1 Real Analysis

1.1 Proving $\mathbb{R}$ Is What We Expect

We've constructed something out of rational numbers that seems to fill in the gaps in $\mathbb{Q}$, but there are reasons not to be convinced that this is $\mathbb{R}$ as we know it, or that we've constructed something worth constructing. For one thing, I justified the "gaps in $\mathbb{Q}$" argument by providing a polynomial that didn't have a zero on $\mathbb{Q}$, but by the same token, there are gaps in $\mathbb{R}$, because $x^2 + 1$ doesn't have any zeros in $\mathbb{R}$. This is admittedly a slightly different situation, because for $x^2 - 2$ we could numerically approximate a rational zero, i.e. for any $\epsilon > 0$ we could find some q such that $|q^2 - 2| < \epsilon$ (choose $q = s_n, n >$ the N corresponding to ϵ, where s_n is the Cauchy sequence of rationals that we identify as $\sqrt{2}$). However, we can't do that in this case, because $|x^2 + 1| \geq 1$, so for any $\epsilon < 1$, we can't get $\epsilon-$close to the zero. Even so, we still have some notion that $\mathbb{R}$ might need to be completed or closed under some operation, so we don't know that there's anything special about $\mathbb{R}$ yet.

Further, I haven't shown you that we can fill all the gaps on the real line through Cauchy sequences of rational numbers; how do we know that there aren't any holes in this extension? How do we know that the zero for $x^2 + 1$ isn't a real number that we just can't get to through this method?

To address these, I'll prove two things: that $\mathbb{R}$ with this definitions forms a field (addressing the "nothing special" concern) and that $\mathbb{R}$ is <u>complete</u>, i.e. every Cauchy sequence of elements of $\mathbb{R}$ converges to something in $\mathbb{R}$.

Fields   A full treatment of groups, rings, and fields would require a completely different note, so I'll just present all of the field axioms. This isn't as compact as defining it in terms of abelian groups, but it's more straightforward when we don't really need to know what an abelian group is.

A field is any set F on which there are two operations that satisfy the following axioms.

The first operation is addition, satisfying

1. associativity, $(a + b) + c = a + (b + c) \ \forall a, b, c \in F$;

2. existence of an additive identity, usually called 0, such that $a + 0 = 0 + a = a \ \forall a \in F$;

3. existence of additive inverses for each element in the field, i.e. $\forall a \in F, \exists -a \in F \, \text{s.t.} \, a + (-a) = (-a) + a = 0$;

4. commutativity ($a + b = b + a \ \forall a, b \in F$).

The second operation is multiplication, satisfying

5. associativity, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

6. existence of a multiplicative identity, usually called 1, such that $a \cdot 1 = 1 \cdot a = a \ \forall a \in F$;

7. existence of multiplicative inverses for each element in the field except possibly 0, i.e. $\forall a \in F/\{0\}, \exists a^{-1} \in F/\{0\} \, \text{s.t.} \, a^{-1} \cdot a = a \cdot a^{-1} = 1$;

8. commutativity ($a \cdot b = b \cdot a \ \forall a, b \in F$).

Finally, the combination of addition and multiplication has to satisfy

9. left and right distributivity: $\forall a, b, c \in F, a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

This might look similar to the axioms for a vector space if you're familiar with those, and it is: it's almost exactly the same, just that associativity of multiplication in a vector space shows an equivalence between field-field and field-vector multiplication whereas this is just a fact about the field.

So we've introduced all this theory to say that we can show that $\mathbb{R}$ is a field. This should help you intuitively think about the construction we just did from $\mathbb{Q}$, i.e. that we've done something nontrivial and made an object that we can do algebra over in more generality than with $\mathbb{Q}$ (which is also a field.)

> **$\mathbb{R}$ is a field**
>
> If we define addition and multiplication by adding and multiplying termwise the rational Cauchy sequences that represent each real number, i.e.,
>
> $$((s + t)_n) = (s_n) + (t_n) = (s_n + t_n)$$
> $$((st)_n) = (s_n)(t_n) = (s_n t_n)$$
>
> and we select $(0, 0, 0, \dots)$ as the additive identity and $(1, 1, 1, \dots)$ as the multiplicative identity, then $\mathbb{R}$ is a field.

Proof. We can show addition satisfies associativity, because addition of rational numbers is associative and we're essentially just doing that infinitely many times. We defined a zero element and we can show that works, because it works termwise. There are additive-inverse sequences, which we get just by inverting every term, i.e. $-(s_n) = (-s_n)$. Addition commutes, because addition of every rational term commutes.

Similarly, we can show multiplication satisfies associativity because it holds termwise. We defined a unit element and we can show that works, because it works termwise. There are multiplicative-inverse sequences, which we get just by inverting every term, i.e. $(s_n)^{-1} = (s_n^{-1})$. Multiplication commutes, which it does because multiplication of every rational term commutes.

