Internal Set Theory and Non-Standard Analytic Methods

Jordan Brown

June 4, 2018

1 Introduction

It is widely known that the differential calculus, as developed by Newton and Leibniz, relied heavily on the concept of the infinitesimal. However, due to the inability of mathematicians to provide an adequate foundation for their manipulations of these mysterious quantities, the infinitesimal approach fell into disrepute over the following centuries and the $\epsilon - \delta$ limit as stated by Weierstrass became the foundation of modern analysis. The tools of modern logic and the axiomatic foundation of mathematics in the twentieth century, however, provided new opportunities for the creation of a rigorous calculus of infinitesimal quantities. The type-theoretic construction of enlargements of $\mathbf{R}$ by Abraham Robinson in the 1950s and 1960s was the first serious attempt to resolve this issue, and resulted in many proofs of (already established) results which were great improvements over the classical proofs. These results were, according to Kurt Gödel, "good reasons to believe that non-standard analysis, in some version or other, will be the analysis of the future." [10]

While non-standard methods are still widely disregarded by working mathematicians, these alternative formulations of the calculus are of great theoretical interest and their application may lead to original results in the future. Here, we attempt to demonstrate the effectiveness of two formulations of non-standard analysis: the type-theoretic, Robinsonian approach already mentioned, and the internal set theory as developed by Edward Nelson. The proofs of elementary results will be provided as illustrations of the power of these methods, and the omission of more complicated results does not reflect the inability of non-standard methods to adequately address such high-level problems; rather, we restrict our attention here to the foundations of these theories, and thus nearly all analytic results will fall outside of the scope of this paper.

We shall begin by performing some informal manipulations to demonstrate the simplicity with which non-standard analytic methods can handle elementary problems of the calculus. We shall then define some terms which shall be significant in both formulations of non-standard analysis (occasionally giving slightly different definitions for the two formulations, which are always substantively equivalent despite syntactical discrepancies due to the differences in the semantics of Nelson's and Robinson's formulations) before outlining the general development of enlargements and the particular properties of enlargements of $\mathbf{N}$ and $\mathbf{R}$ (denoted $^*\mathbf{N}$ and $^*\mathbf{R}$, respectively) important to Robinson's theory. Once the general concepts of non-standard analysis have been satisfactorily developed in this setting in accordance with the line of development in [10], we turn to the primary focus of this paper, the alternative theory of non-standard analysis published by Nelson in the 1970s. The predicate 'standard', as well as the axioms of transfer, idealization, and standardization will be discussed. The uniqueness of constants, the finitude of every standard set, and other elementary results will be explained, and possible objections will be addressed. We attempt to draw attention to the parallels between the theories, and the ways in which Robinson's enlargement $^*\mathbf{R}$ behaves like Nelson's $\mathbf{R}$. We shall discuss some of the results in internal set theory which appear to be contradictory, and some remarks on Powell's proof that IST is a conservative extension of the Zermelo-Fraenkel axioms with the axiom of choice will be in order.

2 Some (In)Formal Manipulations

Consider the function $f : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^2$. To compute the derivative at x_0, we would like to manipulate an infinitesimally small $\epsilon > 0$ in the following way:

$$\begin{aligned}\frac{f(x_0+\epsilon)-f(x_0)}{\epsilon} &= \frac{(x_0+\epsilon)^2 - x_0^2}{\epsilon}\\ &= \frac{2x_0\epsilon + \epsilon^2}{\epsilon}\\ &= 2x_0 + \epsilon\\ &\simeq 2x_0\end{aligned}$$

Note that we have not defined ϵ or given a rigorous exposition of its properties; the same applies to $\simeq$ at this juncture. We simply wish to draw attention to the fact that the limit computation in terms of real ϵ and δ can seemingly be replaced by infinitesimal manipulations which obey intuitive laws. All operations in the calculus can be rewritten in this way. It seems that $\lim_{x\to a} f(x) = L$ is the same as saying $f(x) \simeq L$ for all $x \simeq a$, with $\simeq$ taken to mean that two 'numbers' differ by a merely infinitesimal quantity. Similarly, computations regarding sequences and series can be reduced to considerations of terms with infinite index. While the example of differentiation provided above is not a great improvement on the $\epsilon - \delta$ derivative computation, more complicated cases make the advantages of simply substituting infinitesimal or infinite numbers quite clear. Note that these manipulations are much more closely aligned with human intuition around limiting operations (they were, after all, the basis of the original approach) than the Weierstrass construction.

3 Miscellaneous Definitions

Infinitesimal Numbers. A number $\epsilon \in \mathbf{R}$ shall be called *infinitesimal* if $-r < \epsilon < r$ for all standard $r \in \mathbf{R}^+$. While discussing Robinson's treatment of infinitesimals, we shall take a number in $^*\mathbf{R}$ to be infinitesimal if it satisfies the inequality for all $r \in \mathbf{R}^+$.

Infinitesimally Different Numbers. Two numbers $x, y \in \mathbf{R}$ (or $x, y \in {}^*\mathbf{R}$ in Robinson) are said to be *infinitesimally far apart* if $x - y$ is infinitesimal. We denote this by $x \simeq y$.

Limited Numbers. A number x shall be called *limited* if $|x| \leq r$ for some standard real number r.

Ultrafilters. Suppose S is a set, and $U \in \mathcal{P}(\mathcal{P}(S))$. If U is nonempty, we say U is an *ultrafilter* if (i) $\emptyset \notin U$ (ii) $A \in U \wedge B \in U \implies A \cap B \in U$ (iii) $A \in U \wedge B \in \mathcal{P}(S) \wedge A \subset B \implies B \in U$ (iv) $\forall A \in \mathcal{P}(S)[A \in U \vee S \backslash A \in U]$ are all satisfied. A *filter* is a nonempty subset of $\mathcal{P}(S)$ which satisfies (i), (ii), and (iii).

