Infinitesimal analysis

Also known as nonstandard analysis

David Pierce

September 1, 2015

Matematik Bölümü
Mimar Sinan Güzel Sanatlar Üniversitesi
mat.msgsu.edu.tr/~dpierce/

These notes are based on my course, officially called Rudiments of Nonstandard Analysis, given August 10–22, 2015, at the Nesin Matematik Köyü, Kayser Dağı Mevkii, Şirince, Selçuk, İzmir, Turkey. I gave similar courses in earlier years. Before the course of 2009, I typed up notes on everything that I might possibly want to talk about, if I had time. I revised the notes before the 2010 course. In 2014, after each lecture, I typed up notes of what I actually *had* talked about; later I tried to polish them. I used these notes in 2015, but did not follow them carefully. I created the present notes as I did the 2014 notes, though with (perhaps even more) embellishments, omissions, and rearrangements.

Contents

1 Preliminary discussion

Throughout these notes, the symbols $\mathbb{N}$ and ω represent what I would call the *number-theorist's natural numbers* and the *set-theorist's natural numbers,* respectively:

$$\mathbb{N} = \{1, 2, 3, \dots\}, \qquad \omega = \{0, 1, 2, \dots\}.$$

If the proof of a theorem is not supplied, it is an exercise for the reader.

High-school algebra is largely a study of the properties of arbitrary *ordered fields,* such as the ordered field $\mathbb{Q}$ of *rational numbers* or the ordered field $\mathbb{R}$ of *real numbers.* Calculus is a study of $\mathbb{R}$ as a *complete* ordered field. Here **completeness** is the property whereby every nonempty set of numbers with an upper bound has a *least upper bound,* or **supremum.**

Suppose f is a function from $\mathbb{R}$ to itself (that is, $f\colon \mathbb{R} \to \mathbb{R}$), and a and L are in $\mathbb{R}$. We say f has the **limit** L at a, and we may write either of

$$\lim_{x\to a} f(x) = L, \qquad\qquad \lim_a f = L, \tag{1}$$

provided that $f(x)$ is close to L whenever x is close to, but not equal to, a. This is vague. By the "standard" definition, (1) means we can make $f(x)$ *arbitrarily* close to L by making x *sufficiently* close to (though not equal to) a. Here "arbitrarily close" means "within ε, for any positive ε given to us." Then "sufficiently close" means "within δ, where δ is positive and depends on ε." More precisely then, the "standard" definition of (1) is that, for all positive ε in $\mathbb{R}$,[1] for some positive δ in $\mathbb{R}$,[2] for all x in $\mathbb{R}$, if $0 < |x - a| < \delta$, then $|f(x) - L| < \varepsilon$. Even more symbolically, the definition is,

$$\forall\varepsilon \Big(\varepsilon > 0 \to \exists\delta \big(\delta > 0 \land \forall x \\ \big(0 < |x-a| \land |x-a| < \delta \to |f(x) - L| < \varepsilon\big)\big)\Big),$$

which can also be written as

[1] I do not say "for all $\varepsilon > 0$," because I prefer to treat the expression $\varepsilon > 0$ always as a *sentence* ("ε is greater than 0"), not as a noun phrase ("ε that are greater than 0" or "ε greater than 0").

[2] Alternatively, one may say, "there exists a positive δ in $\mathbb{R}$ such that…"

$$\forall\varepsilon\ \exists\delta\ \forall x\ \Big(\varepsilon > 0 \to \delta > 0 \land \\ \big(0 < |x-a| \land |x-a| < \delta \to |f(x) - L| < \varepsilon\big)\Big). \quad (2)$$

These two expressions are equivalent sentences of *first-order logic,* in which the letters ε, δ, and x are *bound variables.* One could use other letters as bound variables instead; but ε and δ are almost always the letters used, as an aid to the memory, because the sentences are so logically complicated. It is complicated to have *three* blocks of quantifiers, as shown at the head of (2); most definitions in algebra require at most two, as in the axiom ensuring existence of multiplicative inverses in fields,

$$\forall x\ \exists y\ (x \neq 0 \to xy = 1).$$

The axiom of associativity of multiplication,

$$\forall x\ \forall y\ \forall z\ x(yz) = (xy)z,$$

has three quantifiers, but they form one block: they are all universal.

Infinitesimal or **"nonstandard" analysis** provides a way to define limits without using any such letters as ε and δ. By the "nonstandard" definition, $\lim_a f = L$, provided that, for all x, if x is *infinitesimally close* to a, without being equal to a, then $f(x)$ is infinitesimally close to L. Symbolically, the definition is

$$\forall x\ (x \simeq a \land x \neq a \to f(x) \simeq L). \quad (3)$$

This is evidently much simpler than (2), although, as we shall see below, there is a quantifier in the precise definition of $a \simeq x$ and $f(x) \simeq L$.

2 Non-Archimedean fields

The adjective "infinitesimal" is obtained from the adjective "infinite" by adding a Latin ending, which corresponds to the English ordinal ending "-th" seen in "fourth," "fifth," "sixth," and so on. The English ordinals are used also to denote fractional parts: one fourth, or the fourth part, is one of four equal parts, or 1/4. Then an infinitesimal part is an "infinitieth" part, or $1/\infty$.

To be precise then, a number a is **infinitesimal** if, for all n in $\mathbb{N}$,

$$|a| < \frac{1}{n}.$$

A number b is **infinite** if, for all n in $\mathbb{N}$,

$$n < |b|.$$

Then b is infinite if and only if $b \neq 0$ and $1/b$ is infinitesimal. The "nonstandard" definition (3) can be spelled out as

$$\forall x \ \left(\forall n \left(n \in \mathbb{N} \to |x - a| < \frac{1}{n}\right) \wedge x \neq a \to \right.$$
$$\left. \forall n \left(n \in \mathbb{N} \to |f(x) - L| < \frac{1}{n}\right)\right)$$

and hence as

$$\forall x \ \exists m \ \forall n \ \left(\left(m \in \mathbb{N} \wedge |x - a| > \frac{1}{m}\right) \vee \right.$$
$$\left. \left(x \neq a \wedge |f(x) - L| < \frac{1}{n}\right)\right). \qquad (4)$$

Thus there are still three blocks of quantifiers. But again, two of them can be hidden, as in (3); this cannot be done in (2). This makes the "nonstandard" definition of limit simpler than the "standard."

However, there is a new complication with the "nonstandard" definition. The only infinitesimal in $\mathbb{R}$ is 0, and $\mathbb{R}$ contains no infinite numbers. This follows from the theorem below. Thus if the variable x must always represent a real number, then (4) is always true. So we shall have to allow x to be a "nonstandard" number, in a sense that we shall develop.

An ordered field is **Archimedean** if, for every element c of the field, for some n in $\mathbb{N}$,

$$|c| \leqslant n.$$

That is, an ordered field is Archimedean *precisely* when none of its elements is infinite.

Theorem 1. $\mathbb{R}$ *is Archimedean.*

Proof. If $\mathbb{R}$ had an infinite element, then it would have a positive infinite element, and this would be an upper bound of $\mathbb{N}$. Then $\mathbb{N}$ would have a supremum, say d. In this case, $d-1$ would not be an upper bound of $\mathbb{N}$, so $d-1<n$ for some n in $\mathbb{N}$, and then $d<n+1$, which means d is not an upper bound of $\mathbb{N}$ either, since $n+1 \in \mathbb{N}$. $\square$

There *are* ordered fields that contain infinite and (therefore) nonzero infinitesimal elements. (Note that containing infinite elements is not the same as containing infinitely *many* elements.) For one example, first we let

$$\mathbb{R}[X] = \big\{a_0 + a_1X + a_2X^2 + \cdots + a_nX^n : \\ n \in \omega \ \& \ \{a_0, \dots, a_n\} \subseteq \mathbb{R}\big\}.$$

Here $a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$ is also written as

$$\sum_{k=0}^{n} a_kX^k;$$

it is a **polynomial** in the variable X over $\mathbb{R}$, and if $a_n \neq 0$, then a_n is the **leading coefficient** of the polynomial. Also 0 is the leading coefficient of the polynomial 0. Then $\mathbb{R}[X]$ is a **ring** (that is, a commutative unital ring: it has associative, commutative operations of addition and multiplication; multiplication distributes over addition; there are additive and multiplicative identities; and there are additive inverses). We can form fractions with the elements of $\mathbb{R}[X]$, just as the rational numbers are formed from the integers. Thus we define

$$\mathbb{R}(X) = \left\{ \frac{f}{g} \colon f \in \mathbb{R}[X] \ \ \& \ \ g \in \mathbb{R}[X] \smallsetminus \{0\} \right\},$$

which is the field of **rational functions** in X over $\mathbb{R}$. Here f/g is the equivalence class of the ordered pair (f, g) with respect to the equivalence relation $\sim$ given by

$$(f, g) \sim (h, k) \iff fk = hg.$$

Theorem 2. $\mathbb{R}(X)$ *is an ordered field with respect to the relation* $<$ *defined by*

$$\frac{f}{g} > 0 \iff \frac{a}{b} > 0,$$

where $f \in \mathbb{R}[X]$ *and* $g \in \mathbb{R}[X] \smallsetminus \{0\}$, *and* f *and* g *have leading coefficients* a *and* b *respectively. In the ordered field,* X *is positive and infinite.*

