Ultrafilters and the Hyperreals

Lucas Bachmann

March 31, 2024

1 Introduction

The purpose of this article is to introduce a method of constructing number systems with specific properties. We'll introduce the concept of ultrafilters over sets and prove the existence of a special type of ultrafilter known as a non-principal ultrafilter (using the axiom of choice). We'll then build a hyperreal number system that satisfies two important properties:

- Every first-order sentences that is true in the reals also apply to the hyperreals,
- The hyperreals does not satisfy the Archimedean property.

Of course, as stated above, this result will depend on the axiom of choice, and so the resulting number system will be non-constructive.

This document will rely heavily on set notation, so reviewing this first is recommended if you're still unfamiliar with it. Furthermore, while I've tried to simplify aspects of this article, it is still relatively maths heavy, and might only be readable by first-year maths undergrads. This document was written for an article on luckys-corner.com titled "On the topic of 0.999... = 1".

2 Ultrafilters

We'll firstly cover the concept of ultrafilters over sets. We'll begin with it's looser cousin, the filter:

Definition 1. *Let S be a non-empty set. A filter U of S is a collection of subsets of S that satisfies the following properties:*

- ***S and $\emptyset$:*** *$S \in U$ and $\emptyset \notin U$.*

- ***Downwards Directed:*** *If $a, b \in U$, then $a \cap b \in U$.*
- ***Upwards Closed:*** *If $a \in U$ and $b \subseteq S$ such that $a \subseteq b$, then $b \in S$.*

Perhaps it's a little overwhelming, so let's look at a simple example.

Let $S = \mathbb{R}$. Let U be the set of subsets of S that contains $(0, 1)$ as a subset (so $(0, 1)$, $[0, 1]$, $\mathbb{R}_{>0}$ and $\mathbb{R}$ are elements of U, and $\{0\}, \mathbb{Z}$ and $\mathbb{R}/\{\frac{1}{2}\}$ are not elements of U). Then, we notice that U satisfies the following statements:

- **$\mathbb{R}$ and $\emptyset$:** $(0, 1) \subseteq \mathbb{R}$, so $\mathbb{R} \in U$. Similarly, $0 \nsubseteq \emptyset$, so $\emptyset \notin U$.
- **Downwards Directed:** Let $a, b \in U$. Since they both contain $(0, 1)$, therefore $(0, 1) \subseteq a \cap b$, thus $a \cap b \in U$.
- **Upwards Closed:** Let $a \in U$ and b be a subset of $\mathbb{R}$ with $a \subseteq b$. Then, since $(0, 1) \subseteq a$, therefore $(0, 1) \subseteq b$, thus $b \in U$.

So from this, we can conclude that U is a filter of $\mathbb{R}$.

In fact, our choice of set $(0, 1)$ was completely arbitrary. We could have used $[0, 1], \mathbb{Z}$ or any other subset of $\mathbb{R}$. In particular, we could have chosen a singleton set i.e. a set with only one element (e.g. $\{0\}$ or $\{1\}$). In this case, our filter will have an additional property: if U is the filter that contains all of the sets that contains the element 0, then for any subset s of $\mathbb{R}$, $s \in U$, or $s^c \in U$ (remember that $s^c = \mathbb{R}/s$). This additional property is what gives us ultrafilters:

Definition 2. *Let S be a set. An ultrafilter U of S is a maximal filter of S i.e. there doesn't exists a filter T of S such that U is a proper subset of T. Alternatively, an ultrafilter U of S is a filter that satisfies the dichotomy property: for all subsets s of S, either $s \in U$ or $s \notin U$ (but not both simultaneously).*

While the above technically defines ultrafilters in two different ways, they are in fact equivalent definitions. Of course, if both s and s^c were in U, then $s \cap s^c = \emptyset$ would also be in U, which is impossible.

Proof of equivalence of definition 2. Suppose first that U is a maximal filter of S. Suppose for a contradiction that there is a subset $s \subseteq S$ such that $s, s^c \notin U$. We construct the set T whose elements are elements of S expressible in the form $a \cap (s \cup t)$ for any $a \in U$ and $t \subseteq S$. Then, T satisfies the following properties:

- If there were any $a \in U$ such that $a \cap s = \emptyset$, then $a \subseteq s^c$, contradicting our assumption that $s^c \notin U$. Therefore, $a \cap s$ is always non-empty, and therefore $\emptyset \notin T$. Furthermore, since $S \in U$, thus $S \cap (s \cup S) = S \in T$.