Ultraproducts. The notion of an ultraproduct differs slightly between Robinson and Nelson; here we shall give Robinson's definition and postpone our discussion of Nelson's ultraproduct until we have encountered his theory on a more substantial level. An ultraproduct Q_D is a structure with its set of individuals being the set of functions on I satisfying $f(v) \in M_v$ for every $v \in I$, where $\{M_v\}$ is a set of first-order structures. We say the relation $R(f_1, f_2, ..., f_n)$ holds in Q_D if the set $\{v \in I : R(f_1(v), ..., f_n(V))$ holds in $M_v\}$ is an element of the ultrafilter D.

Prenex Normal Form. A statement or relation is said to be in prenex normal form if all of the connectives are in the scope of all of the quantifiers.

Well-Formed Formulae. A well-formed formula (wff) is a formula which can be obtained by atomic formula in natural ways (see section 4 for details on what constitutes a 'natural way').

Internal Formula. An internal formula is a formula which does not include the predicate 'standard'; that is, can be formulated within ZFC.

NS Continuous Functions. We say a standard function f is NS continuous at standard x if $y \simeq x \implies f(y) \simeq f(x)$. We will sometimes denote this condition $< f, x > \in C$.

NS Uniformly Continuous Functions. A standard function f is uniformly NS continuous on a standard set E if, for all $x, y \in E$, the implication $y \simeq x \implies f(y) \simeq f(x)$ holds. We will sometimes denote this condition $< f, E > \in U$.

4 Type Theory, Enlargements, Hyperreal Numbers

We consider a first-order language L with object symbols, constants, individual variables, relation symbols of order $n \geq 1$, the three logical connectives, implication, equivalence, existential quantifiers, universal quantifiers, and brackets as atomic symbols. Atomic formulae are obtained by inserting constants or variables into the places of relative atomic symbols. If X is an atomic formula, then we say $[X]$ is a wff; if X and Y are wff, then $[\neg X], [X \vee Y], [X \wedge Y], [X \equiv Y], [(\exists y)X], [(\forall y)X]$ are wff, where y is some variable which does not appear as a quantifier in X.
We shall omit the details of Robinson's development of the lower predicate calculus through type theory and model theory, but we shall present the major results without proof so as to indicate the general line of development of the theory.

Theorem 4.1

For every sentence X, there is an equivalent sentence X_0 (that is, X and X_0 contain the same constants and relation symbols, and furthermore $X \iff X_0$) such that, in the construction of X_0 from atomic formulae, all connectives are in the scope of all quantifiers. Such an X_0 is said to be in prenex normal form.

Theorem 4.2

Suppose K is a set of sentences and every finite subset of K possesses a model. Then K possesses a model.

Theorem 4.3

If K is a stratified set of sentences which possesses a model, then its enlargement which also possesses a model.

Remark

A wff X is said to be stratified if all constants and variables which occur in X occur only in places of the same type. The enlargement of a stratified set of sentences is the union of K and a set of sentences which would take us too far into the realm of semantic type theory to describe.

This last result is not as precise here as it is presented in Robinson, but the exact nature of the enlargement is not relevant for our purposes. Corresponding to the enlargements of sets of sentences is the concept of enlargements of structures which are, in Robinson's terminology, "B-models" of K. It is at this point which we can introduce $^*\mathbf{N}$ as an enlargement of the natural numbers with strong transference conditions; that is, every true statement about $\mathbf{N}$ is also true about internal entities in $^*\mathbf{N}$ and the definitions of functions and relations remain valid. In particular, order relations are preserved, and we note that the sentence "N is smaller than all other natural numbers except for $0, 1, 2, ..., N-2, N-1$" holds for all $N \in \mathbf{N}$ and is expressible as a sentence of $^*\mathbf{N}$. Changing the language slightly when transferring to $^*\mathbf{N}$, the theorem follows:

Theorem 4.4

If $N \in {}^*\mathbf{N} \backslash \mathbf{N}$, then N is greater than all elements of $\mathbf{N}$.

We have thus shown the existence of 'infinitely large' elements in this extension of $\mathbf{N}$. Both the sets of finite natural numbers and the set of infinite natural numbers are external sets in $^*\mathbf{N}$. But of course, we cannot do analysis on the natural numbers, extension or not. We thus introduce $^*\mathbf{R}$ as an extension of $\mathbf{R}$. Without getting too involved in the machinery of the process, we present the proof of the following important theorem in non-standard analysis:

Theorem 4.5

Suppose $\{s_n\}$ is a standard infinite sequence and s is a standard real number. Then $\{s_n\} \to s$ in the Weierstrass definition if and only if $s_n \simeq s$ for all infinite n.

4.0.1 Proof.

We first suppose that $s_n \to s$ in the usual sense. That is, $(\forall \epsilon \in \mathbf{R}^+)(\exists v \in \mathbf{N})(\forall x \in \mathbf{R})[[x \in \mathbf{N} \wedge x > v] \implies |s_x - s| < \epsilon]$. We can transfer this assertion to $^*\mathbf{R}$ and, since $n > M$ for all infinite n, we conclude that $|s_n - s| < \epsilon$ for all $\epsilon \in \mathbf{R}^+$. Thus, by the definition, $s_n \simeq s$ for all infinite n.
Demonstrating that the implication in the other direction holds takes a bit more work. Suppose $s_n \simeq s$ for all infinite n. Let $\epsilon \in \mathbf{R}^+$ be arbitrary. Note that the corresponding version of the Weierstrass limit can be expressed in $^*\mathbf{R}$ and holds: $(\exists M \in {}^*\mathbf{R})(\forall x \in {}^*\mathbf{R})[[x \in {}^*\mathbf{N} \wedge x > M] \implies |s_x - s| < \epsilon]$ since we can simply take M to be infinite. Now we can transfer the statement and obtain that $s_n \to s$.