3 Ultrafilters

We shall denote by

$$\mathbb{R}^\omega$$

the set of functions from ω to R, that is, the set of real-valued sequences on ω. The same such sequence can be denoted by any of the expressions

$$(a_0, a_1, a_2, \dots), \qquad (a_k \colon k \in \omega), \qquad a.$$

These sequences can be added, subtracted, and multiplied term by term:

$$a \pm b = (a_k \pm b_k \colon k \in \omega), \qquad a \cdot b = (a_k \cdot b_k \colon k \in \omega).$$

Then $\mathbb{R}^\omega$ is a ring, like $\mathbb{R}[X]$. For each n in ω, there is an element χ^n of $\mathbb{R}^\omega$, where

$${\chi^n}_m = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{if } n \neq m. \end{cases}$$

Then the set $\{\chi^n \colon n \in \omega\}$ generates a proper ideal of $\mathbb{R}^\omega$. Let this ideal be included in a maximal ideal M: this exists by the *Maximal Ideal Theorem* (see page 35). We shall do infinitesimal analysis in the field

$$\mathbb{R}^\omega/M.$$

If $a \in \mathbb{R}^\omega$, we define

$$\operatorname{supp}(a) = \{k \in \omega \colon a_k \neq 0\};$$

this is the **support** of a. Thus $\operatorname{supp} \colon \mathbb{R}^\omega \to \mathscr{P}(\omega)$. Like any function, the support function induces two additional functions, as follows. If $A \subseteq \mathbb{R}^\omega$ and $B \subseteq \mathscr{P}(\omega)$, we have

$$\operatorname{supp}[A] = \{\operatorname{supp}(x) \colon x \in A\},$$
$$\operatorname{supp}^{-1}(B) = \{x \in \mathbb{R}^\omega \colon \operatorname{supp}(x) \in B\}.$$

Then

$$A \subseteq \operatorname{supp}^{-1}(\operatorname{supp}[A]), \qquad B = \operatorname{supp}[\operatorname{supp}^{-1}(B)].$$

The ideal of $\mathbb{R}^\omega$ generated by $\{\chi^n \colon n \in \omega\}$ consists of the functions from ω to $\mathbb{R}$ having finite support; also, if this ideal is I, then $\operatorname{supp}[I]$ is the set of all finite subsets of ω.

Theorem 3. *A subset I of $\mathbb{R}^\omega$ is an ideal of $\mathbb{R}^\omega$ if and only if*

$$\operatorname{supp}^{-1}(\operatorname{supp}[I]) = I \tag{5}$$

and, for all X and Y in $\mathscr{P}(\omega)$,
(1) $\varnothing \in \operatorname{supp}[I]$,
(2) $X, Y \in \operatorname{supp}[I] \implies X \cup Y \in \operatorname{supp}[I]$,
(3) $X \subseteq Y \ \& \ Y \in \operatorname{supp}[I] \implies X \in \operatorname{supp}[I]$.
If I and J are distinct ideals of R^ω, then

$$\operatorname{supp}[I] \neq \operatorname{supp}[J].$$

A subset I of $\mathbb{R}^\omega$ is a maximal ideal of $\mathbb{R}^\omega$ if and only if (5) *and, for all X and Y in $\mathscr{P}(\omega)$,*
(1) $\varnothing \in \operatorname{supp}[I]$ *and* $\omega \notin \operatorname{supp}[I]$,
(2) $X, Y \in \operatorname{supp}[I] \implies X \cup Y \in \operatorname{supp}[I]$,
(3) $X \notin \operatorname{supp}[I] \implies \omega \smallsetminus X \in \operatorname{supp}[I]$.
A maximal ideal M of $\mathbb{R}^\omega$ is nonprincipal if and only if $\operatorname{supp}[M]$ *contains every finite subset of ω.*

It may be observed that $\mathscr{P}(\omega)$ is a ring when the **sum** of two elements is defined as their symmetric difference, and their **product** as their intersection:

$$\begin{aligned} X + Y &= X \bigtriangleup Y = (X \smallsetminus Y) \cup (Y \smallsetminus X), \\ X \cdot Y &= X \cap Y. \end{aligned}$$

In this case

$$0 = \varnothing, \qquad 1 = \omega, \qquad -X = X.$$

The ring $\mathscr{P}(\omega)$ is then a **Boolean ring** because it satisfies

$$\forall x\; x^2 = x. \tag{6}$$

From this axiom (along with the ring axioms), it follows that

$$\forall x \; 2x = 0, \tag{7}$$

or $\forall x \; -x = x$. Indeed, from (6) we have $2x = (2x)^2 = 4x^2 = 4x$, and then (7). Now part of the last theorem can be formulated as follows.

Theorem 4. *Let $I \subseteq \mathscr{P}(\omega)$.*

1. *I is an ideal of $\mathscr{P}(\omega)$ if and only if* $\operatorname{supp}^{-1}[I]$ *is an ideal of $\mathbb{R}^\omega$.*
2. *I is a maximal ideal of $\mathscr{P}(\omega)$ if and only if* $\operatorname{supp}^{-1}[I]$ *is a maximal ideal of $\mathbb{R}^\omega$.*

But it will be more useful to think of ideals of $\mathscr{P}(\omega)$ in the terms of Theorem 3, rather than in the usual terms of ring theory. In fact, it will be useful to replace the notion of an ideal with a "dual" notion, as follows. A subset F of $\mathscr{P}(\omega)$ is a **filter of** $\mathscr{P}(\omega)$ and a **filter on** ω if for all X and Y in $\mathscr{P}(\omega)$,

(1) $\omega \in F$,
(2) $X, Y \in F \implies X \cap Y \in F$,
(3) $X \in F \;\&\; X \subseteq Y \implies Y \in F$.

Thus I is an ideal of $\mathscr{P}(\omega)$ if and only if $\{\omega \smallsetminus X \colon X \in I\}$ is a filter on ω. A subset $\mathscr{U}$ of $\mathscr{P}(\omega)$ is an **ultrafilter on** ω if for all X and Y in $\mathscr{P}(\omega)$,

(1) $\omega \in \mathscr{U}$ and $\varnothing \notin \mathscr{U}$,
(2) $X, Y \in \mathscr{U} \implies X \cap Y \in \mathscr{U}$,
(3) $X \notin \mathscr{U} \implies \omega \smallsetminus X \in \mathscr{U}$.

Thus M is a maximal ideal of $\mathscr{P}(\omega)$ if and only if $\{\omega \smallsetminus X \colon X \in M\}$ is an ultrafilter on ω; moreover, in this case,

$$\{\omega \smallsetminus X \colon X \in M\} = \mathscr{P}(\omega) \smallsetminus M.$$

Suppose $\mathscr{U}$ is a *nonprincipal* ultrafilter on ω. By Theorem 3, $\mathscr{U}$ must contain every **cofinite** subset of ω, that is, every subset whose complement is finite. We may think of the elements of $\mathscr{U}$ as **large,** and all other subsets of ω as **small.** Thus:

(1) every set is large or small, but not both;
(2) the small sets are precisely the complements of the large sets;
(3) all finite sets are small, and so their complements are large;
(4) the union of two finite sets is small, and the intersection of two large sets is large.

However, some infinite sets will be small. If ω is a *disjoint* union $A \sqcup B$, then exactly one of A and B is large. Hence if $\omega = A_0 \sqcup \cdots \sqcup A_{n-1}$, then A_i is large for exactly one index i.

4 Application to König's Lemma

In considering ultrafilters, we can replace ω with any infinite set. This allows us to prove theorems like the following, which need not have anything to do with calculus.

For present purposes, a **tree** is an ordered set[3] $(T, <)$ where, for every a in T, the set $\{x \in T \colon x < a\}$ of elements that are *below* a is linearly ordered and, moreover, well-ordered. This well-ordered set then has the order-type of an ordinal, which is called the **height** of a. If a has height $\alpha + 1$, and b has height α, and $b < a$, then a is a **successor** of b. In a **finitely branching** tree, every element has only finitely many successors.[4] A **branch** of a tree is a maximal linearly ordered subset.

Theorem 5 (König's Lemma). *Every infinite finitely branching tree has an infinite branch.*

[3] As I use the term, an *ordered set* is a set equipped with a binary relation that is irreflexive and transitive. If any two distinct elements of the set are comparable by means of the ordering, then this ordering is *linear.* Thus an ordered set is what is often called a "partially" ordered set, even though the ordering might be "total," that is, linear.

[4] We are not going to refer to successors again.