- For $a \cap (s \cup t), b \cap (s \cup q) \in T$ with $a, b \in U$ and $t, q \subseteq S$, we get $(a \cap (s \cup t)) \cap (b \cap (s \cup q)) = (a \cap b) \cap ((s \cup t) \cap (s \cup q)) = (a \cap b) \cap (s \cup (t \cap q)) \in T$, so T is downwards directed.

- For $a \cap (s \cup t)$ with $a \in U$ and $t \subseteq S$, and $b \subseteq S$ with $a \cap (s \cup t) \subseteq b$, we get $b = (a \cup b) \cap (s \cup (t \cup b)) \in T$, so T is upwards closed.

Thus, T is a filter of S. However, U is a proper subset of T, since $s \in T$ but $s \notin U$, contradicting U being maximal. Thus, if U is a maximal filter of S, U must have the dichotomy property.

Conversely, suppose U is a filter of S with the dichotomy property. If T were another filter of S such that U was a proper subset of T, then T would have an element s which isn't in U. By the dichotomy property, this means $s^c \in U$, thus $s^c \in T$, therefore $s \cap s^c = \emptyset \in T$, a contradiction. Thus, any filter with the dichotomy property must be maximal. □

As I've pointed out before, the filter U of $\mathbb{R}$ whose elements are exactly the subsets of $\mathbb{R}$ that contain the element 0 is an ultrafilter, since U would satisfy the dichotomy property. In fact, we can generate a number of ultrafilters in this fashion: pick an element s of S, and generate the ultrafilter U that contains exactly the subsets of S that contains s. We give these ultrafilters a special name: the principal ultrafilters.

The next natural question to ask is whether there are any other ultrafilters. If there were, then importantly due to upwards closure, there wouldn't be any singletons i.e. sets with only one element. In fact, we can say even more: if an ultrafilter has a finite set $\{a_1, a_2, \ldots, a_n\}$ then it must also have a singleton set, otherwise by dichotomy, $\{a_1\}^c, \ldots, \{a_n\}^c$ are all in the ultrafilter, and the intersection of all of these sets with $\{a_1, \ldots, a_n\}$ is the empty set, meaning the empty set is in our ultrafilter, a contradiction. We can therefore conclude that if our ultrafilter is non-principled, it mustn't have any finite sets. This property is vital when we actually construct the hyperreal numbers.

Unfortunately, there isn't a constructive way of proving that non-principled ultrafilters for infinite sets exists. We'll need the axiom of choice, or more specifically, Zorn's Lemma:

Theorem 1 (Zorn's Lemma)**.** *Let S be a non-empty set of sets. If for every chain $\mathcal{C}$ of S the element $\cup\mathcal{C}$ is in S, then S has a maximal element.*

Naturally, the terminology is extremely confusing if this is the first time you're looking at this. I'll break it down here:

- A chain $\mathcal{C}$ is a non-empty collection of sets such that for any sets $a, b \in \mathcal{C}$, either $a = b$, $a \subseteq b$ or $b \subseteq a$. This is called a chain because the elements are ordered by the subset relation; in some cases, we could write $\mathcal{C}$'s elements as $a_1 \subseteq a_2 \subseteq a_3 \ldots$.
- The notation $\cup\mathcal{C}$ just means 'the union of all the sets in $\mathcal{C}$', or

$$\bigcup_{s \in \mathcal{C}} s.$$

- A maximal element m of S is a set such that no other set in S is a superset of m i.e. for all set $s \in S$ not equal to m, $m \nsubseteq s$.

All of this is quite fancy machinery, but the idea for the proof is quite simple. In order to construct our ultrafilter for some infinite set S, we'll do the following: we start with the basic filter that contains all co-finite sets (sets whose complement are all finite), and slowly add elements one at a time, making the filter bigger until it is maximal.

Proposition 1. *Let S be an infinite set. Then, there exists a non-principal ultrafilter of S.*

Proof. Let $\mathcal{U} := \{U \subseteq \mathcal{P}(S) : U \text{ is a filter that contains all co-finite sets}\}$. Clearly, the set of all co-finite subsets of S is a filter, therefore $\mathcal{U}$ is non-empty. We'll now show that $\mathcal{U}$ satisfies the precondition for Zorn's lemma; that is, for any chain $\mathcal{C}$ of $\mathcal{U}$, $\cup\mathcal{C} \in \mathcal{U}$.