Finally, distributivity holds because it holds termwise for rational terms. $\square$

A fun note here is that the additive identity is the sequence that's all zeros, and the multiplicative identity is the sequence that's all ones. We said above that if a sequence is all one constant rational, we can identify it as that element of $\mathbb{Q}$, which means the 0 and 1 from $\mathbb{Q}$ carry over here. This is a nice property that we'd expect from something that contains $\mathbb{Q}$.

Of course, all this is predicated on the rational numbers being a field, which we could also prove from first principles (going back to the successor function and more on set theory). But the major takeaway is that if we define this construction from a field, we'll get another field, which makes it believable that the real numbers are fundamental in some sense, because they're closed under some natural operations.

Closure doesn't necessarily mean the set is special. For example, the set of integers that divide some n are also closed under addition and multiplication (although they're not a field as there's no multiplicative identity unless $n = 1$). Let's make a stronger claim: the reals are special among fields because we've filled in all the gaps; it's the <u>unique complete field</u> that extends the rationals. If we used another construction of the reals, we would be able to show that it is isomorphic to this one. (Technically, it's the unique complete ordered field, but order matters more from a set-theoretic than an analytic perspective, so we'll disregard this.) We've seen it's a field, so now let's show it's complete.

> **Completeness of the reals**
>
> Every Cauchy sequence of real numbers converges to a real number.

This definition will make sense when we talk about a generalization of the reals to metric spaces. For now, note that this definition means that for every Cauchy sequence of real numbers, there exists a Cauchy sequence of rational numbers converging to the same limit. In other words, when we took the set of Cauchy sequences of rational numbers, we didn't leave any gaps: we could get as precise as we wanted by taking a sequence of real numbers converging to whatever we wanted, and we'd still have been able to do it with just the rationals. This is kind of a crazy idea when you think about it: there's uncountably many real numbers (arguments like Cantor diagonalization show this and also clarify what is meant by countably vs. uncountably infinite), and yet we can reach them with just the countable number of rationals.

To prove this, we introduce the concepts of limit superior ($\limsup$) and limit inferior ($\liminf$); these are essentially the supremum and infimum of a sequence for arbitrarily large n.

> **Limit inferior and limit superior**
>
> $\liminf s_n$ is the infimum of the set of values a sequence takes on as n goes to infinity. Let $S_n = \{s_m \mid m \geq n\}$; then $\liminf s_n = \lim_{n\to\infty} \inf S_n$. Similarly, $\limsup s_n$ is the supremum of that same set: $\limsup s_n = \lim_{n\to\infty} \sup S_n$.

These will be useful in proving the reals are complete by bounding a sequence of real numbers below and above, and showing that the bounding sequence (the ones whose limits are the $\liminf$ and $\limsup$ of the sequence) converge to the same limit, so the actual limit of the sequence must be the same as both of those. Let's clarify why this is true.

If a sequence converges, then we expect that under a limit the maximum and minimum attainable values should both converge to the actual limit of the sequence: as you go further and further in the sequence, the supremum of all the sequence values from that point onwards should get closer and closer to the limit, and so should the infimum. For example, $s_n = \frac{1}{2^n}$ has a $\liminf$ of 0 because it's bounded below by 0, and a $\limsup$ of 0 because for every $\epsilon > 0$, I can give an N such that $n > N \implies s_n < \epsilon$. So I can make the sequence elements arbitrarily small positive numbers, meaning that in the limit the sequence is at most 0. Since it's at least 0 and at most 0, it must be exactly 0. We can formalize this:

> **$\liminf s_n = \limsup s_n = \lim s_n$ iff it converges**
>
> If a sequence converges, the $\liminf$ and $\limsup$ are the same and both equal the limit. However, if a sequence doesn't converge, the $\limsup$ and $\liminf$ are not the same (consider $s_n = (-1)^n$). This is an if-and-only-if statement, so we can also say that if $\liminf s_n = \limsup s_n$ then s_n converges to both of them. (I'm skipping the proof for now, but I may come back and add it later.)

Now, we're ready to prove that every Cauchy sequence of real numbers converges to a real number.

Proof of completeness of the reals

Proof. Consider a Cauchy sequence of real numbers (s_n). Because the sequence is Cauchy, we know it must be bounded (its lim sup and lim inf are both finite), so consider the sequences $(l_n) = (\inf\{s_m \mid m \geq n\}), (u_n) = (\sup\{s_m \mid m \geq n\})$. (l_n) is bounded below by the infimum of (s_n), and (u_n) is bounded above by the supremum of (s_n), both of which are finite because (s_n) is bounded. Further, the sequence of infima can only increase and the sequence of suprema can only decrease (i.e. they are both monotonic). Therefore (l_n) converges to $\liminf s_n \in \mathbb{R}$ and (u_n) converges to $\limsup s_n \in \mathbb{R}$.

