Nonstandard Analysis

Armin Straub

10-Dec 2007

Contents

Abstract

These are complementary notes for a *StuVo* talk given at TU Darmstadt, 10-Dec 2007. While this text is supposed to be informal in nature, any corrections as well as suggestions are very welcome.

We discuss some ways to add infinitesimals to our usual numbers, get acquainted with ultrafilters and how they can be used to construct nonstandard extensions, and then provide an axiomatic framework for nonstandard analysis. As basic working examples we present nonstandard characterizations of continuity and uniform continuity. We close with a short external look at the nonstandard integers and point out connections with p-adic integers.

Armin Straub
math@arminstraub.com

1 Introduction

1.1 Approaches to Getting Infinitesimals

One aspect of nonstandard analysis is that it introduces a new kind of numbers, namely infinitely large numbers (and hence infinitesimally small ones as well). This is something that may be philosophically challenging to a Platonist. It's not just that we are talking about infinity but its quality here is of a particular kind. When we have a never-ending process, like counting the natural numbers, we talk of *potential infinity*. This "counting never ends" can be formalized as

$$\forall x \in \mathbb{N} \exists y \quad y > x,$$

and is usually philosophically well accepted by most people. By contrast, the *actual infinity* of nonstandard numbers corresponds to a statement as

$$\exists y \forall x \in \mathbb{N} \quad y > x,$$

and usually troubles people more easily. On the other hand, mathematicians like Euler, Leibniz or Newton had a very good conceptual understanding of infinitesimals. It wasn't before 1961, however, that they were rigorously introduced into mathematics as formal objects by Robinson, see also [Robinson, 1967]. But Robinson's approach does more than just introduce infinitesimals and therefore justifying the non-Archimedean approach prevailing before the ε-δ infection by Weierstrass around 1850. It enriches all infinite structures as well and turns out to be very suitable for applications in fields like topology or probability theory where a metric concept is absent in the first place.

In 1934 Skolem pointed out that the natural numbers can't be characterized as a first-order theory. The unusual structures satisfying the same properties as the natural numbers are therefore called *nonstandard models* of arithmetic. Since we learned a lot about (in)completeness in the previous talk we start by showing how logic leads to nonstandard models. In the sequel, however, we'll try to be more "constructive".

Example 1.1. Recall Gödel's incompleteness theorem. In one guise it states that every consistent first-order theory that includes arithmetic cannot prove its own consistency. Starting with the usual axioms of Peano arithmetic[1.1], PA for short, we can add its inconsistency $\neg\mathrm{Con}(\mathrm{PA})$ as an axiom (we're very informal here). If PA was consistent then so is $\mathrm{PA} \wedge \neg\mathrm{Con}(\mathrm{PA})$. Obviously, a model for the latter can't be our usual (or standard) natural numbers. Hence we must have obtained a nonstandard model of the natural numbers. But well, that was kind of magic. What do we know about this model after all?

Example 1.2. There is another common way to get nonstandard models with the help of logic and the so-called compactness theorem. Gödel's compactness theorem pretty much states that a set of axioms has a model if and only if every finite subset of these axioms has a model. Let's start with PA arithmetic again and consider the theory obtained by adding an element c to the signature of PA as well as the following axioms

$$c > 1, \quad c > 1 + 1, \quad \ldots$$

Every finite subset of these axioms clearly has a model, namely the natural numbers $\mathbb{N}$ with a big element $c \in \mathbb{N}$ exhibited. Thus the compactness theorem asserts that there also exists a model satisfying all the above axioms. Forgetting the special role of c this gives a nonstandard model for PA. While this was still magic in the end, we at least have an infinitely large element c at hand.

1.1. Note that in PA the second-order induction axiom is replaced by a first-order induction schema which consists of one axiom for each of the countably many formulas in PA.

Remark 1.3. We can't resist to demonstrate at least one cool and inspiring implication based on the compactness theorem. Namely, if a statement S (first-order, of course) is true for every field of characteristic 0 then it is true for every field of characteristic p as long as p is big enough. Just note that the assumption implies that

$$\neg S, \quad \text{field axioms}, \quad 1+1\neq 0, \quad 1+1+1\neq 0, \quad \ldots$$

is not satisfiable. By the compactness theorem a finite subset of these statements isn't satisfiable already.

Remark 1.4. By the way, ZF proves that Gödel's compactness theorem as well as Gödel's completeness theorem are equivalent to the ultrafilter theorem. So in the case that you despise our later use of the ultrafilter theorem to construct nonstandard models, you should feel uncomfortable with this part of mathematics as well.

On the other hand, it is easy to introduce infinitely large objects that serve some specific but limited purpose. We present some examples in the hope that they will increase our appreciation of the nonstandard world that Robinson invited us to.

Example 1.5. For example, see [O'Donovan, 2005], we can extend the ordered abelian group $\mathbb{Z}$ of integers to $\mathbb{Z}^2$ by adding a second coordinate and using lexicographic ordering. Then elements with a nonzero second component could be called infinite since each of them is indeed larger than any copy $(n,0)$ of a usual integer $n\in\mathbb{Z}$. While we are able this way to preserve the group structure and the ordering of $\mathbb{Z}$, chances are that these infinite numbers won't make us too happy. They can't even be multiplied.

Example 1.6. We can extend the previous example to allow for multiplication. Note that the previous construction could have been obtained in the following way: Starting with $\mathbb{Z}$ and a symbol c we interpret c as being larger than any integer. Then we fill in all the additional stuff which is needed to get addition to work like $2c+1$. What we end up with is $\mathbb{Z}+\mathbb{Z}\,c\cong\mathbb{Z}^2$, just as before. Since we also want to have multiplication, we again fill in all the additional stuff needed for multiplication, for instance c^2+1. It won't come as a surprise that we get the polynomial ring $\mathbb{Z}[c]$. But, oh well, true happiness is hard to achieve, and this world of infinite numbers still isn't good enough for us. We can't even talk about 2^c. As before we could keep going in the same fashion but the structure will always be too small (even if we kept doing that infinitely often, see [Kossak, 1996] where the inherent and provable complexity of "true" infinitesimals is discussed).

At this point we can already feel that there is going to be some trouble with the nonstandard worlds of integers. Namely, they are going to be very large. In fact, every nonstandard model of the integers that obeys the transfer principle is going to be uncountably large. That's why use of the axiom of choice or a slightly weaker axiom will be needed to shed some light.