This completes our brief summary of the type-theoretic non-standard analysis. We have stressed the intuitive applications of infinitesimals and the reader can likely infer the ease with which they can be applied to more complex situations. However, we have not neglected the dependence of Robinson's theory on type theory and model theory. These are certainly drawbacks, as very little mathematics is founded upon type theory, and model theory is founded upon assumptions not used in other areas of modern mathematics. We say $^*\mathbf{R}$ is a set of hyperreal numbers (the question of uniqueness is outside the scope of this exposition: see [3] and [4]). Some work has been done to produce Robinson's hyperreals from ZFC, and the dependence on type theory is certainly resolvable. However, it seems that the dependence on model theory is unavoidable, and the production of $^*\mathbf{R}$ from standard set theory is less intuitive and perhaps less useful than the type-theoretic formulation. Some credit is due to Robinson on this point: he notes that "one might use axiomatic set theory rather than type theory for the development of higher order Non-standard Analysis" and there is not obviously a "single framework for first or higher order theories which is clearly superior to all others and one's preference may depend on the purpose that one has in mind or may be even only a matter of taste" [10]. Thus, while the use of type theory in the development of non-standard analysis is avoidable, it may still be preferable: the heavy reliance on model theory may prevent its integration into the canon of analytic methods anyways, and the constructions through Zermelo-Fraenkel set theory in the literature are all indisputably clumsier than Robinson's use of type theory. (Robinson was a student of Abraham Fraenkel – no doubt he would have used Fraenkel's axioms if he thought that they provided comparable clarity!) For a constructive description of the hyperreals and discussion of their pedagogical applications, see [4]. A more thorough discussion of the properties of types utilized in Robinson's theory can be found in [1], and a treatment of the relevant model-theoretic concepts is contained in [2].

5 Idealization, Standardization, Transfer

The many defects of non-standard analysis were evident at the time of the publication of Robinson's articles, and non-standard methods were largely ignored by the wider mathematical community. This led to an (even more unsuccessful) attempt to develop non-standard analysis as a useful tool by Edward Nelson in 1977. Nelson's theory is based on Robinson's, but its logical structure is quite different. The most significant difference is its axiomatization: rather than simply being founded on ZFC, internal set theory takes the axioms of idealization, standardization, and transfer in addition to ZFC. These axioms introduce a new, unary predicate called 'standard'. Although it is not immediately clear how the IST axioms are useful in developing non-standard analysis, some immediate results make its power evident. We will follow the natural line of development given in [6], rather than the abbreviated (and slightly more theoretical) discussion in [8]. The natural question is whether IST is a conservative extension of ZFC. If it is to be useful in proving standard results, we must show that any internal theorem in IST can be proven in ZFC. The answer is, luckily, affirmative, and Nelson gives a proof in the original paper of this somewhat surprising fact (although he credits the proof to Powell). We shall postpone addressing this point until we have developed the theory somewhat. Note the care which must be taken throughout to avoid possible illegal set formation. The axioms of ZFC only guarantee that we can form subsets through internal formulae, and if we attempt to form subsets through other formulae in IST we immediately encounter contradictions. This is perhaps the greatest difficulty involved in working with IST. Nelson awards the title of "the most insidious pitfall awaiting the mathematician who wants to use nonstandard analysis" [6] to illegal transfer, a related misinterpretation of the axioms of transfer (illegal set formation being an abuse of the axiom of standardization, as we shall see shortly), but it seems to provide a slightly less insidious challenge than illegal set formation. We leave it to the reader to decide which of these presents more challenges throughout our development of the theory.

6 Axioms of IST

Idealization: Suppose B is an internal formula with free variables x, y, and possibly others. Then

$$\forall^{st fin} z \exists x \forall y \in z B(x,y) \iff \exists x \forall^{st} y B(x,y)$$

Interpretation: If it is true that, for all standard finite sets z, there exists some x such that $B(x,y)$ holds for all $y \in z$, then it is true that there exists an x such that $B(x,y)$ holds for all standard y; furthermore, the implication also holds in the other direction: the statements are equivalent.

Standardization: Suppose C is a formula with free variables z and possibly others. Then

$$\forall^{st} x \exists^{st} y \forall^{st} z (z \in y \iff z \in x \wedge C(z))$$

Interpretation: We can use predicates to define standard subsets with standard elements. That is, for all standard sets x, there is a standard set y such that a standard element z is in y if and only if it is in x and $C(z)$ holds.

Transfer: Suppose t is a k-tuple of free variables and $A(x,t)$ is an internal formula with no undisplayed free variables. Then

$$\forall^{st} t (\forall^{st} x A(x,t) \implies \forall x A(x,t))$$

Interpretation: The transfer principle allows us to draw conclusions about all members of a set from properties of its standard members. It is important here that the axiom only states that the implication holds for standard t.

7 Immediate Results

Theorem 7.1

All usual unique constants defined through ZFC are unique and standard in internal set theory.

Proof

We want to show that all familiar constants are standard. We do this by using transfer where t is a 0-tuple: $\forall^{st} x A(x) \implies \forall x A(x)$. Taking the contrapositive, we find $\exists x(\sim A(x)) \implies \exists^{st} x(\sim A(x))$. The set of internal formulae is, of course, closed over negation, so we have shown that if there is an x which satisfies an internal formula, there is a standard x which satisfies it. If we know that there is a unique object which satisfies the formula, it follows that that object must be standard. Since all usual constants are defined using internal formulae and are unique, it follows that they are still unique and standard in internal set theory.

Theorem 7.2

There is a finite set S such that $\forall^{st} x(x \in S)$.

Proof

Let $B(S, x)$ be the formula ($x \in S \wedge S$ finite). Of course, this formula is internal, and we can apply the idealization principle $\forall^{stfin} z \exists S \forall x \in z B(S, x) \iff \exists S \forall^{st} x B(S, x)$. The second statement, $\exists S \forall^{st} x B(S, x)$, is equivalent to what we are trying to prove. The first statement, $\forall^{stfin} z \exists S \forall x \in z B(S, x)$, must hold because setting $S = z$ clearly satisfies the condition. This demonstrates our result.

Theorem 7.3

X is a standard finite set if and only if every element of X is standard.