Proof. Let $(T, <)$ be such a tree, and let $\mathscr{U}$ be a nonprincipal ultrafilter on T. Being finitely branching, T can have only finitely many elements of each finite height. Suppose $n \in \omega$, and the elements of T of height n are a^0, ..., a^{m-1}. If $i < m$, define

$$A_i = \{x \in T \colon x \geqslant a^i\};$$

and let

$$A_m = \left\{x \in T \colon \bigvee_{i<m} x < a^i\right\}.$$

Then

$$T = A_0 \sqcup \cdots \sqcup A_{m-1} \sqcup A_m,$$

a disjoint union of $m + 1$ sets. Since the set A_m is small (being finite), exactly one of the other sets is large. Supposing A_i is large, let $b_n = a^i$. Then the infinite set $\{b_n \colon n \in \omega\}$ is linearly ordered; for if m and n are in ω, then, being large, the sets $\{x \in T \colon x \geqslant b_m\}$ and $\{x \in T \colon x \geqslant b_n\}$ have a common element c, and then $b_m \leqslant c$ and $b_n \leqslant c$, so b_m and b_n must be comparable, by the definition of a tree. $\square$

An alternative approach to König's Lemma sets the stage for our analysis of $\mathbb{R}$. For any set A, we define A^ω as we did $\mathbb{R}^\omega$.

Theorem 6. *For any set A, for any ultrafilter $\mathscr{U}$ on ω, the binary relation $\sim$ on A^ω given by*

$$a \sim b \iff \{k \in \omega \colon a_k = b_k\} \in \mathscr{U}$$

is an equivalence relation.

Given an ultrafilter $\mathscr{U}$ on A and an element a of A^ω, we may denote by any of

$$[a], \qquad [a_0, a_1, a_2, \dots], \qquad [a_k \colon k \in \omega]$$

the equivalence-class of a with respect to $\sim$ as in the theorem. We could write $[a]_{\mathscr{U}}$ for $[a]$; but we shall not be interested in changing the ultrafilter. We shall however denote the set of all of the classes $[a]$ by

$$A^{\omega}/\mathscr{U}.$$

This is an **ultrapower** of A.

Theorem 7. *For any set A, for any ultrafilter $\mathscr{U}$ on ω, the map $x \mapsto [x, x, x, \dots]$ is an embedding of A in $A^{\omega}/\mathscr{U}$. If A is finite or $\mathscr{U}$ is principal, then this embedding is also surjective; otherwise it is not.*

Proof. If a and b in A^{ω} are *constant* sequences, then the set $\{k \in \omega \colon a_k = b_k\}$ is either ω or $\varnothing$. The former must be true if $[a] = [b]$; but in that case, $a = b$.

Now suppose a is an arbitrary element of A^{ω}. If A is a finite set $\{b^0, \dots, b^{n-1}\}$, then for some unique i less than n, we must have $\{k \in \omega \colon a_k = b^i\} \in \mathscr{U}$. In this case, $[a] = [b^i, b^i, b^i, \dots]$, and so the indicated embedding is surjective.

If $\mathscr{U}$ is principal, this means it is $\{X \in \mathscr{P}(\omega) \colon i \in X\}$ for some i in ω. In particular, $\{i\} \in \mathscr{U}$. Then $[a] = [a_i, a_i, a_i, \dots]$, so the embedding is again surjective.

If A is infinite, then there is an element a of A^{ω} with no repeated terms. If also $\mathscr{U}$ is nonprincipal, then for all b in A, the set $\{k \in \omega \colon a_k = b\}$ has at most one element, so it is not large, and thus $[a]$ is not in the image of the embedding. $\square$

We shall usually treat the embedding of the theorem as an inclusion, identifying a in A with $[a, a, a, \dots]$ in $A^{\omega}/\mathscr{U}$. Thus, by the theorem, $A^{\omega}/\mathscr{U} \smallsetminus A$ is nonempty if and only if A is infinite and $\mathscr{U}$ is nonprincipal. Because of the this, we shall not be interested in principal ultrafilters, though some of our claims will be true for them as well as for nonprincipal ultrafilters.

Theorem 8. *For any ordered set $(A, <)$, for any ultrafilter $\mathscr{U}$ on ω, there is an ordering ${}^*\!<$ on $T^\omega/\mathscr{U}$ given by*

$$[a] \;{}^*\!< [b] \iff \{k \in \omega \colon a_k < b_k\} \in \mathscr{U}.$$

The embedding $x \mapsto [x, x, x, \dots]$ respects the orderings: for all a and b in A,

$$a < b \iff [a, a, a, \dots] \;{}^*\!< [b, b, b, \dots].$$

Finally we have the promised alternative formulation and proof of König's Lemma:

Theorem 9. *Suppose $(T, <)$ is an infinite, finitely branching tree, $\mathscr{U}$ is nonprincipal ultrafilter on ω, and ${}^*\!<$ is ordering of $T^\omega/\mathscr{U}$, induced as in the previous theorem. Then the tree $(T, <)$ has infinite branches, namely the sets*

$$\{x \in T \colon x \;{}^*\!< a\},$$

where $a \in (T^\omega/\mathscr{U}) \smallsetminus T$.

5 Limits

We return to considering fields, whose ultrapowers can be understood in terms of maximal ideals. The following is true for all ordered fields, and not just $\mathbb{R}$; but our main interest is in $\mathbb{R}$.

Theorem 10. *Let M be a nonprincipal maximal ideal of $\mathbb{R}^\omega$, and let $\mathscr{U}$ be the corresponding nonprincipal ultrafilter on ω, so that*

$$\mathscr{U} = \{\omega \smallsetminus \operatorname{supp}(x) \colon x \in M\} = \mathscr{P}(\omega) \smallsetminus \operatorname{supp}[M].$$

The quotient $\mathbb{R}^\omega/M$, which is a field, is precisely the ultrapower $\mathbb{R}^\omega/\mathscr{U}$. This field is an ordered *field with respect to the relation*

${}^*{<}$ of Theorem 8, and the embedding $x \mapsto [x, x, x, \dots]$ of $\mathbb{R}$ in $\mathbb{R}^{\omega}/\mathscr{U}$ is an embedding of ordered fields. The element $[0, 1, 2, \dots]$ of $\mathbb{R}^{\omega}/\mathscr{U}$ is positive and infinite, and the element $[1, 1/2, 1/3, \dots]$ is positive and infinitesimal.

By Łoś's Theorem (Theorem 14 below), all sentences of *first-order logic* will be true in $\mathbb{R}^{\omega}/\mathscr{U}$ if and only if they are true in $\mathbb{R}$. The same will be true when $\mathbb{R}$ is replaced by an arbitrary structure, such as a tree, as in Theorem 9. Then also Theorems 6, 7, 8, and 10 will turn out to have been special cases of Łoś's Theorem.

The formal definition of a first-order logic is recursive. For now, we shall just say that

(1) first-order sentences are *finite,* and
(2) their variables range only over *individuals.*

Here are corresponding examples of non-first-order sentences.

1. By the last theorem, $\mathbb{R}^{\omega}/\mathscr{U}$ differs from $\mathbb{R}$ in having infinite elements. That is, the sentence written as one of

$$\exists x\ (x > 1 \land x > 2 \land x > 3 \land \cdots), \qquad \exists x \bigwedge_{n \in \mathbb{N}} x > n$$

is true in $\mathbb{R}^{\omega}/\mathscr{U}$, but false in $\mathbb{R}$. Being infinite, in the sense of involving the conjunction of infinitely many formulas, this sentence is not first order.

2. In proving Theorem 1 (that $\mathbb{R}$ has no infinite elements), we showed that an ordered field with infinite elements cannot be complete. In particular, the sentence

$$\forall X\ \forall y\ \exists z\ \forall v\ \exists u$$
$$\Bigl(y \in X \to \bigl((v \in X \to v \leqslant z) \land \bigl((u \in X \land u > v) \lor z \leqslant v\bigr)\bigr)\Bigr),$$

which expresses completeness, is true in $\mathbb{R}$, but false in $\mathbb{R}^\omega/\mathscr{U}$. The sentence is not first order, because the variable X ranges over sets as such, not individuals.

We may denote the ordered field $\mathbb{R}^\omega/\mathscr{U}$ by

$$^*\mathbb{R}.$$

We now establish the equivalence of the standard and nonstandard definitions of limit.

Theorem 11. *Suppose f is a function from $\mathbb{R}$ to itself, and a and L are in $\mathbb{R}$. There is a well-defined function $^*f$ from $^*\mathbb{R}$ to itself given by*

$$^*f[x] = [f(x_k) \colon k \in \omega].$$

Then $\lim_a f = L$ *if and only if, for all x in $\mathbb{R}^\omega$,*

$$[x] \simeq a \ \ \& \ \ [x] \neq a \implies {}^*f[x] \simeq L.$$

Proof. For all a and b in $\mathbb{R}^\omega$, we have

$$\{k \in \omega \colon a_k = b_k\} \subseteq \{k \in \omega \colon f(a_k) = f(b_k)\}.$$

If $[a] = [b]$, then the former set is large, and therefore the latter set is large, which means

$$[f(a_k) \colon k \in \omega] = [f(b_k) \colon k \in \omega].$$

Thus $^*f$ is well defined.