Let $\mathcal{C} \subseteq \mathcal{U}$ be a chain. Clearly $\cup\mathcal{C}$ contains all co-finite sets, thus we just need to show $\cup\mathcal{C}$ is a filter.

- Since none of the sets in $\mathcal{C}$ contains the empty set, thus $\emptyset \notin \cup\mathcal{C}$. Similarly, since any set in $\mathcal{C}$ contains S, thus $S \in \cup\mathcal{C}$.
- Let $a, b \in \cup\mathcal{C}$. Then, $a \in A, b \in B$ for some filters $A, B \in \mathcal{C}$. Since either $A \subseteq B$ or $B \subseteq A$, one of A or B must contain both a and b. Let's suppose that $a, b \in A$. Then, since A is a filter, $a \cap b \in A$. Thus, $a \cap b \in \cup\mathcal{C}$, and so $\cup\mathcal{C}$ is downwards directed.
- Let $a \in \cup\mathcal{C}$, and $b \subseteq S$ such that $a \subseteq b$. Then, $a \in A$ for some $A \in \mathcal{C}$, therefore $b \in A$, thus $b \in \mathcal{C}$, and so $\cup\mathcal{C}$ is upwards closed.

Thus, $\cup\mathcal{C}$ is a filter, and so $\cup\mathcal{C} \in \mathcal{U}$.

We've just show that the precondition for Zorn's lemma is satisfied for $\mathcal{U}$. Therefore, the lemma applies, telling us that there is a maximal element U of $\mathcal{U}$. This directly implies U is an ultrafilter, and since U contains no singletons, U is non-principled. $\square$

We'll need the above result in order to build the hyperreal numbers.

Exercise 1. *Why does Proposition 1 need S to be an infinite set? What can we say about ultrafilters over finite sets?*

3 Hyperreals

We are now going to build a model for the hyperreal numbers. In particular, we'll use a technique using ultrafilters that will allow us to conserve first-order sentences.

To make our lives as easy as possible, we start first with a general strategy: we'll consider a mathematical structure that is big enough to embed the real numbers, and then equate different terms until we're left with the structure we want.

So, let's do this: we'll define $\mathcal{H}$ to be the set of countable sequence of real numbers. For example, $(1, 2, 3, 4, 5, \dots)$ and $(0, 0, 0, 0, 0, \dots)$ are both sequences of real number. In the future, we'll use the notation $(a_n)_n$ to be a real sequence such that the n^{th} element is a_n. For instance, $(0)_n$ is the sequence that is all zeroes, and $(n)_n$ is the sequence $1, 2, 3, \dots$.

There is a very natural embedding from $\mathbb{R}$ to $\mathcal{H}$ where we send x to $(x)_n$ i.e. the sequence that is all x. We could also define addition and multiplication component-wise i.e. $(a_n)_n + (b_n)_n = (a_n + b_n)_n$ and $(a_n)_n \cdot (b_n)_n = (a_n b_n)_n$.

However, there are immediately several problems: firstly, we want the hyperreal numbers to be a field, which means we have to define division over non-zero elements. However, our sequences could contain zeroes, in which case component-wise division doesn't work. Furthermore, we also want to order the elements of $\mathcal{H}$, and it's not exactly clear how one might order the two elements $(-1^n)_n$ and $(-1^{n+1})_n$

It would be convenient if, had we a sequence with some but not many zeroes, we could just 'ignore' them, and define division over the non-zero elements of the sequence. It would also be useful if we could define order component-wise, and just say that $(a_n)_n$ is greater than $(b_n)_n$ if $a_n > b_n$ for almost every n (but not necessarily all of them). To do this, we'll need some definition of 'almost everywhere', and as it turns out, an ultrafilter is perfect for this.

For the rest of this article, we are going to fix some non-principal ultrafilter U of $\mathbb{N}$. As discussed before, U doesn't contain any finite sets.

Definition 3. *The relation $\equiv$ is defined on $\mathcal{H}$ such that $(a_n)_n \equiv (b_n)_n$ if and only if the set $\{n \in \mathbb{N} : a_n = b_n\}$ is in U. From here on, we'll write $C((a_n)_n, (b_n)_n) = \{n \in \mathbb{N} : a_n = b_n\}$.*

We ideally want $\equiv$ to be an equivalence relation. Thankfully this is the case.