Example 1.7. Another way to get infinitesimals into play is to consider the ring of dual numbers

$$R\triangleq\mathbb{R}[\varepsilon]/(\varepsilon^2),$$

that is we introduce a new quantity ε that squares to 0. While such a property of a number might seem strange at first, this is actually found in computers when they do numerical calculations. Because of the finite precision there is a smallest positive number which hence has to square to 0. You can then write

$$(x+\varepsilon)^2=x^2+2x\varepsilon.$$

Similarly, for any polynomial we get $f(x+\varepsilon)=f(x)+\ldots\varepsilon$ and we can define ... as the derivative of $f(x)$. To handle a broader class of functions, including the trigonometric ones, in this way we can postulate that

$$\exp(\varepsilon)=1+\varepsilon.$$

Then also $\cos(\varepsilon) = 1$ and $\sin(\varepsilon) = 0$, so that

$$\cos(x+\varepsilon) = \cos(x)\cos(\varepsilon) - \sin(x)\sin(\varepsilon) = \cos(x) - \sin(x)\,\varepsilon$$

whence the derivative of $\cos(x)$ is $-\sin(x)$. This approach has actually been used for example in a freshmen calculus class as described in [Hoorn, 2007]. It can be adapted to handle higher order derivatives or multiple variables. An application to elliptic curves can be found in [Belding, 2007].

1.2 The Use of Infinitesimals

There are several reasons why one would wish to have infinitesimal numbers at hand together with a rigorous foundation. We only state a few.

Resurrection of the Good Old Days. With infinitesimal numbers the arguments of mathematicians like Leibniz, Newton or Euler can finally be made rigorous.

Better Understanding of the Standard World. As quoted from [Kossak, 1996], "we can compute anything we want without irrationals, but we could never understand geometry and calculus without them". In the same spirit nonstandard thinking can illuminate our understanding of the standard world.

Improving the Teaching of Mathematics. Indicated by the fact that mathematicians like Euler used infinitesimals in their reasoning (successfully despite the lack of a firm foundation), it is often argued that a nonstandard approach to teaching calculus may be easier to understand for students. At any rate, there are excellent books, for instance [Keisler, 2000], that demonstrate that calculus can be taught using infinitesimals instead of sequences or ε-δ concepts. As an informal example, continuity of $x \mapsto x^2$ is derived as

$$(x+d)^2 = x^2 + 2dx + d^2 \approx x^2.$$

Enrichment of the Mathematical Environment. Nonstandard objects are worth considering in their own right. We'll for example take a closer look at the hyperintegers ${}^*\mathbb{Z}$ at the end of this text. For another example consider ${}^*\mathbb{R} \times {}^*\mathbb{R}$. It forms a model for plane Euclidean geometry. But when we restrict to finite points in both coordinates we get a plane, called the Dehn plane, that violates the parallel postulate (for instance a line through $(0,1)$ with infinitesimal slope won't meet the x-axis).

Automatic ε-δ Management. As illustrated in [Tao, 2007] the use of nonstandard analysis can make arguments more succinct and understandable. This is in particular true for arguments that make heavy use of interdependent ε_i's and δ_i's.

2 Diving In

2.1 Limits and Ultrafilters

This section is heavily based on [Tao, 2007]. Given a nice sequence (x_n) we can talk about its limit $\lim x_n$ (nice meaning convergent here). This limit can be interpreted, as Tao does, as the outcome of a *voting system*. Every natural number n votes for the candidate x_n, and the voting system then chooses $\lim x_n$ to be the winner. The limit has several properties, including:

(a) lim is an *algebra homomorphism*, that is

$$\begin{aligned} \lim a_n + b_n &= \lim a_n + \lim b_n, \\ \lim c a_n &= c \lim a_n, \\ \lim a_n b_n &= \lim a_n \lim b_n. \end{aligned}$$

(b) lim is *bounded* as follows

$$\inf x_n \leqslant \lim x_n \leqslant \sup x_n.$$

(c) lim is *non-principal*, that is $\lim x_n = \lim y_n$ whenever the sequences (x_n) and (y_n) differ at finitely many indices only.

(d) lim is *shift-invariant*, that is for all shifts $h \in \mathbb{N}$ we have

$$\lim x_{n+h} = \lim x_n.$$

Remark 2.1. Notice that the shift-invariance of lim actually implies that lim is non-principal. However, boundedness, and the non-principality alone suffice to characterize lim . To see this, take a sequence (x_n) that converges to x_∞. For any ε we find $h \in \mathbb{N}$ such that after setting the first h values of (x_n) to x_∞ we have

$$x_\infty - \varepsilon < \inf x_n \leqslant \sup x_n < x_\infty + \varepsilon$$

whence the only candidate for $\lim x_n$ is x_∞.

Can we give a definition of a limit that satisfies the above conditions and that applies to more than just the usually convergent sequences? More specifically, can we define such a limit for merely bounded sequences?

Example 2.2. Consider the sequence $(x_n) = (0, 1, 0, 1, 0, 1, ...)$. It is certainly bounded, and by the algebra homomorphism property of lim , we have $\lim x_n = (\lim x_n)^2$. Hence the only candidates for limits are 0 and 1. On the other hand, using the algebra homomorphism property and the shift-invariance we get the contradiction

$$\lim x_n = \lim x_{n+1} = \lim 1 - x_n = 1 - \lim x_n$$

which would imply that $\lim x_n = 1/2$.

So we see that we have to abandon either the algebra homomorphism property or the shift-invariance of lim . We dismiss shift-invariance. In the voting interpretation this means that we give up the fairness of the election (which means arbitrary choices in the voting system which, as Tao points out, can be taken as a first hint that the Axiom of Choice might get involved).

Suppose we found such an extension of lim to bounded sequences. Let's consider what happens when we apply lim to boolean sequences only, that is to sequences that take only values 0 and 1. By the algebra homomorphism property (as before) we find that for a boolean sequence (x_n) the only candidates for $\lim x_n$ are 0 and 1. Such a sequence can be represented by its indicator set $A = \{n \in \mathbb{N}: x_n = 1\}$, and we can consider the collection F of those indicator sets A such that $\lim x_n = 1$ (those voters that can decide an election by voting in unison). Using the homomorphism property as well as boundedness and non-principality we easily derive the following properties of F:

(a) F is an upper set, that is

$$A \in F, B \supset A \implies B \in F.$$

(b) F is closed under finite intersections,

$$A, B \in F \implies A \cap B \in F.$$

(c) F is dichotomous, that is for $A \subset \mathbb{N}$ exactly one of the following holds

$$A \in F \quad \vee \quad A^c \in F.$$

(d) F is non-principal meaning that changing a set A by finitely many elements doesn't affect whether $A \in F$.