Suppose $\lim_a f = L$, and $[x] \simeq a$, but $[x] \neq a$. For all n in $\mathbb{N}$, for some positive δ in $\mathbb{R}$, for all k in ω, if $0 < |x_k - a| < \delta$, then $|f(x_k) - L| < 1/n$. Thus

$$\{k \in \omega \colon 0 < |x_k - a| < \delta\} \subseteq \{k \in \omega \colon |f(x_k) - L| < 1/n\}. \quad (8)$$

The former set is

$$\{k \in \omega \colon x_k \neq a\} \cap \{k \in \omega \colon x_k - a < \delta\} \cap \{k \in \omega \colon a - x_k < \delta\},$$

the intersection of three sets, all of which are large, since, respectively, $[x] \neq 0$, and $[x]-a < \delta$, and $a-[x] < \delta$. So the intersection itself is large, and therefore the latter set in (8) is large, which means

$$|{}^*f[x] - L| < \frac{1}{n}.$$

This being the case for all n in $\mathbb{N}$, ${}^*f[x] \simeq L$.

Now suppose $\lim_a f \neq L$. Then for some positive ε in $\mathbb{R}$, for every n in ω, there is x_n in $\mathbb{R}$ such that $0 < |x_n - a| < 1/(n+1)$, but $|f(x_n)-L| \geqslant \varepsilon$. Then $[x] \simeq a$, but $[x] \neq a$, and $|{}^*f[x]-L| \geqslant \varepsilon$, so ${}^*f[x] \not\simeq L$. □

The nonstandard definition makes proofs about limits easier, at least once one has the following.

Theorem 12. *In any ordered field, the finite elements compose a sub-ring, and the infinitesimal elements compose a maximal ideal of the ring of finite elements. In fact the ideal is the only maximal ideal of the ring.*

Corollary. *In the ring of finite elements of any ordered field, the relation $\simeq$ of being infinitesimally close is an equivalence relation, and*

$$a \simeq b \ \ \& \ \ c \simeq d \implies a + c \simeq b + d \ \ \& \ \ ac \simeq bd,$$
$$a \simeq b \ \ \& \ \ a \not\simeq 0 \implies b \not\simeq 0 \ \ \& \ \ \frac{1}{a} \simeq \frac{1}{b}.$$

The "standard" proof of the following involves taking different deltas for different epsilons, and then taking the maximum. The "nonstandard" proof is more straightforward.

Theorem 13. *Suppose $\lim_a f = L$ and $\lim_a f = M$. Then*

$$\lim_a (f + g) = L + M, \qquad \lim_a (fg) = LM.$$

If $L \neq 0$, then

$$\lim_a \frac{1}{f} = \frac{1}{L}.$$

Proof. Suppose $[x] \simeq a$, but $[x] \neq a$. By hypothesis and Theorem 11, ${}^*f[x] \simeq L$ and ${}^*g[x] \simeq M$. Then the corollary of Theorem 12 yields the desired results. □

6 Standard parts

Every finite element a of ${}^*\mathbb{R}$ is infinitesimally close to a unique element $\mathrm{st}\,(a)$ of $\mathbb{R}$, namely

$$\sup\{x \in \mathbb{R} \colon x < a\}.$$

This is the *standard part* of a. If a is infinite, we may define $\mathrm{st}\,(a) = \infty$. Then the function $x \mapsto \mathrm{st}\,(x)$ from ${}^*\mathbb{R}$ to $\mathbb{R} \cup \{\infty\}$, is an example of a *place,* as we shall see.

In Theorem 12, if the ordered field is K, and the ring of its finite elements is R, then

$$x \in K \smallsetminus R \implies \frac{1}{x} \in R.$$

This means R is a **valuation ring** of K, at least if R is different from K. The definition does not require K to be ordered. Every valuation ring R of a field K has a unique maximal ideal M, which is the complement of the multiplicative group $R^{\times}$ of invertible elements of R. Then two functions can be defined on K as follows.

1. The function φ with range $R/M \cup \{\infty\}$ given by

$$\varphi(x) = \begin{cases} x + M, & \text{if } x \in R, \\ \infty, & \text{if } x \in K \smallsetminus R \end{cases}$$

is a **place** of K.

2. The function v with range $\{0\} \cup K^\times/R^\times$ given by

$$v(x) = \begin{cases} xR^\times, & \text{if } x \in K^\times, \\ 0, & \text{if } x = 0 \end{cases}$$

is a **valuation** of K. Here

$$v(x) = 0 \iff x = 0,$$

and

$$v(-x) = v(x), \qquad v(xy) = v(x) \cdot v(y).$$

The multiplicative group $K^\times/R^\times$ is linearly ordered by the rule

$$xR^\times \leqslant yR^\times \iff \frac{x}{y} \in R,$$

and 0 is declared to be less than every element of the group. Then a strong triangle inequality holds, namely

$$v(x + y) \leqslant \max(v(x), v(y)).$$

Thus the map $(x, y) \mapsto v(x - y)$ is a *metric* on K.

There are two standard classes of examples.

1. If K is any field, we can form the field $K(X)$ of rational functions in X over K. If $\alpha \in K$, then the function $f \mapsto f(\alpha)$ is a place on $K(X)$. Here, if $f(\alpha)$ is "undefined," because it involves division by 0, then we define $f(\alpha) = \infty$. Thus the range of the place is indeed $K \cup \{\infty\}$. If K is algebraically closed, like $\mathbb{C}$, then every nonzero element of $K(X)$ can be written as

$$\frac{p(X)}{q(X)} \cdot (X - \alpha)^n,$$

where p and q are polynomials that are not zero at α, and $n \in \mathbb{Z}$. If $0 < t < 1$, the valuation can be understood as taking the above element of $K^\times$ to t^n. In particular, according to the associated metric, the sequence $(X^n : n \in \omega)$ converges to 0.

2. If p is a prime number, then $\mathbb{Q}$ has a valuation ring consisting of the fractions having denominators that are indivisible by p. The range of the associated place can be understood as $\mathbb{F}_p \cup \{\infty\}$. Indeed, if $p \nmid b$, then the place can be understood as taking a/b to that element c of $\mathbb{F}_p{}^\times$ such that $a \equiv bc \pmod{p}$. The associated valuation can be understood as the map taking $(a/b) \cdot p^n$ to $1/p^n$ (and 0 to 0). In the associated metric, the sequence $(p^n : n \in \mathbb{N})$ converges to 0, and so every series $\sum_{n=m}^{\infty} a_n p^n$, where $m \in \mathbb{Z}$ and $a_n \in \{0, \ldots, p-1\}$, has a sequence of partial sums that is a Cauchy sequence. Such series then constitute the field of **p-adic numbers.** This is a field that, like $\mathbb{R}$, is complete in the sense that all of its Cauchy sequences converge. This notion of completeness should be distinguished from the completeness of the *ordering* of $\mathbb{R}$.

7 Logic

In the **first-order logic** of ordered fields, the **atomic formulas** are the polynomial equations and inequalities. Then the **formulas** are defined recursively:

1. Every atomic formula is a formula.
2. If φ and ψ are formulas, and x is an individual variable (that is, such a variable as may occur in an atomic formula), then each of the expressions
$$\neg\varphi, \qquad (\varphi \wedge \psi), \qquad \exists x\ \varphi$$
is a formula.

By this scheme, the expressions
$$(\varphi \vee \psi), \qquad (\varphi \to \psi), \qquad (\varphi \leftrightarrow \psi), \qquad \forall x\ \varphi$$
are abbreviations of the formulas
$$\neg(\neg\varphi \wedge \neg\psi), \quad (\neg\varphi \vee \psi), \quad (\varphi \to \psi) \wedge (\psi \to \varphi), \quad \neg\exists x\ \neg\varphi$$

respectively.

An occurrence of a variable x in a formula is **bound** if it enters the formula as occurring in a formula $\exists x\ \varphi$; otherwise the occurrence is **free.** Thus, in the formula $(x = y \lor \exists x \neg\ x = y)$ (usually written as $x \neq y \to \exists x\ x \neq y$), the first occurrence of x is free, but the other two occurrences are bound.

A formula in which no variable occurs freely is a **sentence.** By the definition to be given presently, a sentence σ is either **true** or **false** in a given ordered field K, and we write

$$K \vDash \sigma, \qquad\qquad K \nvDash \sigma$$

accordingly. For now, we take the definition of the truth or falsity of an atomic sentence in K as obvious. Then for all sentences σ and τ,

$$\begin{aligned} K \vDash \neg\sigma &\iff K \nvDash \sigma, \\ K \vDash (\sigma \land \tau) &\iff K \vDash \sigma \ \ \& \ \ K \vDash \tau, \end{aligned}$$

Note here that the two-shafted arrow $\iff$ is just an abbreviation of the English expression "if and only if"; and the ampersand & is an abbreviation of "and." We could replace the abbreviations with the full expressions without changing the mathematical meaning of what we are saying. By contrast, we cannot replace the symbol $\land$ in the formula $(\sigma \land \tau)$ with anything else, because our purpose is to study the uses of such formulas as they are.