Proof

Put $B(x, y) = (x \in X \wedge x \neq y)$. Applying the principle of idealization, $\forall^{stfin} z \exists x \forall y \in z(x \in X \wedge x \neq y) \iff \exists x \forall^{st} y(x \in X \wedge x \neq y)$. Since the implication goes both ways, we can negate both sides to obtain

$$\exists^{stfin} z \forall x \exists y \in z[\sim (x \in X \wedge x \neq y)] \iff \forall x \exists^{st} y[\sim (x \in X \wedge x \neq y)]$$

Rewriting $\sim (x \in X \wedge x \neq y)$ as $x \notin X \vee x = y$, we find the right side of our two-way implication is equivalent to $x \in X \implies$ (x standard). Rewriting the negation on the other side as well, we find that it is equivalent to $\exists^{stfin} z(X \subset z)$. So

$$\exists^{stfin} z(X \subset z) \iff (x \in X \implies (x \text{standard}))$$

From this, both directions of implication are clear: if X is standard finite, then setting $z = X$ tells us that the left side holds, so all elements of X are standard. Going the other way, if every element of X is standard, then there is a standard finite z such that $X \subset z$. So $X \in \mathcal{P}(z)$ for some standard finite z. Subsets of finite sets are obviously finite as well, and thus we know X is finite. Furthermore, by the axiom of the power set and the axiom of extensionality, the power set of X is unique (it is given by an internal formula) and therefore, by (7.1), it is standard. So X is an element of a finite standard set, and is therefore standard by the first part of this result.

Remark

Both (7.2) and (7.3) are rather counterintuitive results. They do not contradict each other or ZFC, but they appear as though they should. Note the distinction between the statements *X is a standard set* and *every element of X is standard*: by (7.3), every element of X is standard if and only if X is standard *and* finite. Additionally, there is no 'smallest' set containing all standard elements: if we were to attempt to take the intersection of all such sets, we could not declare the criteria for the set of sets we are intersecting over without engaging in illegal set formation.

Theorem 7.4

Two standard sets are equal if they have the same standard elements.

Proof

Suppose X and Y are standard sets with the same standard elements. Put $A(x) = ((x \in X \wedge x \in Y) \vee (x \notin X \wedge x \notin Y))$. This formula is clearly internal, and by applying the transfer principle (again with t being a 0-tuple) we find that

$$\forall^{st} x((x \in X \wedge x \in Y) \vee (x \notin X \wedge x \notin Y)) \implies \forall x((x \in X \wedge x \in Y) \vee (x \notin X \wedge x \notin Y))$$

The left side states that X and Y have the same standard elements, and the right that $X = Y$. This completes our proof.

Remark

This shows that, in application of the principle of standardization, the standard set y which is shown to exist is unique. We shall henceforth denote it by ${}^S\{z \in x : C(z)\}$. Note that it may have nonstandard elements, and that the nonstandard elements do not necessarily satisfy $C(z)$.

Theorem 7.5

Every limited real number is infinitesimally far away from a standard real number.

Proof

Let x be an arbitrary limited real number. Then $x \leq r$ for some standard real r. Put $E = {}^S\{t \in \mathbf{R} : t \leq x\}$. It is obvious that r is an upper bound on the standard elements of E and, by transfer, it is an upper bound for E. Furthermore, $-r$ is standard and $-r < x$, so $-r \in E$ and E is nonempty. It follows that E has a least upper bound, and it is standard. It is trivial to verify by use of the transfer principle that $x \simeq \sup E$, which is, as stated above, a standard real number.

Theorem 7.6

There are nonzero infinitesimals in $\mathbf{R}$.

Proof

Put $B(x, y) = (x \in \mathbf{R}\backslash\{0\} \wedge (y \notin \mathbf{R}^+ \vee y > x))$ in the axiom of idealization. The left side clearly holds and the right side states that x is an infinitesimal real number.

Theorem 7.7

The product of an infinitesimal and a standard real number is infinitesimal. The sum of two infinitesimals is infinitesimal. The product of two infinitesimals is infinitesimal.

Proof

Let a be an arbitrary infinitesimal real number, and c be a standard real number. We wish to show that $-r < ac < r$ for all standard $r > 0$. If $c = 0$, then the inequality trivially holds. If $c \neq 0$, note that $-\frac{r}{c} < a < \frac{r}{c}$ if $c > 0$ and $-\frac{r}{c} > a > \frac{r}{c}$ if $c < 0$. These implications hold by our assumption that a is infinitesimal and the trivial consequence of (7.1) that the additive and multiplicative inverses of standard reals, as well as the sum and product of standard reals, are standard. The chains of inequalities presented therefore hold directly from the definition of an infinitesimal number. Multiplying the inequalities through by c, we obtain the desired result $-r < ac < r$.
Now suppose a and b are infinitesimals. We wish to show that $-r < a + b < r$ for all standard positive r. By transfer,

$$\forall^{st}\alpha\forall^{st}\beta(\forall^{st}(x, y)[(x < \alpha \wedge y < \beta) \implies x + y < \alpha + \beta] \implies \forall(x, y)[(x < \alpha \wedge y < \beta) \implies x + y < \alpha + \beta])$$

Since the antecedent clearly holds, it follows that $(x < \alpha \wedge y < \beta) \implies x + y < \alpha + \beta$, where α and β are standard. By the assumption that a is infinitesimal, we know that $-\frac{r}{2} < a < \frac{r}{2}$ and $-\frac{r}{2} < b < \frac{r}{2}$. By the above result obtained through transfer, we find that it follows that $-r < a + b < r$ (strictly speaking, we have only shown that the right inequality must hold, but the argument for the left inequality is nearly identical and we shall omit it). We have therefore shown that the sum of two infinitesimals is infinitesimal. The proof that the product of infinitesimals is infinitesimal is nearly identical and left to the reader.

8 Basic Results in the Calculus

We begin by giving definitions.

Definition 8.1

A function $f \in \mathbf{R}^{\mathbf{R}}$ is NS-continuous at $x \in \mathbf{R}$ if $< f, x > \in C$ where $C = {}^S\{< f, x > \in \mathbf{R}^{\mathbf{R}} \times \mathbf{R} : \forall y[y \simeq x \implies f(y) \simeq f(x)]\}$

Definition 8.2

A function $f \in \mathbf{R}^{\mathbf{R}}$ is uniformly NS-continuous on a set $E \in \mathbf{R}$ if $< f, E > \in U$ where $U = {}^S\{< f, E > \in \mathbf{R}^{\mathbf{R}} \times \mathcal{P}(\mathbf{R} : \forall x \in E \forall y \in E[x \simeq y \implies f(x) \simeq f(y)]\}$

Remark

We shall eventually show that, if f, x, and E are standard, then the notions of NS-continuity and uniform NS-continuity directly correspond to continuity and uniform continuity. However, the sets given to us through the axiom of standardization do not necessarily have all of their elements satisfying the defining predicate. Thus, while C and U are unique and standard, they are infinite and by (7.3) contain elements which do not satisfy their respective predicates. We can therefore say very little about the nonstandard elements of U and C.