Proposition 2. *$\equiv$ is an equivalence relation.*

Proof. We check the three properties of an equivalence relation:

- **Transitive** Clearly, $(a_n)_n \equiv (a_n)_n$ since $C((a_n)_n, (a_n)_n)) = \mathbb{N} \in U$.
- **Symmetry** Since by definition, $C((a_n)_n, (b_n)_n) = C((b_n)_n, (a_n)_n)$, we therefore get $(a_n)_n \equiv (b_n)_n$ if and only if $(b_n)_n \equiv (a_n)_n$.
- **Transitivity** Let $(a_n)_n \equiv (b_n)_n$ and $(b_n)_n \equiv (c_n)_n$. If $n \in C((a_n)_n, (b_n)_n) \cap C((b_n)_n, (c_n)_n)$, then $a_n = b_n$ and $b_n = c_n$, thus $a_n = c_n$ or $n \in C((a_n)_n, (c_n)_n)$. Therefore, since $C((a_n)_n, (b_n)_n), C((b_n)_n, (c_n)_n) \in U$, and $C((a_n)_n, (b_n)_n) \cap C((b_n)_n, (c_n)_n) \subseteq C((a_n)_n, (c_n)_n)$, thus by the property of filters, $C((a_n)_n, (c_n)_n) \in U$, or $(a_n)_n \equiv (c_n)_n$.

$\square$

Perfect, so $\equiv$ is an equivalence relation. In particular, we have the following:

Definition 4. *Define $[a]$ to be the equivalence class of a under the relation $\equiv$ (i.e. the set of elements b of $\mathcal{H}$ such that $a \equiv b$). Let ${}^*\mathbb{R}$ be the set of equivalence classes of $\equiv$. For any class $c \in {}^*\mathbb{R}$, if $a \in c$, we say that a is a class representative of c, and $c = [a]$.*

Next, we want to define our ordered field operations and relations.

Definition 5. *We define the following operations and relations on ${}^*\mathbb{R}$:*

- *(Addition of classes)* $[(a_n)_n] + [(b_n)_n] = [(a_n + b_n)_n]$,
- *(Multiplication of classes)* $[(a_n)_n] \cdot [(b_n)_n] = [(a_n \cdot b_n)_n]$,

- *(Additive inverse)* $-[(a_n)_n] = [(-a_n)_n]$,
- *(Multiplicative inverse) If* $c \in {}^*\mathbb{R}$ *with* $c \neq [0]$, *and* $a_n \in c$ *with* $a_n \neq 0$ *for all* n, *then* $c^{-1} = [(a_n^{-1})_n]$,
- *(Order)* $[(a_n)_n] < [(b_n)_n]$ *if and only if the set* $\{n \in \mathbb{N} : a_n < b_n\} \in U$ *(We'll write* $C_<((a_n)_n, (b_n)_n) = \{n \in \mathbb{N} : a_n < b_n\}$*).*

These definitions rely on our choice of class representative, which means they might not be valid (i.e. if $[a] = [c]$, we might not have the case that $[a+b] = [c+b]$). We will now prove that these operations are in fact perfectly fine.

Proposition 3. *The previously defined operations on* ${}^*\mathbb{R}$ *are well-defined.*

Proof.

- (Addition of classes) Let $[(a_n)_n] = [(c_n)_n]$ and $[(b_n)_n] = [(d_n)_n]$. We want to show $[(a_n)_n + (b_n)_n] = [(c_n)_n + (d_n)_n]$.

  If $i \in C((a_n)_n, (c_n)_n) \cap C((b_n)_n, (d_n)_n)$, we get $a_i = c_i$ and $b_i = d_i$, thus $a_i + b_i = c_i + d_i$, thus $i \in C((a_n + b_n)_n, (c_n + d_n)_n)$. Therefore, $C((a_n)_n, (c_n)_n) \cap C((b_n)_n, (d_n)_n) \subseteq C((a_n + c_n)_n, (b_n + d_n)_n) \in U$, which gives us $[(a_n)_n + (b_n)_n] = [(c_n)_n + (d_n)_n]$ as required.

- (Multiplication of classes) The proof is similar to the previous case, with the addition symbol $+$ replaced with the multiplication symbol $\cdot$.