The first two conditions state that F is a *filter*, the third adds that it is in fact an *ultrafilter* and the last one exclude the *principal filters* (which are just all the sets containing a particular index, and which correspond to a dictator in the voting interpretation).

On the other hand, given an ultrafilter F on a set I we can define $F\text{-lim}\, x_n$ for a sequence (x_n) taking values in $\{0,1\}$ by setting

$$F\text{-lim}\, x_n = 1 \quad \Longleftrightarrow \quad \{n \in I: x_n = 1\} \in F,$$

and $F\text{-lim}\, x_n = 0$ otherwise. Then $F\text{-lim}$ satisfies the properties discussed above. Whenever we have a property $p(n)$, like $x_n = 1$ here, we say that $p(n)$ holds w.r.t. F if $\{n \in I: p(n)\} \in F$. The properties of an ultrafilter translate into:

(a) (Modus-Ponens) If $p(n) \Rightarrow q(n)$ then

$$p(n) \text{ w.r.t. } F \quad \Longrightarrow \quad q(n) \text{ w.r.t. } F.$$

(b) The closure of F under finite intersections implies that

$$p(n) \text{ w.r.t. } F, q(n) \text{ w.r.t. } F \quad \Longrightarrow \quad p(n) \wedge q(n) \text{ w.r.t. } F.$$

(c) (Law of the Excluded Middle) Exactly one of the following holds,

$$p(n) \text{ w.r.t. } F \quad \vee \quad \neg p(n) \text{ w.r.t. } F.$$

Now, given any bounded sequence (x_n) we can define $F\text{-lim}\, x_n$ as the unique x such that

$$|x_n - x| < \varepsilon \text{ w.r.t. } F$$

for all $\varepsilon > 0$.

But wait! We actually obtained an ultrafilter F only under the assumption that we found an extension of lim to bounded sequences taking values in $\{0,1\}$. So do ultrafilters exist after all?

Theorem 2.3. ***(UFT)*** *Every filter is contained in an ultrafilter.*

Proof. Well, this is independent from ZF but it can be proved very straightforward using Zorn's Lemma (which is equivalent to the Axiom of Choice). Note, however, that the ultrafilter theorem is strictly weaker than both. □

To find a non-principal ultrafilter we can just start with the filter of cofinite sets. We now present two statements which assert that actually there are incredibly many ultrafilters on an infinite set (on a finite set X there are only the principal ultrafilters and hence only $|X|$ many).

Theorem 2.4. *An infinite set X allows $2^{2^{|X|}}$ many (non-principal) ultrafilters.*

Proof. More is not possible any way. The hard part is to construct a set S of $2^{|X|}$ subsets of X such that any Boolean combination

$$s_1 \cap s_2 \cap \ldots \cap s_n \neq \emptyset, \quad s_i \in S \vee s_i^c \in S$$

is nonempty. Then any subset $T \subset S$ gives rise to a filter basis

$$T \cup \{s^c: s \in S \backslash T\}$$

which in turn define distinct ultrafilters. Of course, there are $2^S = 2^{2^{|X|}}$ many such subsets T. Finally, note that are only exactly $|X|$ many principal ultrafilters on X. □

Corollary 2.5. *There are non-isomorphic non-principal ultrafilters on any infinite set X (two ultrafilters being isomorphic if one arises from the other by a bijection on X).*

Proof. There are no more than $2^{|X|}$ many bijections on X but $2^{2^{|X|}}$ many non-principal ultrafilters. □

2.2 A First Nonstandard Peak

For the field extension from $\mathbb{Q}$ to $\mathbb{R}$ what you do is consider all sequences $\mathbb{Q}^{\mathbb{N}}$, take the ring of Cauchy sequences, and factor by the maximal ideal $\{a\colon \lim a = 0\}$. Now, what do we get if we start with $\mathbb{R}$, consider the ring of all sequences $\mathbb{R}^{\mathbb{N}}$, and factor by a maximal ideal? Ok, of course we get a field but what else can we say. Is it for example going to be ordered again? And what do maximal ideals of $\mathbb{R}^{\mathbb{N}}$ look like?

Theorem 2.6. *Ultrafilters on I are in a* 1-1 *correspondence with maximal ideals of $\mathbb{R}^I$ via*

$$F \mapsto \{(x_n) \in \mathbb{R}^I \colon x_n = 0 \text{ w.r.t. } F\}.$$

Whenever F is a principal ultrafilter on $\mathbb{N}$ then $\mathbb{R}^{\mathbb{N}}/F \cong \mathbb{R}$ and we don't win anything. That's why we need a non-principal ultrafilter F. Those exist by the ultrafilter theorem though we'll never be able to describe one explicitly.

Let F be an ultrafilter, and set ${}^*\mathbb{R} = \mathbb{R}^{\mathbb{N}}/F$. Elements of ${}^*\mathbb{R}$ are equivalence classes of sequences where two sequences (x_n) and (y_n) are identified whenever $x_n = y_n$ w.r.t. F. We denote the equivalence class of a sequence (x_n) by $[x_n]$. $\mathbb{R}$ is embedded in ${}^*\mathbb{R}$ by identifying a real number x with the sequence that is constantly x. We write $[x_n] \leqslant [y_n]$ whenever

$$x_n \leqslant y_n \text{ w.r.t. } F$$

which gives a well-defined binary relation on ${}^*\mathbb{R}$.