It remains to declare that $K \vDash \exists x\ \varphi$ if and only if, for some a in K,

$$K \vDash \varphi(a).$$

Here $\varphi(a)$ is the result of replacing each free occurrence of x in φ with a.

We can prove Łoś's Theorem just for $\mathbb{R}$, but a more general form may actually be easier to understand. Suppose $(K_n \colon n \in \omega)$

is an indexed family of ordered fields. Then we can form the product

$$\prod_{n\in\omega} K_n,$$

namely the set of sequences a or $(a_n\colon n \in \omega)$, where $a_n \in K_n$ for each n in ω. Let us denote the product also by R. Each element a of R can be understood as a **constant symbol,** whose **interpretation** in the ordered field K_n is just a_n, for each n in ω. In other words, a *means* a_0 in K_0, and a_1 in K_1, and so forth. Likewise, the identity of an abelian group is usually 0, and this *means* 0 in the additive group $\mathbb{Z}$; but it means 1 in multiplicative groups like $\mathbb{R}^\times$. For an arbitrary element b of K_n, with a in R as before,

$$K_n \vDash a = b \iff a_n = b.$$

Now let $\mathscr{S}$ be the set

$$\{+, 0, -, \,\cdot\,, 1, <\}$$

of symbols: this is the **signature** of ordered fields. The constant symbols in $\mathscr{S}$ are only 0 and 1. Using these, we can form the polynomials

$$\Bigl(\cdots\bigl((1+1)+1\bigr)+\cdots+1\Bigr), \quad -\Bigl(\cdots\bigl((1+1)+1\bigr)+\cdots+1\Bigr).$$

According to the associativity of addition in every field, these polynomials can be written unambiguously as

$$1+\cdots+1, \qquad\qquad -(1+\cdots+1).$$

Every integer can be denoted by such an expression. (Even 0 can be so denoted, if we allow the "empty" sum.) Thus, every equation or inequality of polynomials with coefficients from $\mathbb{Z}$ can be understood as an atomic formula in the pure signature of ordered fields. We may also introduce, as coefficients, new constant symbols standing for elements of a particular ordered field or fields. The new constant symbols can be called **parameters.** With R being $\prod_{n\in\omega} K_n$ as above, if σ is a sentence in the signature $\mathscr{S}\cup R$ of ordered fields with parameters from R, we define

$$\|\sigma\| = \{n \in \omega\colon K_n \vDash \sigma\}.$$

(In his *Model Theory,* Hodges calls this the *Boolean value* of σ.) Then by the definition of truth and falsity,

$$\begin{gathered}\|\neg\sigma\| = \omega \smallsetminus \|\sigma\|,\\ \|(\sigma \wedge \tau)\| = \|\sigma\| \cap \|\tau\|,\end{gathered}$$

and for all a in R,

$$\|\varphi(a)\| \subseteq \|\exists x\ \varphi\|.$$

Moreover, given the formula φ in which only x occurs freely, we can define a in R so that, for all n in ω, if $K_n \vDash \exists x\ \varphi$, then $K_n \vDash \varphi(a_n)$, but if $K_n \nvDash \exists x\ \varphi$, then a_n is an arbitrary element of K_n (for example, 0 or 1). In this case

$$\|\exists x\ \varphi\| = \|\varphi(a)\|. \tag{9}$$

Now let $\mathscr{U}$ be an ultrafilter on ω, and let T be the set of sentences σ of $\mathscr{S}\cup R$ such that $\|\sigma\| \in \mathscr{U}$. Then T contains every sentence of $\mathscr{S}\cup R$ that is true in every ordered field (in which the parameters have interpretations). Moreover,

$$\neg\sigma \in T \iff \sigma \notin T, \tag{10a}$$

$$(\sigma \wedge \tau) \in T \iff \sigma \in T \ \&\ \tau \in T, \tag{10b}$$

$$\exists x\ \varphi \in T \iff \text{for some } a \text{ in } R,\ \varphi(a) \in T. \tag{10c}$$

Now, suppose F is some ordered field, and every element of R has an interpretation in F, and every element of F is the interpretation of an element of R. The set of sentences of $\mathscr{S}\cup R$ that are true in F is the **(complete) theory of F** in the indicated signature. But then, by the definition of truth, this theory has the properties of T that we have just determined. The converse will turn out to true as well: by having the indicated properties, T must itself be the complete theory of some ordered field (in which the elements of R have interpretations). We shall denote this ordered field (whose theory is T) by

$$\prod_{n\in\omega} K_n \Big/ \mathscr{U}$$

or just $R/\mathscr{U}$: it is an **ultraproduct** of the ordered fields K_n. For all a and b in R, since

$$R/\mathscr{U} \models a = b \iff \{k \in \omega \colon a_k = b_k\},$$

we can understand the elements of $R/\mathscr{U}$ as being the equivalence classes $[a]$ or

$$\bigl\{x \in R \colon \{k \in \omega \colon a_k = x_k\}\bigr\},$$

where $a \in R$.

In the special case where each ordered field K_n is $\mathbb{R}$, then the ultraproduct $\prod_{n\in\omega} K_n/\mathscr{U}$ will be the ultrapower ${}^*\mathbb{R}$, that is, $\mathbb{R}^\omega/\mathscr{U}$. This is an ordered field by Theorem 10, and we may consider $\mathbb{R}$ as a subfield. Then a (first-order) sentence σ with parameters from $\mathbb{R}$ will be true in $\mathbb{R}$ if and only if it is true in ${}^*\mathbb{R}$, because $\|\sigma\|$ in this case is either $\mathscr{U}$ or $\varnothing$.

A structure in which every sentence of a set of sentences is true is a **model** of that set. So we are going to show that the set T above has a model. This will give us (a special case of) Łoś's Theorem. There is a sense in which we need not prove this theorem though, just to do "nonstandard," infinitesimal analysis. Analysis is a study of $\mathbb{R}$, not ${}^*\mathbb{R}$. We shall not really use the latter itself, but only its theory, and we have this already.

8 Łoś's Theorem

Strictly speaking though, even to understand the theory T above, we need to do more work. We should understand the structure of polynomials. The symbols $+$, $-$, and $\cdot$ in $\mathscr{S}$ are examples of **operation symbols,** taking two, one, and two arguments respectively; and polynomials are examples of *terms.*

In an arbitrary signature, the **terms** are defined recursively as follows.

1. every individual variable is a term.
2. every constant symbol is a term.
3. for every n in $\mathbb{N}$, if F is an n-ary operation symbol (that is, an operation symbol taking n arguments), and t_i is a term when $i < n$, then the expression
$$Ft_0 \cdots t_{n-1}$$
is a term. In case $n = 2$, it is conventional to write $t_0 \; F \; t_1$ instead of Ft_0t_1.

A term is called **closed** if no variable occurs in it. Thus closed terms are obtained by omitting the first condition in the definition of terms.

An **interpretation** of an n-ary operation symbol on a set A is a function from A^n to A (where A^n can be understood as the set of functions from $\{0, \dots, n-1\}$ to A).[5] If each constant symbol and operation symbol in a signature has an interpretation on A, then, recursively, each closed term of $\mathscr{S}$ has an interpretation as an element of A.[6]

[5] I think this is the best way to understand A^n, but many people understand it as the set of functions from the set $\{1, \dots, n\}$ to A.

[6] An arbitrary term can be interpreted as an operation on A; but the best way of making this precise is not clear. It might be best to interpret the term $x + y$ as a function from $A^{\{x,y\}}$ to A; but then if this term occurs in a formula where the variable z also occurs, then we should interpret $x + y$ as a function from $A^{\{x,y,z\}}$ to A.

The symbol $<$ in the signature of ordered fields is a binary **relation symbol.** In an arbitrary signature, the atomic formulas are of two kinds:

1. There are equations $t_0 = t_1$ of terms t_0 and t_1.
2. For every n in $\mathbb{N}$, for every n-ary relation symbol in the signature, if t_i is a term when $i < n$, there is an atomic formula
$$Rt_0 \cdots t_{n-1}.$$

If $n = 2$, it is conventional to write $t_0 \; R \; t_1$ instead of Rt_0t_1. An **interpretation** of an n-ary relation symbol on a set A is a subset of A^n. An interpretation of all symbols in a signature $\mathscr{S}$, on a set, is a **structure** whose **universe** is the set. If the set is A, then the structure itself may be $\mathfrak{A}$ (the same letter in a more elaborate font), and the interpretation of a symbol or closed term s can be denoted by
$$s^{\mathfrak{A}}.$$

Then the truth or falsity of an atomic sentence in $\mathfrak{A}$ has the obvious definition:

$$\mathfrak{A} \models t_0 = t_1 \iff {t_0}^{\mathfrak{A}} = {t_1}^{\mathfrak{A}},$$
$$\mathfrak{A} \models Rt_0 \cdots t_{n-1} \iff ({t_0}^{\mathfrak{A}}, \ldots, {t_{n-1}}^{\mathfrak{A}}) \in R^{\mathfrak{A}}.$$

Now arbitrary sentences, and their truth or falsity, are defined as before. We just have to note that, unless otherwise specified, every element of a set can be understood as a constant symbol whose interpretation is itself.