- (Additive inverse) Let $[(a_n)_n] = [(b_n)_n]$. Then, since $C((-a_n)_n, (-b_n)_n) = C((a_n)_n, (b_n)_n)$, we get $[(-a_n)_n] = ([(-b_n)_n]$.

- (Multiplicative inverse) Firstly, we need to check if $c \in {}^*\mathbb{R}$ has a non-zero sequence if $c \neq [0]$. Let $(a_n)_n$ be any sequence in c. Let $(b_n)_n$ be the real sequence such that $b_i = a_i$ whenever $a_i \neq 0$, and $b_i = 1$ otherwise. Then, since $C((a_n)_n, (0)_n) \notin U$, thus $C((a_n)_n, (b_n)_n) = C((a_n)_n, (0)_n)^c \in U$.

  Then, if for some class $c \neq [0]$, we have non-zero sequences $(a_n)_n, (b_n)_n \in c$, then since $C((a_n^{-1})_n, (b_n^{-1})_n) = C((a_n)_n, (b_n)_n) \in U$, we get $[(a_n^{-1})_n] = [(b_n^{-1})_n]$.

- (Order) Let $[(a_n)_n] = [(c_n)_n]$ and $[(b_n)_n] = [(d_n)_n]$, and $C_<((a_n)_n, (b_n)_n) \in U$. Then, if $i \in C((a_n)_n, (c_n)_n) \cap C_<((a_n)_n, (b_n)_n) \cap C((b_n)_n, (d_n)_n)$, then $c_i = a_i < b_i = d_i$, thus $C((a_n)_n, (c_n)_n) \cap C_<((a_n)_n, (b_n)_n) \cap C((b_n)_n, (d_n)_n) \subseteq C_<((c_n)_n, (d_n)_n) \in U$.

$\square$

The above demonstrates that our operators on ${}^*\mathbb{R}$ are well defined. Of course, since we want ${}^*\mathbb{R}$ to be an ordered field, we also need to check that our above operations satisfy the ordered field axioms. Most of the axioms are straightforward enough to prove, and so I'll omit them for this article (but feel free to attempt the proof on your own). This is perhaps the only interesting case:

Proposition 4. $<$ *is a total order.*

Proof. We'll need to prove that $<$ is transitive, and satisfies the trichotomy statement i.e. for any classes $c, d \in {}^*\mathbb{R}$, exactly one of the following is true: $c < d$, $d < c$ or $c = d$.

First with transitivity: let $[(a_n)_n], [(b_n)_n], [(c_n)_n] \in {}^*\mathbb{R}$ with $[(a_n)_n] < [(b_n)_n]$ and $[(b_n)_n] < [(c_n)_n]$. By transitivity of $<$, we necessarily get that $C_<((a_n)_n, (b_n)_n) \cap C_<((b_n)_n, (c_n)_n) \subseteq C_<((a_n)_n, (c_n)_n) \in U$, thus $[(a_n)_n] < [(c_n)_n]$.

Second, with trichotomy: let $[(a_n)_n], [(b_n)_n] \in {}^*\mathbb{R}$. We get by the definitions of C and $C_<$ that $C((a_n)_n, (b_n)_n)$, $C_<((a_n)_n, (b_n)_n)$ and $C_<((b_n)_n, (a_n)_n)$ partitions $\mathbb{N}$; these are three mutually disjoint sets whose union is $\mathbb{N}$. Since they are mutually disjoint, it is definitely not possible for two of them to be in U, since downward directedness would imply $\emptyset \in U$. If none of $C((a_n)_n, (b_n)_n)$, $C_<((a_n)_n, (b_n)_n)$ or $C_<((b_n)_n, (a_n)_n)$ were in U, then by maximality, their complement would be in U, and so by downward closure, $C((a_n)_n, (b_n)_n)^c \cap C_<((a_n)_n, (b_n)_n)^c \cap C_<((b_n)_n, (a_n)_n)^c \in U$, the right hand side being equal to the empty set. This therefore means that exactly one of $C((a_n)_n, (b_n)_n)$, $C_<((a_n)_n, (b_n)_n)$ or $C_<((b_n)_n, (a_n)_n)$ must be in U, which means exactly one of the following is true: $[(a_n)_n] = [(b_n)_n]$, $[(a_n)_n] < [(b_n)_n]$ or $[(b_n)_n] < [(a_n)_n]$. $\square$

So, we've successfully constructed an ordered field ${}^*\mathbb{R}$. Furthermore, we can embed the real number $\mathbb{R}$ into ${}^*\mathbb{R}$ by sending $r \in \mathbb{R}$ to the class $[(r)_n] \in {}^*\mathbb{R}$. One thing that's left for us to check is whether or not ${}^*\mathbb{R}$ satisfies our notion of the hyperreal numbers: namely, does ${}^*\mathbb{R}$ have elements greater than any elements in $\mathbb{R}$ (or more precisely, the image of $\mathbb{R}$ under our embedding). The answer is yes.