Corollary 2.7. ${}^*\mathbb{R}$ *is a totally ordered field.*

Proof. By the preceding theorem, ${}^*\mathbb{R}$ is the quotient of the ring $\mathbb{R}^{\mathbb{N}}$ by a maximal ideal and hence a field. Alternatively, we can check the axioms for a field, use their validity in $\mathbb{R}$ and then use the logic w.r.t. F to deduce them for ${}^*\mathbb{R}$. We only prove the existence of inverses. Given a sequence (x_n) such that $[x_n] \neq 0$, we define (y_n) by $y_n = x_n^{-1}$ if $x_n \neq 0$ and $y_n = 0$ otherwise. Then for all n

$$x_n \neq 0 \implies x_n y_n = 1.$$

Thus the logic w.r.t. F implies

$$\begin{aligned} [x_n] \neq 0 &\implies x_n \neq 0 \text{ w.r.t. } F \\ &\implies x_n y_n = 1 \text{ w.r.t. } F \\ &\implies [x_n][y_n] = 1 \end{aligned}$$

whence $[y_n]$ is the inverse of $[x_n]$.

That $\leqslant$ is a total order is again just checking the axioms for a total order. For instance, given sequences (x_n) and (y_n) we know that for each n either $x_n < y_n$, $x_n = y_n$, or $x_n > y_n$. Thus it follows that we have exactly one of the cases

$$x_n < y_n \text{ w.r.t. } F, \quad x_n = y_n \text{ w.r.t. } F, \quad x_n > y_n \text{ w.r.t. } F,$$

that is exactly one of $[x_n] < [y_n]$, $[x_n] = [y_n]$, $[x_n] > [y_n]$.

Compatibility with the field operations follows from the logic w.r.t. F as well. Let for example $(x_n), (y_n), (z_n)$ be sequences. Since $\mathbb{R}$ is an ordered field we have for each n that

$$x_n \leqslant y_n \wedge z_n \geqslant 0 \implies x_n z_n \leqslant y_n z_n.$$

Thus we employ the logic w.r.t. F to get

$$\begin{aligned}[x_n] \leqslant [y_n] \ \wedge \ [z_n] \geqslant 0 &\implies x_n \leqslant y_n \text{ w.r.t. } F \ \wedge \ z_n \geqslant 0 \text{ w.r.t. } F \\ &\implies x_n \leqslant y_n \wedge z_n \geqslant 0 \text{ w.r.t. } F \\ &\implies x_n z_n \leqslant y_n z_n \text{ w.r.t. } F \\ &\implies [x_n][z_n] \leqslant [y_n][z_n].\end{aligned}$$

Basically what we observe is that statements in $\mathbb{R}$ pass over to statements in $^*\mathbb{R}$. □

Our initial goal was to extend the real numbers so that we have infinitesimals at hand. Did we achieve this goal? Well, consider the sequence $(x_n) = (1, 1/2, 1/3, \ldots)$. If F is a non-principal ultrafilter then

$$(\forall \varepsilon \in \mathbb{R}_+) \ [x_n] < \varepsilon$$

because only finitely many of the x_n are larger than any fixed ε. Hence $[x_n]$ is an infinitesimal.

So we succeeded in extending the real numbers as an ordered field to the nonstandard field $^*\mathbb{R}$ which contains infinitesimal and infinite numbers. We could go on from here, for instance lift functions $\mathbb{R} \to \mathbb{R}$ to functions $^*\mathbb{R} \to {}^*\mathbb{R}$, make explicit what kind of statements over $\mathbb{R}$ transfer to statements over $^*\mathbb{R}$ and lots more. This *ultrapower* construction is described in detail in [Landers and Rogge, 1994]. Here we'll switch to an axiomatic approach now that hopefully some credibility in the possibility of such a construction is won.

2.3 An Axiomatic Approach

Starting with our usual mathematical world, we want to construct a nonstandard extension of this world, in which it is more fun to fool around. Since we know about the sneaky trap of playing with the set of all sets, we cave in and reduce our usual mathematical world to be based on the set of all "reasonable" sets (in a sense that most mathematicians stay reasonable most of the time). More details of the axiomatic approach exposited here and how to work with it can be found in [Salbany and Todorov, 1999].

Definition 2.8. ***(Superstructure)*** *Let S be a set, and define $V_0(S) = S$ and inductively $V_{n+1}(S) = V_n(S) \cup \mathcal{P}(V_n(S))$. The* superstructure $V(S)$ *of S is*

$$V(S) = \bigcup_{n \in \mathbb{N}} V_n(S).$$

We usually choose $S = \mathbb{R}$. The superstructure $V(\mathbb{R})$ entails lots of the mathematical world that we use to work in (recall that tuples, relations, functions and pretty much anything else can be encoded as a set). Our goal is to extend this superstructure to some strictly larger one, namely $V(^*\mathbb{R})$ such that true statements in $V(\mathbb{R})$ transfer to true statements in $V(^*\mathbb{R})$. $^*\mathbb{R}$ can be constructed as in the preceding section, and some details need to be considered to define how to embed the whole of $V(\mathbb{R})$. We have to be a bit careful about what statements we can transfer. These are formulas in which quantification is always bounded, that is quantification appears in the forms $(\forall x \in A)$ and $(\exists x \in A)$ only. The forms $(\forall x)$ or $(\exists x)$ are said to be unbounded and are not allowed. Such formulas are called *bounded quantifier formulas*. We write $\Phi(x_1, \ldots, x_n)$ to denote a formula with free variables $x_1, \ldots, x_n$.

Definition 2.9. ***(Nonstandard Model)*** *Let S be an infinite set. A map*

$$^*\colon V(S) \to V(^*S), \quad A \mapsto {}^*A$$

is a nonstandard extension *if it satisfies the following three axioms.*

Extension Principle. *$^*s = s$ for all $s \in S$.*

Transfer Principle. *For every bounded quantifier formula $\Phi(x_1, \ldots, x_n)$ and all $A_1, \ldots, A_n \in V(S)$,*

$$\begin{aligned} &\Phi(A_1, \ldots, A_n) \text{ is true in } \mathcal{L}(V(S)) \\ \Longleftrightarrow\ &\Phi(^*A_1, \ldots, {}^*A_n) \text{ is true in } \mathcal{L}(V(^*S)). \end{aligned}$$

Saturation Principle. *The extension is nontrivial, that is $^*S \supsetneq S$.*

To get a feeling how to work in the setting of a nonstandard extension we prove some basic properties.