Fix $\epsilon > 0$; because s_n is Cauchy, we can find $N \in \mathbb{N}$ such that $m \geq N$ implies $|s_m - s_N| < \frac{\epsilon}{2}$, that is, $s_N - \frac{\epsilon}{2} < s_m < s_N + \frac{\epsilon}{2}$. The sequence of infima (l_n) is strictly less than the sequence itself, and the sequence of suprema (u_n) is strictly greater, but both are the tightest possible bounds which means each infimum must be greater than this bound we've placed on s_n, and each supremum must be less than the bound. That is, for all $m \geq N$, $l_m > s_N - \frac{\epsilon}{2}$ and $u_m < s_N + \frac{\epsilon}{2}$. Therefore, consider the difference between the inf and sup:

$$|u_m - l_m| = u_m - l_m < s_N + \frac{\epsilon}{2} - s_N + \frac{\epsilon}{2} = \epsilon$$

i.e. for all $\epsilon > 0$ there exists an N such that $m \geq N \implies |u_m - l_m| < \epsilon$. Therefore $(u_n - l_n)$ is a sequence converging to 0, so $\lim(u_n - l_n) = 0 \implies \limsup s_n - \liminf s_n = 0$. Therefore the sequence converges to $\lim s_n = \limsup s_n = \liminf s_n$. $\square$

This should convince you that $\mathbb{R}$ is special and is a useful thing to have constructed. We could still come up with an extension of $\mathbb{R}$ based on the previous argument that $x^2 + 1$ doesn't have any zeros in $\mathbb{R}$, but we've shown that we can't get arbitrarily close to a zero, so this does not represent a "gap" to be filled in $\mathbb{R}$. (We can still construct $\mathbb{C}$ from $\mathbb{R}$ using the idea of a field extension, but again, that's a topic for the Algebra version of this document.)

1.2 Continuous Functions on $\mathbb{R}$

We have a formal definition of what it means to be complete, and we know that $\mathbb{R}$ satisfies it, which is enough for us to be able to do calculus. To start with this, let's define what a function is, and define what it means for a function to be continuous.

Functions   A function from set S to set T is a choice for every $s \in S$ of some $f(s) \in T$. This is denoted $f : S \to T$. (Sometimes, a function is denoted along with its action on an arbitrary element of S, as in $f : S \to T, s \to f(s)$.)

We call S the domain of f and T the codomain of f. T is sometimes also referred to as the range or image. Terminology is confusing here; personally, I like codomain as the full space in which the output values live that is somehow analogous to the domain (for example, if S is a field then the codomain of $f : S \to T$). The image is the set of values that are actually attained, i.e. $\{f(s) \mid s \in S\}$. The range is used as either of these. I'll try not to use range, but usually I like to use it interchangeably with image.

Note that this is only partially an aesthetic choice: once you've defined a function and define the sets it acts between, the codomain and image are well defined and not up to interpretation. The choice comes in setting what T is. For example, consider the two functions

$$f : \mathbb{R} \to \mathbb{R}, x \to x^2$$
$$g : \mathbb{R} \to \mathbb{R}^+, x \to x^2$$

where $\mathbb{R}^+ = [0, \infty) \subset \mathbb{R}$. The codomain of f is $\mathbb{R}$, but the image is $\mathbb{R}^+$ because squaring a real number can only get you a positive real number. The codomain of g is $\mathbb{R}^+$, and the image is also $\mathbb{R}^+$. The "aesthetic choice" part of this comes in choosing whether you want to consider f or g as the function that squares real numbers. I prefer f because you have the terminology to deal with both the field where the output values live, and the actually-attained output values, but g is also valid.

> **Continuous functions on $\mathbb{R}$**
>
> We say a function $f : S \to \mathbb{R}, x \to f(x)$ is continuous at some point $a \in S$ if either of the following (equivalent) properties hold:
>
> - for all sequences (a_n) such that $a_n \in S$ for all n, we can say $\lim f(a_n) = f(a) = f(\lim a_n)$.
>
> - for all $\epsilon > 0$, there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.
>
> We say f is continuous if it is continuous at a for all $a \in S$.

The first property (the sequence definition of continuity) is probably the more fundamental one, because it comes clearly out of the idea of convergence of a sequence that we established. Essentially, whenever we choose a sequence that converges to the input point we care about, applying the function termwise to the sequence should give us another sequence that converges, and we say that sequence converges to $f(a)$. This is a reasonable notion of what it means to be continuous: you could imagine a geometric mapping of the real line to some curve in space such that you could draw it without lifting your pen. This would have the property that if two sequences represent real numbers that are arbitrarily close together, then the points on the curve they map to also represent real numbers whose closeness is related to that. Put more simply, a continuous function should preserve the idea that there are no gaps in the real line when we map every point in its domain to something in its range.