Consider the sequence $(n)_n$. Let $r \in \mathbb{R}$. There exists some positive integer m such that $r < m$ (since $\mathbb{R}$ satisfies the Archimedean principle). Then, $\{m, m+1, m+2, \dots\} \subseteq C_<((r)_n, (n)_n)$, and so since the set $\{m, m+1, m+2, \dots\}$ is co-finite (i.e. its complement $\{1, 2, \dots, m-1\}$ is finite), this set is in U (since U is non-principal). Therefore, for all real numbers r, $[(r)_n] < [(n)_n]$, so ${}^*\mathbb{R}$ is non-Archimedean.

Exercise 2. *Throughout this section, we have taken U to be a non-principal ultrafilter of $\mathbb{N}$. What would happen if we instead had U be principal? Would we still be able do define $+$ and $\cdot$ on $^*\mathbb{R}$? Do we lose some important property?*

4 Transfer Principal

If you've read and understood everything so far, congratulations. This last part is going to be a lot more technical and laborious than any other section.

We've constructed the hyperreal numbers and showed that it is non-Archimedean. Cool. Of course, we should ask ourselves what have we kept. The answer is quite a lot; any **first-order sentences** that are true in the real numbers are also true in the hyperreal numbers, and vice versa. This is known as the 'transfer principle'.

The key phrase in the above statement is "first-order sentences". What exactly does that mean? Informally, a first-order statement is one that is constructed using:

- Variables, like $x, y, \ldots$,
- Constants, like $0, 1, \ldots$,
- Operators, like $+, \cdot, \ldots$,
- Propositions, like $=, <, \ldots$,
- Boolean Operators, namely $\wedge$, $\vee$, $\rightarrow$ and $\neg$,
- Quantifiers, namely the universal quantifier $\forall$ and the existential quantifier $\exists$.

For instance, we can write the statement "0 is an additive identity" as follows: $\forall x, 0 + x = x$. We can also write the statement "There is a multiplicative inverse for any non-zero number" as follows: $\forall x, \neg x = 0 \rightarrow (\exists y, x \cdot y = 1)$.

The language of first-order logic is quite powerful, and can be used to express a lot of different statements. However, they aren't capable of expressing everything: for instance, we cannot characterise the Archimedean property using first-order statements.

In any case, we want to now show that any first-order sentence that is true in the real numbers is also true in the hyperreal numbers. Keen eyed readers might have noticed that I oscillate between the words "first-order statements" and "first-order sentences". This isn't by accident; a first-order sentence is simply a first-order statement that has all of its variables bounded by some quantifier i.e. if a variable x appears in the sentence, then at some point earlier on, there was a quantifier $\forall x$ or $\exists x$ that 'bound' the variable. A variable that isn't bound is called a free variable.

Alright. In order to prove the transfer principal, we'll first formally define what a first-order statement is.

Definition 6. *A term in variables $v_1, \ldots, v_n$ is an expression that is built using the following rules:*

- *A variable v_i is a term for all i,*
- *A constant is a term,*
- *For any function $f(x_1, \ldots, x_m)$ and terms $t_1, \ldots, t_m$, $f(t_1, \ldots, t_m)$ is a term.*

The above expression just allows us to formally describe expressions. For instance, $-x_2 \cdot (1 + x_1)^{-1}$ is a term in variables x_1, x_2, since

- x_1 is a term,
- 1 is a a term,
- $1 + x_1$ is a term,
- $(1 + x_1)^{-1}$ is a term,
- x_2 is a term,
- $-x_2$ is a term,
- $-x_2 \cdot (1 + x_1)^{-1}$ is a term.

where each term is sequentially built using our rules and previously constructed term. You can, for now, just think of terms as an unambiguous expressions using variables, constants, and functions.