We now have all of the ingredients for the full statement of Łoś's Theorem. As we did when proving König's Lemma, we can replace our usual index set ω with an arbitrary index set.

Theorem 14 (Łoś)**.** *In a signature $\mathscr{S}$, let $(\mathfrak{A}_i \colon i \in \Omega)$ be a family of structures indexed by elements of some infinite set Ω.*

In the product $\prod_{i\in\Omega} A_i$ of the universes of the structures $\mathfrak{A}_i$, let each element $(a_i\colon i\in\Omega)$ be considered as a new constant symbol, interpreted in $\mathfrak{A}_i$ as a_i for each i in Ω. Let $\mathscr{U}$ be an ultrafilter on Ω. In the signature $\mathscr{S}\cup\prod_{i\in\Omega} A_i$, there is a new structure

(1) *whose complete theory consists precisely of those σ such that*

$$\{i\in\Omega\colon \mathfrak{A}_i\vDash\sigma\}\in\mathscr{U}, \tag{11}$$

and

(2) *whose universe consists of the interpretations of the elements of*

$$\prod_{i\in\Omega} A_i.$$

In case each of the structures $\mathfrak{A}_i$ is the same structure $\mathfrak{A}$, then this embeds in the new structure under the map sending each a in A to the interpretation of the constant sequence $(a\colon i\in\Omega)$. In this case, a sentence of $\mathscr{S}\cup A$ is true in the new structure if and only if it is true in $\mathfrak{A}$.

Proof. Let $B=\prod_{i\in\Omega} A_i$, and let T be the set of sentences of $\mathscr{S}\cup B$ such that (11) holds. As we have shown, T must have the properties in (10). In particular, by (10a), if T has a model, then T is the complete theory of this model. So we want to show T has a model whose universe consists of the interpretations of the elements of B. In fact, this will follow from (10), along with the observation that T contains all sentences that are **logically true,** that is, true in all structures of the signature $\mathscr{S}\cup B$. Indeed, for all a, b, and c in B, T contains the sentences

$$\begin{gathered} a=a,\\ a=b\to b=a,\\ a=b\land b=c\to a=c. \end{gathered}$$

Therefore the relation $\sim$ on B given by

$$a \sim b \iff \text{the sentence } a = b \text{ is in } T$$

is an equivalence relation. Let us denote by $B/\mathscr{U}$ the set of equivalence classes of elements of B. For every n in $\mathbb{N}$, for every n-ary operation symbol F of $\mathscr{S}$, for all $(a_0, \dots, a_{n-1})$ and $(b_0, \dots, b_{n-1})$ in B^n, T contains

$$\bigwedge_{i<n} a_i = b_i \to Fa_0 \cdots a_{n-1} = Fb_0 \cdots b_{n-1},$$
$$\exists x \; Fa_0 \cdots a_{n-1} = x.$$

Therefore, by (10c), T determines an interpretation of F on $B/\mathscr{U}$. Similarly, T determines an interpretation of every relation symbol in $\mathscr{S}$. Likewise, every atomic sentence is true in the resulting structure if and only if it belongs to T. Then the same is true for *every* sentence, by (10). $\square$

The new structure given by the theorem can be denoted by

$$\prod_{i\in\Omega} \mathfrak{A}_i \Big/ \mathscr{U};$$

it is the **ultraproduct** of the indexed family $(\mathfrak{A}_i \colon i \in \Omega)$ with respect to the ultrafilter $\mathscr{U}$. In case $\mathscr{U}$ is a *principal* ultrafilter $\{X \in \mathscr{P}(\Omega) \colon i \in X\}$ for some i in Ω, then T is just the complete theory of $\mathfrak{A}_i$, and so the theorem is trivial. It is the existence of nonprincipal ultrafilters that makes the theorem interesting and useful.

In the special case of the theorem where each $\mathfrak{A}_i$ is the same structure $\mathfrak{A}$, the ultraproduct is an **ultrapower,** denoted by

$$\mathfrak{A}^\Omega/\mathscr{U}.$$

In this case, the conclusion of Łoś's Theorem is that $\mathfrak{A}$ is an **elementary substructure** of the ultrapower. We denote this situation by

$$\mathfrak{A} \preccurlyeq \mathfrak{A}^{\Omega}/\mathscr{U}.$$

9 Applications

We can now understand Theorem 11 as a consequence of Łoś's Theorem in the following way. We are given a function f from $\mathbb{R}$ to $\mathbb{R}$. this means f is a singulary operation on $\mathbb{R}$, and as such it can considered as the interpretation of some singulary operation symbol (which we also write as f). This symbol is interpreted in the ultrapower $^*\mathbb{R}$ as $^*f$. We assume $\lim_a f = L$, that is, for all n in $\mathbb{N}$, for some positive δ in $\mathbb{R}$, the sentence

$$\forall x \ \left(0 < |x - a| < \delta \rightarrow |f(x) - L| < \frac{1}{n}\right)$$

is true in $\mathbb{R}$. Then the same sentence is true in $^*\mathbb{R}$. In particular, in $^*\mathbb{R}$, suppose $x \simeq a$, but $x \neq a$. Then $0 < |x - a| < \delta$, so $|f(x) - L| < 1/n$, where the symbol f means $^*f$. This being so for all n in $\mathbb{N}$, we have $^*f(x) \simeq L$.

If instead $\lim_a f \neq L$, then for some positive ε in $\mathbb{R}$, the sentence

$$\forall \delta \ \exists x \ (\delta > 0 \rightarrow 0 < |x - a| < \delta \wedge |f(x) - L| \geqslant \varepsilon)$$

is true in $\mathbb{R}$, hence in $^*\mathbb{R}$. In particular, letting δ be a positive infinitesimal, we obtain x in $^*\mathbb{R}$ such that $x \simeq a$, but $x \neq a$, and $^*f(x) \not\simeq L$.

If we call $^*\mathbb{R}$ a **(nonstandard) extension** of $\mathbb{R}$, then $^*f$ as above is the corresponding extension of the function f. Similarly, being a subset of $\mathbb{R}$, the set $\mathbb{N}$ is a singulary relation on $\mathbb{R}$, and so it has an extension $^*\mathbb{N}$, which is the set of all a in $^*\mathbb{R}$ such that the sentence $a \in \mathbb{N}$ is true in ${}^*\mathbb{R}$. (By the formal definition above, the sentence $a \in \mathbb{N}$ would be written as $\mathbb{N}a$.) This means

$$ {}^*\mathbb{N} = \big\{[x] \colon x \in \mathbb{R}^\omega \;\;\&\;\; \{k \in \omega \colon x_i \in \mathbb{N}\} \in \mathscr{U}\big\}. $$

Theorem 15. $\mathbb{N} \subset {}^*\mathbb{N}$, *and the finite elements of* ${}^*\mathbb{N}$ *are precisely the elements of* $\mathbb{N}$.

Proof. For every n in $\mathbb{N}$, the sentence $n \in \mathbb{N}$ is true in $\mathbb{R}$, hence in ${}^*\mathbb{R}$; so $\mathbb{N} \subseteq {}^*\mathbb{N}$. Since the sentence

$$ \forall x \,\big(x \in \mathbb{N} \to 1 \leqslant x \land (n \leqslant x < n + 1 \to n = x)\big) $$

is true in $\mathbb{R}$, it is true in ${}^*\mathbb{R}$, and so there are no new finite elements of ${}^*\mathbb{N}$. Finally, the sentence $\forall x \;\exists y \;(y \in \mathbb{N} \land y > x)$ is true in $\mathbb{R}$, hence in ${}^*\mathbb{R}$, so letting x be a positive infinite element of ${}^*\mathbb{R}$ gives us an infinite element of ${}^*\mathbb{N}$; this element is not in $\mathbb{N}$, so $\mathbb{N} \subset {}^*\mathbb{N}$. □

If now a is a sequence $(a_k \colon k \in \mathbb{N})$ of elements of $\mathbb{R}$, this means a is the binary relation $\{(k, a_k) \colon k \in \mathbb{N}\}$ on $\mathbb{R}$, so it has an extension ${}^*a$. This extension must be a sequence $({}^*a_n \colon n \in {}^*\mathbb{N})$, where ${}^*a_n = a_n$ if $n \in \mathbb{N}$; for the sentences

$$ \forall x \;\forall y \;(x \; a \; y \to x \in \mathbb{N}), $$
$$ \forall x \;\forall y \;\forall z \;(x \; a \; y \land x \; a \; z \to y = z) $$

are true in $\mathbb{R}$, hence in ${}^*\mathbb{R}$.