The second property (the epsilon-delta definition of continuity) is more algebraically useful, and we'll use it more when we move from the real line to metric spaces. We claim that the two definitions are equivalent.

Proof. First, we show that if the sequence definition holds, then the epsilon-delta definition holds. We do this by contraposition, that is, by showing that if the epsilon-delta definition does <u>not</u> hold, then the sequence definition does not hold either.

If the epsilon-delta definition doesn't hold then there's some $\epsilon > 0$ such that no valid choice of δ exists; that is, for every $\delta > 0$ there's some $x \in S$ such that $|x - a| < \delta$, but $|f(x) - f(a)| \geq \epsilon$. For each $n \in \mathbb{N}^+$ (new notation I'm introducing for the natural numbers without zero), choose $\delta = \frac{1}{n}$ and set a_n equal to the corresponding $x \in S$ that's $\delta-$close to a but whose function value is <u>not</u> $\epsilon-$close to $f(a)$. Then $(a_n) \to a$ because $\lim \frac{1}{n} = 0$, so a_n gets arbitrarily close to a for sufficiently high n. However, $f(a_n)$ does not converge to $f(a)$ because $|f(a_n) - f(a)| < \epsilon$ is never true for any n. Therefore the sequence definition does not hold.

Second, we show that if the epsilon-delta definition holds, then the sequence definition holds. We do this by identifying δ as playing the usual role of ϵ in the convergence of a sequence. Pick an arbitrary sequence $(a_n) \to a$. If for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\forall x \in S, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$, then we can say there exists some N such that $n > N$ implies $|a_n - a| < \delta$. That is, we can eliminate δ in the definition:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in S, |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$
$$\forall \epsilon > 0, \exists N \in \mathbb{R} \text{ s.t. } n > N \implies |a_n - a| < \delta$$
$$\implies \forall \epsilon > 0, \exists N \in \mathbb{R} \text{ s.t. } n > N \implies |f(a_n) - f(a)| < \epsilon$$

What we've ended up with is the definition of the convergence of a sequence, which is what we wanted! □

Analogous to the idea that a sequence being Cauchy meant we didn't need to know what a sequence converged to in order to say it was convergent, we can define the idea of <u>uniform continuity</u>, a way of saying a function is continuous at some a without having to write down what $f(a)$ is.

> **Uniform continuity**
>
> $f : S \to \mathbb{R}$ is uniformly continuous on a subset $T \subseteq S$ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in T$,
>
> $$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

This isn't completely the same idea as the relationship between being Cauchy and convergent: in the case of the reals, uniform continuity is only the same as continuity if the subset T is a closed interval. (If we take T to be an open interval, we can define a unique extension of the function to the corresponding closed interval such that it is continuous.)

> **Equivalence of continuity and uniform continuity**
>
> Let $f : [a, b] \to \mathbb{R}$, where $[a, b]$ represents the set $\{r \in \mathbb{R} \mid a \leq r \leq b\}$. Then f is uniformly continuous (abbreviated u.c.) if and only if f is continuous.

Proof. The direction "u.c. implies continuous" can be proved by translating the u.c. condition to a statement about sequences. Take an arbitrary point $c \in [a, b]$. Suppose f is u.c.; then, for all $\epsilon > 0$, if x and c are $\delta-$close, we can look at the Cauchy sequence converging to x and say it approaches being $\delta-$close to c. Call this (x_n); then there exists some N such that $n > N \implies |x_n - c| < \delta \implies |f(x_n) - f(c)| < \epsilon$. This is the sequence definition of continuity, therefore a u.c. function must be continuous.

In the direction "continuous implies u.c.', suppose, towards a contradiction, that f is continuous but not u.c. Then there exists some $\epsilon > 0$ such that for every δ, there exist $x, y \in [a, b]$ s.t. $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$.

For all n, take $\delta = \frac{1}{n}$. Then, there exist $x_n, y_n \in [a, b]$ such that $|x_n - y_n| < \frac{1}{n}$, but $|f(x_n) - f(y_n)| \geq \epsilon$. (x_n) is bounded, so by the Bolzano-Weierstrass theorem, there exists (x_{n_k}) convergent. Let $z = \lim\limits_{k \to +\infty} x_{n_k}$. Then (y_{n_k}) converges to z as well because it's $\delta-$close to (x_{n_k}). $x_{n_k}, y_{n_k} \in [a, b] \forall k$ and $z \in [a, b]$. So f is continuous at z, which implies

$$\lim f(x_{n_k}) = f(z) = \lim f(y_{n_k})$$
$$\therefore \lim f(x_{n_k}) - f(y_{n_k}) = 0$$

But $|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon$ for all k, which is a contradiction. Therefore f must be uniformly continuous. $\qquad\square$