Next, we need to define statements:

Definition 7. *A statement in variables $v_1, \ldots, v_n$ is an expression that is built using the following rules:*

- *For any propositions $p(x_1, \ldots, x_m)$ and terms $t_1, \ldots, t_m$, $p(t_1, \ldots, t_m)$ is a statement (this is known as an atomic statement),*

- *If s_1, s_2 are statements, then so is $s_1 \vee s_2$, $s_1 \wedge s_2$, $s_1 \rightarrow s_2$, and $\neg s_1$,*
- *If t is a statements in variables $x, v_1, \ldots, v_n$, then $\exists x.t$ and $\forall x.t$ are statements in variables $v_1, \ldots, v_n$.*

To clarify, propositions are statements that take in values and return true or false. For instance, equality is a proposition: if x_1 and x_2 are variables, then $x_1 = x_2$ is an atomic statement.

The point of showing these two definitions is because I want to show that first-order statements are built starting with small simple components (the atomic statements), and then combined using logical operations to create more complex statements. In other words, the definition of terms and statements are recursive. The reason to show you this is because, when we eventually write proofs on these statements, we'll use recursion.

For this paper, we should specify the symbols we'll be using.

- We have two constants, 0 and 1. In $\mathbb{R}$, these constants have their usual value. In ${}^*\mathbb{R}$, $0 = [(0)_n]$ and $1 = [(1)_n]$.
- We have four functions we can use to build terms: $t_1 + t_2$, $t_1 \cdot t_2$, $-t_1$ and t_1^{-1} for any terms t_1, t_2.
- We have two propositions we can use to build logical statements: $t1 < t_2$ and $t_1 = t_2$ for any terms t_1, t_2.

Now that that's out of the way, how might one prove the transfer principle. It's actually kind of hard to prove it directly. However, there is an intermediary result that is surprisingly powerful, and relatively simple to understand. In order to do so, we'll first need to generalise the functions C and $C_<$ that we've been using the previous section:

Definition 8. *Let $p(v_1, \ldots, v_n)$ be a statement with variables $v_1, \ldots, v_n$, and let $(x_{1,n})_n, \ldots, (x_{m,n})_n \in \mathcal{H}$. We define $C_p((x_{1,n})_n, \ldots, (x_{m,n})_n = \{n \in \mathbb{N} : p(x_{1,n}, \ldots, x_{m,n})$ is true$\}$. In other words, for some values $(x_{1,n})_n, \ldots, (x_{m,n})_n \in \mathcal{H}$, the function C_p evaluates p over these sequences component-wise, and returns a set containing the indexes where p is true.*

So, for instance, if our proposition $p(v_1)$ is "$v_1 + 1 = 1$", then $C_p((0)_n) = \mathbb{N}$, $C_p((1)_n) = \emptyset$, and $C_p((n-1)_n) = \{0\}$.

Proposition 5. *For any statement $s(v_1, \ldots, v_n)$ with variables $v_1, \ldots, v_n$ and $[x_1], \ldots, [x_n] \in {}^*\mathbb{R}$, $s([x_1], \ldots, [x_n])$ is true in ${}^*\mathbb{R}$ if and only if $C_s(x_1, \ldots, x_n) \in U$.*

Proof. We'll proceed by induction on the structure of s. We can actually reduce the number of cases we need to consider, since we technically can write any logical statements using only the logical symbols $\neg$, $\wedge$, and $\forall$ (prove it!).

- s is an atomic statement: We've already proved this; this case just boils down to showing $=$ and $<$ behave as we expect.

- $s = \neg s'$ for some statement s': Suppose s' satisfy our hypothesis. Then, $s([x_1], \ldots, [x_n])$ is true $\Leftrightarrow s'([x_1], \ldots, [x_n])$ is false $\Leftrightarrow C_{s'}(x_1, \ldots, x_n) \notin U \Leftrightarrow C_s(x_1, \ldots, x_n) = \mathbb{N}/C_{s'}(x_1, \ldots, x_n) \in U$.