Theorem 16. *A real-valued sequence a on $\mathbb{N}$ is bounded if and only if, for each infinite n in ${}^*\mathbb{N}$, ${}^*a_n$ is finite.*

Theorem 17. *For all real-valued sequences a on $\mathbb{N}$, for all L in $\mathbb{R}$,*

$$ \lim_{n \to \infty} a_n = L $$

if and only if, for all infinite n in ${}^*\mathbb{N}$,

$$ {}^*a_n \simeq L. $$

Recall that a real-valued sequence a on $\mathbb{N}$ is a **Cauchy sequence** if for all positive ε in $\mathbb{R}$, for some ℓ in $\mathbb{N}$,

$$m > \ell \ \& \ n > \ell \implies |a_m - a_n| < \varepsilon.$$

Easily, every convergent sequence is a Cauchy sequence. The converse is more difficult. The following is the "nonstandard" version of the theorem: the proof will use standard parts as defined above.

Theorem 18. *A real-valued sequence a on $\mathbb{N}$ is convergent if and only if, for all infinite m and n,*

$$a_m \simeq a_n.$$

10 Mock second-order logic

We may consider a structure as having a universe partitioned into several subsets, called **sorts.** For example, a vector space has a sort of vectors and a sort of scalars. Correspondingly, in the logic of vector spaces, there will be vector variables and scalar variables. In a many-sorted structure, operations and relations do not simply take n arguments for some n in $\mathbb{N}$, but the sort of each argument must be specified. Thus, in a vector space, two vectors can be added together, and two scalars can be added together, but not a vector and a scalar.

We can consider the natural numbers as composing a two-sorted structure, the sorts being $\mathbb{N}$ and $\mathscr{P}(\mathbb{N})$. The variables for these sorts are minuscule and capital letters, respectively. In addition to the relation symbol $<$, which takes two arguments from $\mathbb{N}$, let the signature contain the membership relation $\in$, taking an argument from $\mathbb{N}$ and an argument from $\mathscr{P}(\mathbb{N})$. Then we can express the well-ordering property of $(\mathbb{N}, <)$ as a *first-order* sentence in this signature:

$$\forall X\ \forall y\Big(y \in X \to \exists z\ \big(z \in X \land \forall w\ (w \in X \to z \leqslant w)\big)\Big). \tag{12}$$

Denoting our two-sorted structure by $\mathbb{N} \sqcup \mathscr{P}(\mathbb{N})$, we can form the extension $^*(\mathbb{N} \sqcup \mathscr{P}(\mathbb{N}))$, which can be understood as having the two sorts $^*\mathbb{N}$ and $^*\mathscr{P}(\mathbb{N})$. Then the same sentence (12) is true in the extension. However, this does not mean that $^*\mathbb{N}$ is well ordered. This is because $^*\mathscr{P}(\mathbb{N})$ is not $\mathscr{P}(^*\mathbb{N})$. If $a \in \mathbb{N}^\omega$ and $B \in \mathscr{P}(\mathbb{N})^\omega$, we have

$$[a] \mathrel{^*{\in}} [B] \iff \{k \in \omega \colon a_k \in B_k\} \in \mathscr{U}.$$

Thus we can identify $[B]$ with the subset

$$\bigl\{[x] \in {}^*\mathbb{N} \colon \{k \in \omega \colon x_k \in B_k\} \in \mathscr{U}\bigr\} \tag{13}$$

of $^*\mathbb{N}$. We can denote this subset of $^*\mathbb{N}$ by

$$\prod_{k \in \omega} B_k \Big/ \mathscr{U};$$

however, it must be understood that if $[a]$ belongs to this set, this does not mean $a \in \prod_{k\in\omega} B_k$, but only that $\{k \in \omega \colon a_k \in B_k\} \in \mathscr{U}$, as in (13). For example, if some B_k is empty, then $\prod_{k\in\omega} B_k = \varnothing$, but this does not imply that the set in (13) is empty.

The embedding $[X] \mapsto \prod_{k\in\omega} X_k / \mathscr{U}$ of $^*\mathscr{P}(\mathbb{N})$ in $\mathscr{P}(^*\mathbb{N})$ is not surjective. For example, $\mathbb{N}$ itself (considered as a subset of $^*\mathbb{N}$) is not in the image of $^*\mathscr{P}(\mathbb{N})$. We can see this in two ways.

1. Every nonempty subset of $^*\mathscr{P}(\mathbb{N})$ has a least element, because (12) is true in $^*(\mathbb{N} \sqcup \mathscr{P}(\mathbb{N}))$; but $^*\mathbb{N} \smallsetminus \mathbb{N}$ has no least element.

2. More directly, suppose B in $\mathscr{P}(\mathbb{N})^\omega$ is such that the element of $\mathscr{P}(^*\mathbb{N})$ in (13) includes $\mathbb{N}$. This means, for each n in $\mathbb{N}$,

$$\{k \in \omega \colon n \in B_k\} \in \mathscr{U}.$$

In particular, there are infinitely many k in ω such that $n \in B_k$. Thus if, for all k in ω, we define

$$a_k = \max\{n \in B_k \colon n \leqslant k+1\},$$

then the sequence $(a_k\colon k \in \omega)$ is unbounded, and so $[a_k\colon k \in \omega]$ is not in (the image of) $\mathbb{N}$, though it is in $\prod_{k\in\omega} B_k/\mathscr{U}$.

We can introduce power sets as sorts in order to define *integration.* By one standard definition (attributed to Riemann), a real-valued function f whose domain includes a closed interval $[a, b]$ is **integrable** on $[a, b]$ if, for some I in $\mathbb{R}$, for all positive ε in $\mathbb{R}$, there is a positive δ in $\mathbb{R}$ such that, for all n in $\mathbb{N}$, for all subsets $\{a_0, \dots, a_n\}$ of $[a, b]$, for all subsets $\{\xi_0, \dots, \xi_{n-1}\}$ of $[a, b]$, if

$$\left.\begin{gathered} a = a_0 < \dots < a_n = b, \\ a_0 \leqslant \xi_0 \leqslant a_1 \;\;\&\;\; \cdots \;\;\&\;\; a_{n-1} \leqslant \xi_{n-1} \leqslant a_n, \end{gathered}\right\} \tag{14}$$

and

$$\max\{a_{k+1} - a_k\colon k < n\} < \delta,$$

then

$$\left|\sum_{k<n} f(\xi_k)\cdot(a_{k+1} - a_k) - I\right| < \varepsilon.$$

In this case I is unique and is the **integral** of f on $[a, b]$; the integral is denoted by

$$\int_a^b f.$$

If f is given only in the form $x \mapsto f(x)$, we may write the integral as

$$\int_a^b f(x)\,\mathrm{d}\,x.$$

The idea of the notation is that $\mathrm{d}\,x$ is infinitesimal, so that the integral itself is the *sum* (abbreviated by $\int$, an elongated letter S) of rectangles of infinitesimal width. Given f as before, and I in $\mathbb{R}$, suppose that, for all n in ${}^*\mathbb{N}$, for all subsets $\{a_0, \dots, a_n\}$

of ${}^*[a,b]$ and for all subsets $\{\xi_0, \dots, \xi_{n-1}\}$ of ${}^*[a,b]$ such that (14) and

$$a_0 \simeq a_1 \ \ \& \ \ \cdots \ \ \& \ \ a_{n-1} \simeq a_n$$

hold, it follows that

$$\sum_{k<n} f(\xi_k) \cdot (a_{k+1} - a_k) \simeq I.$$

In this case, by the nonstandard definition, $I = \int_a^b f(x)\,\mathrm{d}\,x$.

11 Set-theoretic considerations

The signature of set theory as $\{\in\}$. The **Extension Axiom** is that sets with the same elements are equal:

$$\forall x\ \forall y\ \big(\forall z\ (z \in x \leftrightarrow z \in y) \to x = y\big).$$

If φ is a formula of the signature $\{\in\}$ in which only the variable x occurs freely, then those sets a such that $\varphi(a)$ is true compose a **class,** denoted by

$$\{x \colon \varphi\}.$$

Such an expression is not literally part of the logic of sets. However, in what is known as Zermelo–Fraenkel set theory, every axiom besides the Foundation Axiom is that certain classes are sets.[7] For example, the "power class" of a set a is the class

$$\{x \colon \forall y\ (y \in x \to y \in a)\};$$

[7]This is not quite true. The **Foundation Axiom** is that every nonempty set has a minimal element with respect to $\in$, that is, $\forall x\ \forall y\ \exists z\ \forall w\ \big(y \in x \to z \in x \land (w \in z \to w \notin x)\big)$. There could in fact be sets without minimal elements; all the Foundation Axiom does is to decline to consider such sets *as* sets.

the Power Set Axiom is accordingly

$$\forall z\,\exists w\,\forall x\,\bigl(x \in w \to \forall y\,(y \in x \to y \in z)\bigr).$$

The Axiom of Choice is not of this form, but is that, for every set a, there is a function f such that, for every nonempty subset b of a,

$$f(b) \in b.$$

We used a version of this in proving Łoś's Theorem, when we found $(a_n : n \in \omega)$ so that (9) on page 23 holds.