Proposition 2.10. ***(Basic Properties of *)*** *Let * be a nonstandard extension.*

(a) ** is injective.*

(b) ** preserves the usual boolean operations, i.e. for sets $A, B \in V(S)$*

$$^*(A \cup B) = {}^*A \cup {}^*B, \quad ^*(A \cap B) = {}^*A \cap {}^*B, \quad ^*(A \backslash B) = {}^*A \backslash {}^*B.$$

(c) *$^*\{a_1, \ldots, a_n\} = \{^*a_1, \ldots, {}^*a_n\}$ for all $a_i \in V(S)$.*

(d) *If $f\colon A \to B$ is a function between sets $A, B \in V(S)$ then $^*f$ is a function $^*A \to {}^*B$.*

Proof.

(a) Let $^*A = {}^*B$ for $A, B \in V(S)$. By transfer $A = B$.

(b) We prove the claim for unions. The other statements can be handled the same way. Let $C = A \cup B$. Choose n such that $A, B, C \subset V_n(S)$. Then the following is true in $\mathcal{L}(V(S))$

$$(\forall x \in V_n(S))\ (x \in C \quad \Leftrightarrow \quad (x \in A \vee x \in B))$$

and transfer yields

$$(\forall x \in {}^*V_n(S))\ (x \in {}^*C \quad \Leftrightarrow \quad (x \in {}^*A \vee x \in {}^*B))$$

in $\mathcal{L}(^*V(S))$. Since (also by transfer) $^*A$, $^*B$, $^*C \subset {}^*V_n(S)$, the latter amounts to saying $^*C = {}^*A \cup {}^*B$.

(c) Let $A = \{a_1, \ldots, a_n\}$. Then

$$(\forall a \in A) a = a_1 \vee \ldots \vee a = a_n$$

is true. By transfer

$$(\forall a \in {}^*A) a = {}^*a_1 \vee \ldots \vee a = {}^*a_n$$

is true as well which shows that $^*\{a_1, \ldots, a_n\} = {}^*A = \{^*a_1, \ldots, {}^*a_n\}$.

(d) Since we didn't even discuss how to encode a function as a set we won't provide a proof but hope for the reader's belief (or willingness to work it out as an exercise). □

2.4 Basic Usage

As for the explicit construction based on ultrafilters we have the following.

Theorem 2.11. *${}^*\mathbb{R}$ is a totally ordered field which properly extends $\mathbb{R}$.*

Proof. We have to check lots of properties of ${}^*\mathbb{R}$ but all of them follow from the respective property of $\mathbb{R}$. For instance the commutativity of addition in ${}^*\mathbb{R}$ formalized as

$$(\forall x, y \in {}^*\mathbb{R}) x \,{}^*\!+ y = y \,{}^*\!+ x$$

follows from transfer of

$$(\forall x, y \in \mathbb{R}) x + y = y + x. \qquad \square$$

Since ${}^*\mathbb{R}$ was build to contain infinitely large numbers, we don't expect it to be Archimedean. But why doesn't this property follow by transfer as well? Because the statement

$$(\forall x \in \mathbb{R})(\exists n \in \mathbb{N}) n > x$$

transfers to

$$(\forall x \in {}^*\mathbb{R})(\exists n \in {}^*\mathbb{N}) n \,{}^*\!\!> x$$

which just asserts that ${}^*\mathbb{R}$ is "hyper"-Archimedean.

Since it shouldn't lead to any confusion we'll just write $+$ instead of ${}^*+$, and similarly for other symbols like $<$, $|\cdot|$ and so on. We introduce the following intuitive language.

Definition 2.12. *Let $x, y \in {}^*\mathbb{R}$.*

- *x is called* finite *if $|x| \leqslant n$ for some $n \in \mathbb{N}$, and* infinite *otherwise.*
- *x is called* infinitesimal *if $1/x$ is infinite.*
- *x and y are said to be* infinitesimally close*, denoted by $x \approx y$, if $x - y$ is infinitesimal.*

We just note that intuitively true statements like "finite + finite = finite" are just as easily derived from these definitions. But one question lingers: Do there exist infinite numbers at all, or are we talking in a sophisticated way about empty attributes here?

Theorem 2.13. *${}^*\mathbb{N} - \mathbb{N} \neq \emptyset$ and every $h \in {}^*\mathbb{N} - \mathbb{N}$ is infinite.*

Proof. Let $h \in {}^*\mathbb{N} - \mathbb{N}$. To prove that h is infinite we have to show that $h > n$ for all $n \in \mathbb{N}$. We prove $h > n$ inductively. Clearly, $h > 1$ since $1 \in \mathbb{N}$. Suppose that $h > n$. By transfer, $h \geqslant n + 1$ (consider the formula $(\forall m \in \mathbb{N}) m > n \Rightarrow m \geqslant n + 1$ and recall that ${}^*n = n$). But $n + 1 \in \mathbb{N}$ and hence $h > n + 1$.

We now prove that ${}^*\mathbb{N} - \mathbb{N} \neq \emptyset$. Assume otherwise that ${}^*\mathbb{N} = \mathbb{N}$. Since the rationals are dense in $\mathbb{R}$ we have the following true statement

$$(\forall x \in \mathbb{R})(\forall n \in \mathbb{N})(\exists r, s, t \in \mathbb{N}) \left| x - \frac{r - s}{t} \right| < \frac{1}{n}$$

which under the assumption ${}^*\mathbb{N} = \mathbb{N}$ transfers to

$$(\forall x \in {}^*\mathbb{R})(\forall n \in \mathbb{N})(\exists r, s, t \in \mathbb{N}) \left| x - \frac{r - s}{t} \right| < \frac{1}{n}.$$

Let $x \in {}^*\mathbb{R}$. We thus find a sequence $x_k \in \mathbb{Q}$ such that $|x - x_n| < 1/n$. This makes (x_n) a Cauchy sequence which hence converges to some $x_\infty \in \mathbb{R}$. Note that

$$|x - x_\infty| \leqslant |x - x_n| + |x_n - x_\infty| \to 0 \quad \text{as } n \to \infty.$$

In particular, $|x - x_\infty| < 1/n$ for all $n \in \mathbb{N} = {}^*\mathbb{N}$. The transfer of

$$(\forall x, y \in \mathbb{R})((\forall n \in \mathbb{N})|x - y| < 1/n) \implies x = y$$

now yields that $x = x_\infty \in \mathbb{R}$. But then ${}^*\mathbb{R} \subset \mathbb{R}$ which is a contradiction. □

This language is very nice for defining notions as limits, accumulation points, continuity or uniform continuity.