Lemma

If f and x are standard, then $f(x)$ is standard.

Proof

Put $A(y, f, x) = [y \neq f(x)]$. Evidently, A is an internal formula. For f and x standard, we can apply the transfer axiom and take the contrapositive. That is, $\exists y[y = f(x)] \implies \exists^{st} y[y = f(x)]$. Since $f(x)$ is unique, it follows by the argument given for (7.1) that $f(x)$ is standard.

Theorem 8.3

Suppose $E \subset \mathbf{R}$ is closed and bounded. If $< f, x > \in C$ for all $x \in E$, then $< f, E > \in U$.

Proof

We will show that this holds when f and E are standard, and then proceed to demonstrate that we can do so without loss of generality. Let $x, y \in E$ such that $x \simeq y$. Assume for contradiction that $\mathrm{st}(x) \notin E$. Since E is closed, $\mathbf{R} \backslash E$ is open. Therefore, there is a standard positive real number r such that ${}^S\{z : |\mathrm{st}(x) - z| < r\} \subset \mathbf{R} \backslash E$ (note that this is only true because we can apply the transfer axiom). Since E is standard by assumption, so is $\mathbf{R} \backslash E$, and it satisfies the predicate construction of the standard topology on $\mathbf{R}$ through our construction of our topology vis-à-vis the axiom of standardization. But for all such r, we know $|\mathrm{st}(x) - x| < r$, so $\mathrm{st}(x) \in \mathbf{R} \backslash E$, which is a contradiction. Thus, $\mathrm{st}(x) \in E$. Of course, $x \simeq \mathrm{st}(x)$, and it follows that $\mathrm{st}(x) \simeq y$. Therefore, by continuity, $f(\mathrm{st}(x)) \simeq f(y)$. (Note that we could not apply this argument directly to show $f(x) \simeq f(y)$ by continuity, since we only know that continuous f satisfy the predicate when f and the point at which f is being considered are both standard.) Also by the continuity of f at $\mathrm{st}(x)$, we see that $f(x) \simeq f(\mathrm{st}(x))$. Therefore, $f(x) \simeq f(y)$. This demonstrates uniform continuity for f and E standard. By definition, C and U are standard sets, and the generalization to cases where f or E may be nonstandard follows by an application of the transfer principle with the statement A being that the implication of uniform NS continuity holds.

Remark

As part of our proof that $\mathrm{st}(x) \in E$, we considered the set ${}^S\{z : |\mathrm{st}(x) - z| < r\}$. We could not form the set $\{z : |\mathrm{st}(x) - z| < r\}$ since the formula is external and this is illegal set formation. Although it is tempting (and correct to some degree) to think of 'standard' objects as corresponding to conventional mathematical objects and internal formulae as conventional mathematical formulas, they do not correspond to each other: we can form nonstandard sets using internal formulae, and we can use the principle of standardization to form standard sets through external predicates. The formula $x \leq n$ is internal regardless of whether n is standard, but if we wish to apply transfer to determine whether all members of a set satisfy $x \leq n$ based on whether the standard elements satisfy the condition, then we are out of luck if n is nonstandard.

Theorem 8.4 (Extreme Value Theorem)

Suppose $f : E \to \mathbf{R}$ is NS-continuous, where E is closed and bounded. Then f achieves its maximum.

Proof

By the transfer principle, if we can show that this is true for all standard f and E, it must hold for all f and E. Thus, we take f and E to be standard without a loss of generality. Let F be a finite subset of E containing all standard points in E. Let $x \in F$ be a point at which the restricted function $f|_F$ attains its maximum. Then st $x \in F$, since st $x \in E$ and st x is standard by construction. By continuity, $f(x) \simeq f(\text{st}(x))$. Thus, for all standard $y \in E$, we have $f(\text{st}(x)) \geq f(y) + \alpha$ for some infinitesimal α (which may depend on y). By assumption, $f(\text{st}(x))$ and $f(y)$ are standard, so $f(\text{st}(x)) \geq f(y)$. (This is not difficult to see: if $f(y) > f(\text{st}(x))$, then $f(y) - f(\text{st}(x))$ would be a standard number greater than zero and there could not be any infinitesimal that bridges the difference.) Since $\forall^{st} y \in E[f(y) \leq f(\text{st}(x))]$, we can apply the transfer principle to see that $\forall y \in E[f(y) \leq f(\text{st}(x))]$. Thus, f achieves its maximum at st x.

Theorem 8.5 (Intermediate Value Theorem)

Suppose $f : [a, b] \to \mathbf{R}$ is NS-continuous, $f(a) < 0$, and $f(b) > 0$. Then f attains the value zero in the interval.

Proof

As before, we take f, a, b standard without a loss of generality, and set $F \subset [a, b]$ to be a finite set containing all standard elements in the interval. Let $x \in F$ be the smallest number in F for which $f(x) \geq 0$. It is easy to see that st $x \in F$. Furthermore, we know that $f(\text{st}(x)) \geq 0$, since $f(\text{st}(x))$ is standard and by continuity it is infinitely close to $f(x)$, which is greater than or equal to zero. Let y be the element in F preceding x. Assume for contradiction that it is not true that $y \simeq x$. Then it is not true that $\text{st}(y) \simeq \text{st}(x)$, so there is a standard $r > 0$ such that $\text{st}(y) + r < \text{st}(x)$. But $\text{st}(y) + r/2$ is standard and is thus contained in F. Additionally, it is greater than y and less than x. This contradicts our assumption that y is the element of F directly preceding x. Therefore, $y \simeq x$. But $f(y) < 0$ by construction and, by continuity, $f(y) \simeq f(\text{st}(x))$, since $y \simeq \text{st}(x)$. It follows that $f(\text{st}(x)) \leq 0$, since $f(\text{st}(x))$ is standard. We have shown that $0 \leq f(\text{st}(x)) \leq 0$, so $f(\text{st}(x)) = 0$. This completes our proof.