Łoś's Theorem requires also the existence of nonprincipal ultrafilters on arbitrary sets. As we know, ultrafilters are complements of maximal ideals of Boolean rings. The **Maximal Ideal Theorem** is that, in arbitrary ring R, if I is a proper ideal, then I is included in a maximal ideal.

Theorem 19. *The Maximal Ideal Theorem is a consequence of the Axiom of Choice.*

Proof. This is commonly proved by means of *Zorn's Lemma,* which is equivalent to the Axiom of Choice. A more useful approach for present purposes is the following. The Axiom of Choice allows us to well-order any ring R, writing it as $\{a_\alpha : \alpha < \kappa\}$, where κ is now the *cardinality* of R. We define an increasing sequence of ideals of R by *transfinite recursion*:

1. I_0 is the given proper ideal I of R.
2. For all α in κ, If $I_\alpha \cup \{a_\alpha\}$ generates a proper ideal of R, this ideal is defined to be $I_{\alpha+1}$; otherwise $I_{\alpha+1} = I_\alpha$.
3. If β is a limit ordinal and $\beta \leqslant \kappa$, then $I_\beta = \bigcup_{\alpha<\beta} I_\alpha$.

Then I_κ is a proper ideal of R; it is maximal, since any larger ideal J will contain some a_α that is not in I_κ, but this means $I_\alpha \cup \{a_\alpha\}$ must generate R, and thus $J = R$. □

The proper ideal I of R is a **prime ideal** if

$$xy \in I \And x \notin I \implies y \in I.$$

For example, every element n of $\mathbb{Z}$ generates the ideal (n), namely $\{nx \colon x \in \mathbb{Z}\}$. Then

$$x \in (n) \iff n \mid x.$$

Euclid's Lemma is that for all primes p,

$$p \mid xy \And p \nmid x \implies p \mid y,$$

that is, (p) is a prime ideal of $\mathbb{Z}$. But (0) is also a prime ideal of $\mathbb{Z}$, although 0 is not considered to be a prime number.

Any ring in which (0) is a prime ideal is an **integral domain.** In general, the proper ideal I of R is

- maximal, if and only if R/I is a field;
- prime, if and only if R/I is an integral domain.

Since all fields are integral domains, all maximal ideals are prime. The converse holds for *Boolean* rings, since the only Boolean integral domain has the two elements 0 and 1 (namely the solutions to $x^2 = x$) and is therefore the two-element field.

The **Prime Ideal Theorem** is that every proper ideal of a ring is included in a prime ideal. It is known that the Prime Ideal Theorem does not imply the Maximal Ideal Theorem in Zermelo–Fraenkel set theory. (Both the Prime and the Maximal Ideal Theorem become real theorems when the Axiom of Choice is introduced.) Moreover, the Maximal Ideal Theorem implies the Axiom of Choice [4].

We derived the Compactness Theorem from Łoś's Theorem, but this turns out to be stronger than needed for the job: for the following is true.

Theorem 20. *The Compactness Theorem is a consequence of the Prime Ideal Theorem.*

Proof. In proving Łoś's Theorem, in a certain signature we find a set T of sentences such that, for all sentences σ and τ, and all formulas φ in which only x occurs freely, the properties (10) on page 23 hold; moreover, T contains all logically true sentences of its signature. These properties ensure that T is the set of all sentences that are true in some structure, whose universe consists of the interpretations of the constant symbols in the signature. (This result can be called the **Canonical Model Theorem.**)

Given a set Γ of sentences such that every finite subset of Γ has a model, we want to enlarge Γ to a set like T above. We do this first by adding to the signature, for every φ as above, a new constant symbol, c_φ. Now let Γ' consist of the sentences in Γ along with, for every φ as above, the sentence $\exists x\ \varphi \to \varphi(c_\varphi)$. Then every finite subset of Γ' has a model.

This means, as it were, Γ' generates a proper filter of the set of sentences of its signature. Now, this is not quite true, because this set of sentences is not exactly a Boolean ring; but it becomes a Boolean ring when we replace each of its sentences with the *logical equivalence class* of the sentence. Two sentences σ and τ are **logically equivalent** if the sentence $\sigma \leftrightarrow \tau$ is true in all structures of its signature. As equipped with the operations $\wedge$, $\vee$, and $\neg$ given by

$$\begin{aligned}
[\sigma] \wedge [\tau] &= [(\sigma \wedge \tau)],\\
[\sigma] \vee [\tau] &= [(\sigma \vee \tau)],\\
\neg[\sigma] &= [\neg\sigma],
\end{aligned}$$

the set of logical equivalence classes of sentences of a signature is the **Lindenbaum algebra** of the signature.[8] A Lindenbaum algebra is a **Boolean algebra,** which is just a Boolean ring with different operations emphasized. Addition in the Boolean ring is

[8] Actually, for any list of variables, there is a Lindenbaum algebra of formulas in which only the variables on that list occur freely.

given by $[\sigma]+[\tau] = [((\sigma\vee\tau)\wedge\neg(\sigma\wedge\tau))]$; multiplication is $\wedge$. Then the image of Γ' under $\sigma \mapsto [\sigma]$ generates a proper filter, which is included in an ultrafilter $\mathscr{U}$ by the Prime Ideal Theorem. Then the set $\{\sigma\colon [\sigma] \in \mathscr{U}\}$ is of the desired form of T as above, and so it has a model, which is a model of Γ since $\Gamma \subseteq T$. □

We also have the converse. To prove this, we define the **diagram** of any structure $\mathfrak{A}$ to be the set of all *quantifier-free* sentences, with parameters from A, that are true in $\mathfrak{A}$.

Theorem 21. *The Prime Ideal Theorem is a consequence of the Compactness Theorem.*

Proof. Suppose I is a proper ideal of a ring R. We consider I as a singulary relation on R, and we introduce a new singulary relation symbol P. Let Γ consist of the sentences of the diagram of (R, I), together with, for all a and b in R, the sentences

$$a \in P \wedge b \in P \to a + b \in P,$$
$$a \in P \vee b \in P \to ab \in P,$$
$$1 \notin P,$$
$$a \in I \to a \in P.$$

If $(R', I', P') \models \Gamma$, then R' is not necessarily a ring; but R embeds in R', since the sentence $a \neq b$ is in Γ, and is therefore true in R', whenever a and b are distinct elements of R. So we may assume $R \subseteq R'$. In this case $I = I' \cap R$, and $P' \cap R$ is a proper prime ideal of R that includes I.

The set Γ does have a model, by the Compactness Theorem. For suppose Δ is a finite subset of Γ. Then only finitely many elements of R occur in the sentences of Δ. These elements generate a *countable* sub-ring S of R. Then the Maximal Ideal Theorem is true for this sub-ring, as in the proof of Theorem 19. In particular, since $I \cap S$ is a proper ideal of S, it is included in a maximal ideal, which is therefore a prime ideal, and this provides a model of Δ. $\square$

Finally, it is possible to derive the Axiom of Choice from Łoś's Theorem. Using obvious abbreviations, we have

$$\begin{tikzcd}
\text{ŁOŚ} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \text{AC} \arrow[r, Leftrightarrow] & \text{MAX} \arrow[d, Rightarrow] \\
\text{COM} \arrow[rr, Leftrightarrow] & & \text{PRI}
\end{tikzcd}$$

References

When I set out to learn nonstandard analysis for this course, I used the book of its creator, Abraham Robinson [7]. A professor had referred me to this book when I was starting graduate school; but I could not make sense of the book then. There are textbooks of nonstandard analysis, but I confess to not having sought them out. I have consulted two *calculus* textbooks for beginners that use the nonstandard approach: Keisler [6] and Henle and Kleinberg [3]. Most books on model theory will say something about ultraproducts; the ones that I have used the most are Hodges [5], followed by Chang and Keisler [2] and Bell and Slomson [1].

[1] J. L. Bell and A. B. Slomson. *Models and ultraproducts: An introduction.* North-Holland Publishing Co., Amsterdam, 1969. Reissued by Dover, 2006.

[2] C. C. Chang and H. J. Keisler. *Model theory.* North-Holland Publishing Co., Amsterdam, third edition, 1990. First edition 1973.

[3] James M. Henle and Eugene M. Kleinberg. *Infinitesimal calculus*. Dover Publications Inc., Mineola, NY, 2003. Reprint of the 1979 original [MIT Press, Cambridge, MA.

[4] Wilfrid Hodges. Krull implies Zorn. *J. London Math. Soc. (2)*, 19(2):285–287, 1979.

[5] Wilfrid Hodges. *Model theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.

[6] H. Jerome Keisler. *Elementary Calculus: An Infinitesimal Approach*. 2013. First edition 1976, second edition 1986, both by Prindle, Weber & Schmidt. Published online at `www.math.wisc.edu/~keisler/calc.html`.

[7] Abraham Robinson. *Non-standard analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1996. Reprint of the second (1974) edition. With a foreword by Wilhelmus A. J. Luxemburg. First edition 1965.