ lies in τ_d – is a set that contains $e^2 = f(x_0)$.

Consider $f^{-1}(G) = f^{-1}(\{e^2\}) = \{\text{telescope}\}$. Does there exist any open set M in X such that

$$\text{telescope} \in M \subseteq f^{-1}(G) = \{\text{telescope}\}?$$

Note that (X, τ) does not have *any* open sets contained in $\{\text{telescope}\}$, so the answer is "No".

Conclusion. f is not continuous at telescope. You are invited to verify that f *is* continuous at dog and at kangaroo.

1.16. HA HA HA HA HA HA HA!!! No ε's nor δ's, and yet we can talk about f being continuous at a point in X!!! HA HA HA HA HA!!!

1.17. Example. Let (X, τ_d) and (Y, τ_Y) be topological spaces, where τ_d is the discrete topology on X. Let $f : X \to Y$ be a function and $x_0 \in X$. Let $G \in \tau_Y$ be any τ_Y-open subset of Y that contains $f(x_0)$. Consider $f^{-1}(G) \subseteq X$.

By the definition of the discrete topology, $M := f^{-1}(G)$ is open in (X, τ_d), and thus

$$x_0 \in M \subseteq f^{-1}(G)$$

with $M \in \tau_d$ shows that f is continuous at x_0.

Again - no discussion of ε's, δ's, and we have no idea whatsoever what an element of X or Y looks like.

Conclusion. Topologies rock.