Remark

It is not necessary here to investigate the infinitesimal differential calculus to any extended degree. However, we will briefly discuss the relevant definitions and provide proofs of results regarding derivatives which are easily obtained using Nelson's theory. In particular, we aim to demonstrate the ease with which one can establish the chain rule within internal set theory, a crucial theoretical result in the calculus which requires a good deal of manipulation to prove within the classical theory. Note that, when we define the derivative, we do not hold one of the points we are using to compute the slope at the point where we are computing the derivative. We require that the derivative is infinitesimally far away from the slope between any two distinct points which are each infinitesimally far away from the point at which we are computing the derivative. The theoretical superiority of this more restrive definition of the derivative is argued in [9], and its properties mesh nicely with the techniques of internal set theory.

Definition

Suppose $I \subset \mathbf{R}$ is an interval. We say a function $f \in \mathbf{R}^I$ is differentiable at $x \in I$ if $< f, x > \in \mathfrak{D}$, where

$$\mathfrak{D} =^S \{< \text{f}, \text{x} > \in \mathbf{R}^I \times I : \exists^{st} y[(x_1 \simeq x \wedge x_2 \simeq x \wedge x_1 \neq x_2 \wedge x_1 \in I \wedge x_2 \in I) \implies \frac{f(x_2) - f(x_1)}{x_2 - x_1} \simeq y]\}$$

Definition

If f and x are standard, $< f, x > \in \mathfrak{D}$, then we say y is a derivative of f at x and denote y by $f'(x)$.

Theorem 8.6

If $< f, x > \in \mathfrak{D}$, x lies in the interior of I, and y and z are derivatives of f at x, then $y = z$; that is, derivatives are unique.

Proof

By the transfer principle, we need only to demonstrate the result for standard f and x, as the set $\mathfrak{D}$ is standard and the statement of uniqueness is an internal formula. Thus, we take f and x to be standard without loss of generality. Since I is an interval which contains at least two points other than x, we can take $x_1 \in I$ and $x_2 \in I$ such that $x_1 \simeq x \wedge x_2 \simeq x \wedge x_1 \neq x_2$. Thus, $\frac{f(x_2)-f(x_1)}{x_2-x_1} \simeq y$ must hold for x_1 and x_2. Therefore, $\frac{f(x_2)-f(x_1)}{x_2-x_1} \simeq z$ and, by transitivity, $y \simeq z$. Since y and z are standard, it follows that $y = z$.

Lemma

Suppose I is an interval. Then

$$\mathfrak{D} =^S \{<\mathrm{f,x}>\in \mathbf{R}^I \times I : \exists^{st} f'(x)[(x_1 \simeq x \wedge h \simeq 0 \wedge h \neq 0 \wedge x_1 \in I \wedge x_1 + h \in I) \implies (\frac{1}{h}(f(x_1+h) - f(x_1)) \simeq f'(x))]\}$$

Furthermore, $y = f'(x)$ if $<\mathrm{f,x}>$ is in either set.

Proof

Since $\mathfrak{D}$ and this new set, which we shall call $\mathfrak{D}'$ for the moment, are both standard, we need only to show that their standard values agree to ascertain that they are the same set by the transfer principle. Since the standard values of $<\mathrm{f,x}>$ satisfy the predicates, we will simply show

$$\exists^{st} y[(x_1 \simeq x \wedge x_2 \simeq x \wedge x_1 \neq x_2 \wedge x_1 \in I \wedge x_2 \in I) \implies \frac{f(x_2) - f(x_1)}{x_2 - x_1} \simeq y] \iff$$

$$\exists^{st} f'(x)[(x_1 \simeq x \wedge h \simeq 0 \wedge h \neq 0 \wedge x_1 + h \in I) \implies (\frac{1}{h}(f(x_1+h) - f(x_1)) \simeq f'(x))]$$

But of course, both directions are trivial: simply put $x_2 = x_1 + h$, or equivalently $h = x_2 - x_1$ in the other direction, and $y = f'(x)$. Thus, the sets of differentiable functions are the same, and the derivatives are the same. We shall henceforth use the two definitions of the derivative interchangeably, depending on which formulation is more convenient for our purposes.

Theorem 8.7 (Derivative as Linear Operator)

Differentiation is a linear operator. That is, if $< f, x >, < g, x > \in \mathfrak{D}$, then for all $\alpha, \beta \in \mathbf{R}$, $< \alpha f + \beta g > \in \mathfrak{D}$ and $(\alpha f + \beta g)' = \alpha f' + \beta g'$.

Proof

The proof of this result is not difficult. It is only given the status of a theorem because of its tremendous significance. As usual when demonstrating results about standard sets, we take our values f, g, x to be standard and rely on the transfer principle for generalization to all elements. We let $x_1 \simeq x$ and $h \simeq 0$ and apply the formulation from the previous lemma to write the following equalities for some infinitesimal a and b:

$$f(x_1 + h) = f(x_1) + f'(x)h + ah$$

$$g(x_1 + h) = g(x_1) + g'(x)h + bh$$

It follows immediately that

$$\alpha f(x_1 + h) = \alpha f(x_1) + \alpha f'(x)h + (\alpha a)h$$
$$\beta g(x_1 + h) = \beta g(x_1) + \beta g'(x)h + (\beta b)h$$

By (7.7), the product of two infinitesimals is infinitesimal, an thus we have shown that $(\alpha f)' = \alpha f'$ and $(\beta g)' = \beta g'$. Combining the equations yields

$$\alpha f(x_1 + h) + \beta g(x_1 + h) = \alpha f(x_1) + \beta g(x_1) + \alpha f'(x)h + \beta g'(x)h + (\alpha a)h + (\beta b)h$$

Rewriting this equality yields

$$(\alpha f + \beta g)(x_1 + h) = (\alpha f + \beta g)(x_1) + (\alpha f'(x) + \beta g'(x))h + (\alpha a + \beta b)h$$

The sum of two infinitesimals is also infinitesimal by (7.7). This completes our proof.