Theorem 2.14. *Let $(x_n) \subset \mathbb{R}$ be a sequence and $x \in \mathbb{R}$. The following are equivalent.*

- $\lim x_n = x$ *if and only if* ${}^*x_h \approx x$ *for all* $h \in {}^*\mathbb{N} - \mathbb{N}$.
- x *is an accumulation point of* (x_n) *if and only if* ${}^*x_h \approx x$ *for some* $h \in {}^*\mathbb{N} - \mathbb{N}$.

Proof. We only prove the first statement. Suppose first that $\lim x_n = x$, and let $h \in {}^*\mathbb{N} - \mathbb{N}$. Then for any $\varepsilon \in \mathbb{R}_+$ we find n_0 such that

$$(\forall n \in \mathbb{N})\ n > n_0 \implies |x_n - x| < \varepsilon.$$

This transfers to

$$(\forall n \in {}^*\mathbb{N})\ n > n_0 \implies |{}^*x_n - x| < \varepsilon.$$

Since $h > n_0$ and ε was arbitrary this implies that $1/|{}^*x_h - x| > n$ for any $n \in \mathbb{N}$, and hence ${}^*x_h \approx x$ as claimed.

Now suppose that ${}^*x_h \approx x$ for all $h \in {}^*\mathbb{N} \backslash \mathbb{N}$. Let $\varepsilon \in \mathbb{R}_+$. Then

$$(\exists n_0 \in {}^*\mathbb{N})(\forall n \in {}^*\mathbb{N})\ n > n_0 \implies |{}^*x_n - x| < \varepsilon$$

holds, and hence by transfer

$$(\exists n_0 \in \mathbb{N})(\forall n \in \mathbb{N})\ n > n_0 \implies |x_n - x| < \varepsilon$$

which proves that $\lim x_n = x$. □

Theorem 2.15. *Let $D \subset \mathbb{R}$, $x_0 \in D$, and $f\colon D \to \mathbb{R}$ be a function. The following are equivalent.*

- f *is continuous in* x_0.
- $(\forall x \in {}^*D)\ x \approx x_0 \implies {}^*f(x) \approx f(x_0)$.

Proof. The proof is along the lines of the previous one. We give it anyway.

First, let f be continuous in x_0. Let $\varepsilon \in \mathbb{R}_+$. We find $\delta \in \mathbb{R}_+$ such that

$$(\forall x \in D)\ |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

This transfers to

$$(\forall x \in {}^*D)\ |x - x_0| < \delta \implies |{}^*f(x) - f(x_0)| < \varepsilon.$$

Let $x \in {}^*D$ such that $x \approx x_0$. The condition $|x - x_0| < \delta$ is then trivially satisfied for any choice of ε. We thus have that $|{}^*f(x) - f(x_0)| < \varepsilon$ for all $\varepsilon \in \mathbb{R}_+$, and hence ${}^*f(x) \approx f(x_0)$.

Now suppose that ${}^*f(x) \approx f(x_0)$ for all $x \in {}^*D$ such that $x \approx x_0$. Let $\varepsilon \in \mathbb{R}_+$. Then

$$(\exists \delta \in {}^*\mathbb{R}_+)(\forall x \in {}^*D)\ |x - x_0| < \delta \implies |{}^*f(x) - f(x_0)| < \varepsilon$$

holds, and hence by transfer

$$(\exists \delta \in \mathbb{R}_+)(\forall x \in D)\ |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon$$

which proves that f is continuous at x_0. □

Theorem 2.16. *Let $D \subset \mathbb{R}$, $x_0 \in D$, and $f: D \to \mathbb{R}$ be a function. The following are equivalent.*

- *f is uniformly continuous.*
- *$(\forall x, y \in {}^*D)\ x \approx y \implies {}^*f(x) \approx f(y)$.*

Proof. The proof is almost exactly the previous one. Again, we give it anyway.

First, let f be uniformly continuous. Let $\varepsilon \in \mathbb{R}_+$. We find $\delta \in \mathbb{R}_+$ such that

$$(\forall x, y \in D)\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

This transfers to

$$(\forall x, y \in {}^*D)\ |x - y| < \delta \implies |{}^*f(x) - {}^*f(y)| < \varepsilon.$$

Let $x, y \in {}^*D$ such that $x \approx y$. The condition $|x - y| < \delta$ is then trivially satisfied for any choice of ε. We thus have that $|{}^*f(x) - {}^*f(y)| < \varepsilon$ for all $\varepsilon \in \mathbb{R}_+$, and hence ${}^*f(x) \approx {}^*f(y)$.

Now suppose that ${}^*f(x) \approx {}^*f(y)$ for all $x, y \in {}^*D$ such that $x \approx y$. Let $\varepsilon \in \mathbb{R}_+$. Then

$$(\exists \delta \in {}^*\mathbb{R}_+)(\forall x, y \in {}^*D)\ |x - y| < \delta \implies |{}^*f(x) - {}^*f(y)| < \varepsilon$$

holds, and hence by transfer

$$(\exists \delta \in \mathbb{R}_+)(\forall x, y \in D)\ |x - y| < \delta \implies |f(x) - f(y)| < \varepsilon$$

which proves that f is uniformly continuous. □

All these examples show that using nonstandard language we can define standard notions not only in an intuitive way but even in a more succinct form that allows to focus on the essentials.

2.5 The Source of Nonstandard Strength

Let's take another look at the *hypernatural numbers* ${}^*\mathbb{N}$. We proved before that ${}^*\mathbb{N} - \mathbb{N} \neq \emptyset$ and every $h \in {}^*\mathbb{N} - \mathbb{N}$ is infinite. In particular, for infinite h also $h - 1 \in {}^*\mathbb{N} - \mathbb{N}$ is infinite. Therefore, ${}^*\mathbb{N} - \mathbb{N}$ has no minimal element!

What is going on here? Any subset of $\mathbb{N}$ has a minimal element. Shouldn't this by transfer be true for ${}^*\mathbb{N}$ as well? To understand this, look what happens when we transfer statements from $\mathbb{R}$ to ${}^*\mathbb{R}$. For example

$$(\forall m \in A)...(\exists n \in B)...$$

transfers to

$$(\forall m \in {}^*A)...(\exists n \in {}^*B)...$$

Objects of the form ${}^*A$ are called *standard*, and we observe that after transfer we talk about elements of standard objects ${}^*A$ only. Those are called *internal*. Since ${}^*\{A\} = \{{}^*A\}$ standard objects are internal.