- $s = s_1 \wedge s_2$ for statements s_1, s_2: Suppose s_1 and s_2 satisfy our hypothesis. Then, $s([x_1], \ldots, [x_n])$ is true $\Leftrightarrow$ both $s_1([x_1], \ldots, [x_n])$ and $s_2([x_1], \ldots, [x_n])$ are true $\Leftrightarrow C_{s_1}(x_1, \ldots, x_n) \in U$ and $C_{s_2}(x_1, \ldots, x_n) \in U \Leftrightarrow C_s(x_1, \ldots, x_n) \in U$ (the final equivalence is true due to the fact that $C_s(x_1, \ldots, x_n) = C_{s_1}(x_1, \ldots, x_n) \cap C_{s_2}(x_1, \ldots, x_n)$ and because U is an ultrafilter).

- $s = \forall y.s'$ for some statement s' with variables $y, v_1, \ldots, v_n$: suppose s' satisfy our hypothesis. Let $x_i = (a_{i,n})_n$ for $i = 1, \ldots, m$. Suppose $s([x_1], \ldots, [x_n])$ is true. Then, for all $[(b_n)_n] \in {}^*\mathbb{R}$, we have $s'([(b_n)_n], [x_1], \ldots, [x_n])$ is true, thus $C_{s'}((b_n)_n, x_1, \ldots, x_n) \in U$ for any choice of $(b_n)_n \in \mathcal{H}$.

  Now, suppose for a contradiction that $C_s(x_1, \ldots, x_n) \notin U$. This means that for every $i \notin C_s(x_1, \ldots, x_n)$, we have the sentence $\forall y.s'(y, a_{1,i}, \ldots, a_{n,i})$ is false in $\mathbb{R}$, therefore there exists some $b_i \in \mathbb{R}$ such that $s'(b_i, a_{1,i}, \ldots, a_{n,i})$ is false. Therefore, if we take all such b_i, and extend this to some sequence $(b_n)_n$ in $\mathcal{H}$ (by setting all of the missing values to some arbitrary number, say 0), then we get $C_{s'}((b_n)_n, x_1, \ldots, x_n) = C_s(x_1, \ldots, x_n) \notin U$, a contradiction. Thus, we must have $C_s(x_1, \ldots, x_n) \in U$.

  Conversely, suppose $C_s(x_1, \ldots, x_n) \in U$. Then, for all choices $[(b_n)_n] \in {}^*\mathbb{R}$, we get $C_{s'}((b_n)_n, x_1, \ldots, x_n) \supseteq C_s(x_1, \ldots, x_n)$, thus $C_{s'}((b_n)_n, x_1, \ldots, x_n) \in U$, thus by our hypothesis, $s'([(b_n)_n], [x_1], \ldots, [x_n])$ is true for any choice of $(b_n)_n \in \mathcal{H}$, thus $s([x_1], \ldots, [x_n])$ is true.

$\square$

This is by far the hardest part of this entire exercise. Now comes the easy part:

Proposition 6. *(The Transfer Principle) The function* $f : \mathbb{R} \to {}^*\mathbb{R}$ *sending* $r \in \mathbb{R}$ *to* $[(r)_n]$ *is an elementary embedding i.e. for all statements* s *with variables* $v_1, \ldots, v_n$*,* $s(f(r_1), \ldots, f(r_n))$ *is true in* ${}^*\mathbb{R}$ *if and only if* $s(r_1, \ldots, r_n)$ *is true in* $\mathbb{R}$*. In particular, every sentence (i.e. statements with zero free variables) that is true in* $\mathbb{R}$ *is true in* ${}^*\mathbb{R}$ *and vice versa.*

Proof. Let s be a statement with variables $v_1, \ldots, v_n$, and $r_1, \ldots, r_n \in \mathbb{R}$. If $s(f(r_1), \ldots, f(r_n))$ is true, then $C_s((r_1)_m, \ldots, (r_n)_m) \in U$, thus $C_s((r_1)_m, \ldots, (r_n)_m)$ is non-empty. Therefore, by the definition of C_s, we must have $s(r_1, \ldots, r_n)$ is true.

Conversely, if $s(r_1, \ldots, r_n)$ is true, then $C_s((r_1)_m, \ldots, (r_n)_m) = \mathbb{N} \in U$, therefore $s(f(r_1), \ldots, f(r_n))$ is true. □

5 Final Thoughts

If you made it to this part of the document, well done!

If you enjoy this type of mathematics, you will certainly be interested in looking into Model Theory. If you do, you might find another method of proving the existence of a hyperreal field using Upward Löwenheim-Skolem. While this prove will be even less constructive, it is much easier and faster that all of the things I've done in this document.