Theorem 8.8

The usual rules for differentiation hold in internal set theory: the product rule, quotient rule, and power rule are all valid relationships between elements of $\mathfrak{D}$.

Proof

The proofs are trivial and left to the reader.

Theorem 8.9 (Single-Variable Chain Rule)

Suppose $< g, x >\in \mathfrak{D}$ and $< f, g(x) >\in \mathfrak{D}$. Then $< f \circ g, x >\in \mathfrak{D} >$ and $(f \circ g)'(x) = f'(g(x))g'(x)$.

Proof

Let $x_1 \simeq x$ and $h \simeq 0$ be given, satisfying the conditions of our alternate formulation of the derivative. Then there are infinitesimals α and β such that

$$f(g(x_1) + h) = f(g(x_1) + g'(x)h + \beta h)$$

The above statement only used the differentiability of g at x. Now we use the fact that $g'(x)h + \beta h \simeq 0$ to write

$$f(g(x_1) + h) = f(g(x_1)) + f'(g(x))[g'(x)h + \beta h] + \alpha[g'(x)h + \beta h]$$

Simply rearranging the terms yields

$$f(g(x_1) + h) = f(g(x_1)) + hf'(g(x_1))g'(x) + h[\alpha g'(x) + \alpha\beta + f'(g(x_1))\beta]$$

This completes our proof.

9 The Reduction Algorithm

The reduction algorithm is an important procedure in IST which allows us to reduce statements to internal statements which are equivalent for all standard values of the free variables. While it is not particularly complex, it is theoretically critical to understanding the relationship between IST and ZFC. We follow [6] in dividing the algorithm into six steps:

1. **Reduction to Definitions.** Replace all external predicates with their definitions, so that the only remaining external predicate is 'standard'.

2. **Expression Through Quantifiers.** Rewrite the formula such that the only external predicates are the quantifiers $\forall^{st}$ and $\exists^{st}$, e.g. replacing $[x\text{standard}]$ with $\exists^{st}y[y = x]$ and $(x\text{standard}) \implies A(x)$ with $\forall^{st}xA(x)$.

3. **Prenex Normal Form.** Using the usual rules for logical manipulation and their analogues for the external quantifiers, rewrite the formula such that all quantifiers occur before all connectives. (This procedure is detailed in [6] and identical to the one outlined in [10].)

4. **Rank Reduction.** We say a relation in prenex normal form is of rank j if there are j internal quantifiers which include external quantifiers in their scope. If $j > 0$, we obtain an equivalent formula of rank $j - 1$ in the following way. Without loss of generality, assume that the rightmost internal quantifier with an external quantifier in its scope is $\forall$. Call this quantifier Q_i. (The fact that this is without loss of generality comes from the fact that $\exists xA(x) \equiv \neg\forall x[\neg A(x)]$ and dealing with the negations is trivial.) We can rearrange the external quantifiers in the scope of Q_i such that the universal external quantifiers occur to the left of the existential external quantifiers by the equivalence $\exists^{st}\forall^{st}yA(x, y) \equiv \forall^{st}y'\exists^{st}xA(x, y'(x))$. Nelson proves this equivalence by applying a result (Theorem 1.3 in his article) which we have omitted here due to its intuitive nature and the tedious question of its proof. Once we have moved all of the universal external quantifiers to the right of Q_i to the left of all of the existential external quantifiers to the right of Q_i, we can simply pull all of these external universal quantifiers to the other side of Q_i, since $\forall x\forall^{st}yA(x, y) \equiv \forall^{st}y\forall xA(x, y)$. We now simply have existential external quantifiers in the scope of Q_i. We can take this string of existential external quantifiers to be an ordered tuple with only one existential external quantifier remaining, stating the existence of such a tuple; i.e. $\exists^{st}x\exists^{st}yB(x, y) \equiv \exists^{st}(x, y)B(x, y)$. Taking the contrapositive of the axiom of idealization, we find that $\forall x\exists^{st}yB(x, y) \implies \exists^{st fin}z\forall x\exists y \in zB(x, y)$ for internal B. To the right of our existential external quantifier for the tuple, there are only internal statements, so we can apply this implication to rewrite the statement with the existential quantifier to the left of Q_i. There are now only internal quantifiers in the scope of Q_i and the rank of the new, equivalent formula is $j - 1$.

5. **Induction.** Repeat the previous process until the formula has rank 0.

6. **Transfer.** Replace all external quantifiers with the corresponding internal quantifiers. By transfer, this is equivalent to the original formula for all standard values of the free variables.

This completes our description of the reduction algorithm. We shall now apply it to obtain the promised proof that NS-continuity and uniform NS-continuity are equivalent to continuity and uniform continuity for standard values of the free variables. We know that, for standard values of f and x, the condition $< f, x >\in C$ is equivalent to the statement $\forall y(y \simeq x \implies f(y) \simeq f(x))$. We reduce the 'infinitesimally close' condition to its definition to obtain the statement $\forall y([\forall^{st}\delta|y - x| < \delta] \implies [\forall^{st}\epsilon|f(y) - f(x)| < \epsilon])$, where δ and ϵ are permitted to range over the standard positive reals. Rewriting in prenex normal form and reducing the rank, we ultimately obtain that our statement is equivalent to $\forall\epsilon\exists^{fin}\delta'\forall y(\forall\delta \in \delta'|y-x| < \delta \implies |f(y)-f(x)| < \epsilon)$ for all standard values of f and x. Since δ' is finite, we can put $\delta = \min \delta'$ and our formula further reduces to $\forall\epsilon\exists\delta\forall y(|y-x| < \delta \implies |f(y)-f(x)| < \epsilon)$, as the condition that $|y-x| < \delta$ for all $\delta \in \delta'$ is simply equivalent to saying that $|y-x|$ is less than the minimum value of elements in δ', as δ' is finite. We have thus reduced our description of NS-continuity to the condition of continuity for standard f and x. A similar procedure demonstrates that uniform NS-continuity and uniform continuity agree for standard f and E.