Remark 2.17. It's not hard to see that the set $V_{\text{int}}({}^*S)$ of internal objects is

$$V_{\text{int}}({}^*S) = \bigcup_{n \in \mathbb{N}} {}^*V_n(S).$$

Objects that are not internal are called *external*. Vaguely, external objects are those that exist in the nonstandard model but which can not be accessed by transfer. $^*\mathbb{N} - \mathbb{N}$ is an example of an external object. While this might seem like an obstruction it actually is source for the strength of working with nonstandard models. What would we win when it behaved exactly the same as our standard world? No, it's great that it doesn't, and essentially what we gain is the possibility to talk about an object being standard, that is of the form $^*A$, or not. In ZF we only have the binary relation $\in$ but in the nonstandard model we also have this unary relation st that expresses whether an object is standard.

Remark 2.18. In fact, nonstandard analysis can be build axiomatically around this unary relation st and some axioms how to work with it. This approach was first used by Nelson who introduced his internal set theory IST as an extension of ZF. See for instance [Nelson, 1987] where he uses IST to develop probability theory.

2.6 Stronger Nonstandard Models

The saturation property which asserts that $^*S \supsetneq S$ can be reformulated in the following equivalent way, see [Landers and Rogge, 1994]. For any countable family of standard sets $(^*A_n)_{n\in I}$ that has the finite intersection property,

$$\bigcap_{n\in I} {}^*A_n \neq \emptyset.$$

Example 2.19. Contrast this with the situation in the standard world. For instance the family $(\mathbb{N} - \{1, ..., n\})_{n\in\mathbb{N}}$ clearly has the finite intersection property but still has empty intersection. Set $B_n \triangleq \mathbb{N} - \{1, ..., n\}$. Observe that

$$(\forall n, m \in \mathbb{N})\ n > m \implies B_n \subset B_m.$$

Let $h \in {}^*\mathbb{N} - \mathbb{N}$. By transfer, $^*B_h \subset {}^*B_n$ for any $n \in \mathbb{N}$. Hence,

$$^*B_h \subset \bigcap_{n\in\mathbb{N}} {}^*B_n.$$

Also by transfer $^*B_h \neq \emptyset$.

For a countable family of standard sets that has the finite intersection property we can do the first-entry decomposition to obtain a decreasing sequence of sets. By the same argument as in this example we then find that the intersection of the nonstandard sets is nonempty.

Especially in branches like topology or probability theory, we would like to have a stronger version of the above saturation property.

Definition 2.20. *Let κ be an infinite cardinal. A nonstandard extension*

$$^*: V(S) \to V(^*S), \quad A \mapsto {}^*A$$

is said to be κ-saturated if the following holds.

κ-Saturation Principle. *For any family of internal sets $(A_n)_{n\in I}$ of cardinality* $\operatorname{card} I \leqslant \kappa$ *that has the finite intersection property,*

$$\bigcap_{n\in I} A_n \neq \emptyset.$$

We wish to present an easy example that makes use of a κ-saturated nonstandard extension. First recall the Riesz' representation theorem in the following guise.

Theorem 2.21. *Let $f: X \to \mathbb{R}$ be a continuous linear functional on a Hilbert space X with inner product $(\cdot, \cdot)$. Then there exists $y \in X$ such that*

$$f(x) = (y, x)$$

for all $x \in X$.

What if we drop the condition that f is continuous and that we're working in a Hilbert space? Suppose that we have a κ-saturated nonstandard model where $\kappa \geqslant |X|$. We then have the following extension of the representation theorem, see [Landers and Rogge, 1994].

Theorem 2.22. *Let $f: X \to \mathbb{R}$ be a linear functional on a pre-Hilbert space X with inner product $(\cdot, \cdot)$. Then there exists $y \in {}^*X$ such that*

$$f(x) = {}^*(y, {}^*x)$$

for all $x \in X$.

Proof. The sets

$$C_x \triangleq \{y \in X: f(x) = (y, x)\}$$

where $x \in X$ are easily checked to have the finite intersection property by restricting to finite dimensions where Riesz' representation theorem trivially applies. Thus we find

$$y \in \bigcap_{x \in X} {}^*C_x,$$

or equivalently $y \in {}^*X$ such that

$$(\forall x \in X) \quad {}^*f(x) = {}^*(y, {}^*x).$$

Since ${}^*f(x) = f(x)$ the claim follows. □

2.7 Studying ${}^*\mathbb{Z}$ in its Own Right

The purpose of this section is at least two-fold. Firstly, to demonstrate that nonstandard objects are interesting objects in their own right. And secondly, to (hopefully) wet the readers appetite to play with p-adic numbers.

Let p be a prime. The p-adic integers $\mathbb{Z}_p$ are the inverse limit of the rings $\mathbb{Z}/p^n$. From a practical point of view they can be regarded as sequences $(a_n)_{n \in \mathbb{N}}$ such that $a_n \in \mathbb{Z}/p^n$ and $a_m \equiv a_n \pmod{p^n}$ whenever $m > n$. Of course, addition and multiplication are then given pointwisely.

Remark 2.23. As a set we can also think of $\mathbb{Z}_p$ as $(\mathbb{Z}/p)^{\mathbb{N}}$ but then addition and multiplication are no longer just the pointwise ones. At any rate, we see that $\mathbb{Z}_p$ is uncountable.

The integers $\mathbb{Z}$ canonically embed into $\mathbb{Z}_p$. A number $z \in \mathbb{Z}$ corresponding to the sequence of its remainders modulo p^n, that is

$$\mathbb{Z} \to \mathbb{Z}_p, \quad z \mapsto (z \bmod p^n)_{n \in \mathbb{N}}.$$

Example 2.24. The number 23 corresponds to $(1, 3, 7, 7, 23, 23, \ldots)$ in $\mathbb{Z}_2$.

Remark 2.25. The invertible elements in $\mathbb{Z}_p$ are exactly those sequences not starting with 0. For this reason, the field of fractions $\mathbb{Q}_p$ of $\mathbb{Z}_p$ consists of the elements that can be written as $p^{-\alpha} z$ for some $z \in \mathbb{Z}_p$. $\mathbb{Q}_p$ is a field extension of $\mathbb{Q}$ which can't be turned into an ordered field.

Remark 2.26. One importance of considering $\mathbb{Q}_p$ is rooted in number-theory. Certain equations (eg. quadratic forms) satisfy the Hasse principle, that is they have solutions over $\mathbb{Q}$ if and only if they have solutions over $\mathbb{R}$ and all the fields $\mathbb{Q}_p$ (this is called a local-global principle because a solution in $\mathbb{Q}$ is global in the sense that $\mathbb{Q}$ embeds in the $\mathbb{Q}_p$ and $\mathbb{R}$). Here as well as in other occasions the p-adic fields naturally occur alongside $\mathbb{R}$.

For every $n \in \mathbb{N}$ we have the projections $\mathbb{Z} \to \mathbb{Z}/n$ and hence also the induced projections ${}^*\mathbb{Z} \to \mathbb{Z}/n$. By the universal property of $\mathbb{Z}_p$ as the inverse limit of the $\mathbb{Z}/p^n$ we have unique morphisms $\mathbb{Z} \to \mathbb{Z}_p$ and ${}^*\mathbb{Z} \to \mathbb{Z}_p$ which commute with the canonical projections onto $\mathbb{Z}/p^n$. These morphisms are

$$\mathbb{Z} \to \mathbb{Z}_p, \quad z \mapsto (z \bmod p^n)_{n \in \mathbb{N}},$$

and

$${}^*\mathbb{Z} \to \mathbb{Z}_p, \quad z \mapsto (z \bmod p^n)_{n \in \mathbb{N}}.$$

While we saw that the embedding of $\mathbb{Z}$ into $\mathbb{Z}_p$ is injective but definitely not surjective, the canonical morphism ${}^*\mathbb{Z} \to \mathbb{Z}_p$ is no longer injective but surjective. We thus get that $\mathbb{Z}_p$ is a quotient of the hyperintegers. Using some analog of the Chinese remainder theorem, we actually have that

$$\hat{\mathbb{Z}} \triangleq \prod_p \mathbb{Z}_p$$

is a quotient as well, see [Fesenko, 2006]. The kernel are those elements of ${}^*\mathbb{Z}$ that are 0 modulo any p^α. Equivalently, they are multiples of any natural number, namely

$$\bigcap_{n \in \mathbb{Z}} n\,{}^*\mathbb{Z}$$

which is a torsion-free divisible abelian group and hence isomorphic to $\mathbb{Q}^{(\kappa)}$ (because it can be viewed as a vector space over $\mathbb{Q}$). In fact,

$${}^*\mathbb{Z} \cong \hat{\mathbb{Z}} \times \mathbb{Q}^{(\kappa)}$$

for some κ. We thus find that, in particular, the endomorphism ring of ${}^*\mathbb{Z}$ is noncommutative which is a drastic difference to the standard case where the endomorphisms of $\mathbb{Z}$ are isomorphic to $\mathbb{Z}$ again.

3 Conclusions

Given a single non-principal ultrafilter (the existence statement is called the *weak ultrafilter theorem*) we are able to construct a nonstandard extension as defined by the above axioms. To further construct κ-saturated nonstandard extensions we can use ultrafilters with special properties which do exist if we assume the axiom of choice. On the other hand, we could ask if the existence of nonstandard extensions really requires ultrafilters. Maybe there is another ultrafilter-free way to construct nonstandard extensions. Well, there is not. Recall our initial discussion of extending lim to bounded sequences. We concluded that such an extension would give rise to a non-principal ultrafilter. But given a nonstandard extension * and some infinite $h \in {}^*\mathbb{N}$ we can define for a bounded real sequence (x_n)

$$\text{x-lim}\, x_n \triangleq \text{st}({}^*x_h)$$

where $\text{st}(x)$ denotes the unique real number $x_0 \in \mathbb{R}$ such that $x \approx x_0$. It's easy to check that the so-defined x-lim is an extension of lim in the above sense. Thus we canonically get a non-principal ultrafilter from any nonstandard extension.

Yet, if we are willing to give up things like commutativity and invertibility we can do without ultrafilters. Basically, we can replace the non-principal ultrafilter on $\mathbb{N}$ in our construction by the filter of cofinite sets. A few more details can be found in [Connes, 2007] and [Tao, 2007].

Bibliography

[Belding, 2007] Belding, J. V. (2007). A weil pairing on the p-torsion of ordinary elliptic curves over the dual numbers of k. Published online at http://arxiv.org/abs/math/0703906.

[Connes, 2007] Connes, A. (2007). Noncommutative geometry: Infinitesimal variables. Published online at http://noncommutativegeometry.blogspot.com/2007/06/infinitesimal-variables.html.

[Fesenko, 2006] Fesenko, I. (2006). Several nonstandard remarks. *American Mathematical Society Translations*, 217(2).

[Hoorn, 2007] Hoorn, M. (2007). Non-nonstandard calculus, I. Published online at http://cornell-math.wordpress.com/2007/08/28/non-nonstandard-calculus-i/.

[Keisler, 2000] Keisler, H. J. (2000). *Elementary Calculus: An Approach Using Infinitesimals*.

[Kossak, 1996] Kossak, R. (1996). What are infinitesimals and why they cannot be seen. *The American Mathematical Monthly*, 103(10):846–853.

[Landers and Rogge, 1994] Landers, D. and Rogge, L. (1994). *Nichtstandard Analysis*.

[Nelson, 1987] Nelson, E. (1987). *Radically Elementary Probability Theory*. Annals of Mathematics Studies. Princeton University Press.

[O'Donovan, 2005] O'Donovan, R. (2005). An additive group containing infinities. *Seminarberichte aus dem Fachbereich Mathematik, Band 76, FernUniversität in Hagen*.

[Robinson, 1967] Robinson, A. (1967). Nonstandard arithmetic. *Bull. Amer. Math. Soc.*, 73(6):818–843.

[Salbany and Todorov, 1999] Salbany, S. and Todorov, T. (1999). Nonstandard analysis in point-set topology. *The Erwin Schrödinger International Institute for Mathematical Physics*.

[Tao, 2007] Tao, T. (2007). Ultrafilters, nonstandard analysis, and epsilon management. Published online at http://terrytao.wordpress.com/2007/06/25/ultrafilters-nonstandard-analysis-and-epsilon-management/.