10 Consistency and Applications

We do not have space here to describe all of the possible applications of the internal set theory. Nelson discusses more advanced analytic results in a nonstandard fashion in [6], and his development of probability theory through internal set theory in [7] also provides a good idea of the advantages of internal set theory. Differentiation and integration are easily handled, and it does not take much work to transfer usual topological concepts. Here, we shall not discuss the applications of the theory any further, opting instead to prove that the theory can be applied. That is, we want to show that it is legitimate to use IST to prove results over ZFC. The most succinct statement of this result is

Theorem 10.1 (Conservativity of Extension)

Every internal theorem provable in internal set theory is provable in ZFC.

We shall not prove this result in its entirety, but rather prove several intermediate results to give an outline of the proof and to demonstrate the lines of argumentation which are used to obtain the intermediate steps omitted here. We note here that this means that, if ZFC is consistent, then so is IST. If IST is inconsistent, then by the principle of explosion we can prove any internal result, including statements which directly contradict the axioms of ZFC. By the conservativity of the extension to IST (which we shall shortly prove), it follows that the axioms of ZFC can be used to demonstrate that the axioms of ZFC are false. Thus, if IST is inconsistent, then so is ZFC; we take no additional risk by appending the axioms of IST to the usual ones.
To move towards a proof of (10.1), we shall develop the theory of ultrafilters and ultraproducts, primarily in the single-variable setting. We shall omit the proof of the restricted idealization principle for adequate ultrapowers and results regarding ultralimits, as well as the final steps using transfinite induction and techniques from model theory. The complete proof, as presented in [6], demonstrates that it is possible to rewrite the proof of any internal theorem in IST as a proof in ZFC, explicitly demonstrating that IST is a conservative extension of ZFC.

Definition 10.2

Let V be a set and $\mathfrak{U}$ an ultrafilter on an index set I. We write two functions $f, g \in V^I$ are equivalent modulo $\mathfrak{U}$ if the set of elements of I for which f and g agree is an element of $\mathfrak{U}$.

Lemma

Let $\mathfrak{F}$ be a filter on a set S. Then $S \in \mathfrak{F}$ and $\forall I \in \mathcal{P}(S)[I \in \mathfrak{F} \implies S\backslash I \notin \mathfrak{F}]$

Proof

Filters must be nonempty by definition, so there is some $U \in \mathcal{P}(S)$ such that $U \in \mathfrak{F}$. Naturally, $U \subset S$ and, since filters are closed over the taking of supersets, $S \in \mathfrak{F}$. Furthermore, filters are closed over intersection and if I and $S\backslash I$ were both in $\mathfrak{F}$ for some $I \in \mathcal{P}(S)$, their intersection would also be in $\mathfrak{F}$. But their intersection is the empty set, which cannnot be contained in $\mathfrak{F}$ by definition.

Remark

The lemma implies that there is no filter on a set which properly contains an ultrafilter. That is, ultrafilters are maximal filters. If $\mathfrak{U}$ is an ultrafilter on I, and $S \subset \mathcal{P}(I)$ such that $\mathfrak{U} \subset S$ and $\mathfrak{U} \neq S$, then there is some element $G \in \mathcal{P}(I)$ such that $G \in S$ and $G \notin \mathfrak{U}$. By our characterization of ultrafilters, this implies that $G^c \in \mathfrak{U}$ and thus $G^c \in S$. So $G \in S$ and $G^c \in S$, and we must either conclude that S is not closed over intersection or S contains the empty set, both of which disqualify it from being a filter. It is easy to see that the converse holds as well – if a filter is maximal, it is an ultrafilter – and we use this characterization of ultrafilters when we prove that every filter can be enlarged to an ultrafilter.

Theorem 10.3

Let V be a set and $\mathfrak{U}$ an ultrafilter on an index set I. If $f, g \in V^I$ are equivalent modulo $\mathfrak{U}$ and $E \subset V$, then $\{i \in I : f(i) \in E\} \in \mathfrak{U}$ if and only if $\{i \in I : g(i) \in E\} \in \mathfrak{U}$.

Proof

By symmetry, we need only prove one direction. Suppose $\{i \in I : f(i) \in E\} \in \mathfrak{U}$. Let $\mathfrak{F}$ be the set of elements in I on which f and g agree. By assumption, $\mathfrak{F} \in \mathfrak{U}$. Since ultrafilters are closed over intersection, $\{i \in I : f(i) \in E\} \cap \mathfrak{F} \in \mathfrak{U}$. It is obvious that $\{i \in I : f(i) \in E\} \cap \mathfrak{F} = \{i \in I : g(i) \in E\} \cap \mathfrak{F}$. Thus, $\{i \in I : g(i) \in E\} \cap \mathfrak{F} \in \mathfrak{U}$. Since ultrafilters are closed over the taking of supersets, it follows that $\{i \in I : g(i) \in E\} \in \mathfrak{U}$. This completes our proof.

Definition 10.4

Let V be a set, $\mathfrak{U}$ an ultrafilter on an index set I. We define the ultrapower $^*V = V^I/\mathfrak{U}$ as the set of all equivalence classes modulo $\mathfrak{U}$ of functions in V^I. Note that this makes sense by (10.3).

Definition 10.5

Let $V, I, \mathfrak{U}$ be as above. We write $^*E$ is the set of all equivalence classes of functions over modulo $\mathfrak{U}$ such that $\{i \in I : f(i) \in E\} \in \mathfrak{U}$ and we say $^*E$ is the extension of E.

Remark

The terminology may seem strange here. Note that there is a natural map from E to $^*E$: none of the constant functions $f(i) = a$ for a fixed $a \in E$ and all $i \in I$ are equivalent to each other, and thus represent distinct equivalence classes. If we consider such a constant function to correspond to the element $a \in E$, then we see that it makes sense to call $^*E$ the extension of E. There is an injective function mapping elements of E to (some subset of) $^*E$.

Theorem 10.6

Suppose $^*V$ is an ultrapower of V. Then the map from $\mathcal{P}(V)$ to $\mathcal{P}(^*V)$ which takes E to $^*E$ is an isomorphism of the Boolean algebra $\mathcal{P}(V)$ onto a Boolean subalgebra of $\mathcal{P}(\ ^*V)$.

Proof

We first show that $^*\emptyset = \emptyset$. For all f, we see that $\{i \in I : f(i) \in \emptyset\} = \emptyset$. Of course, by the definition of the ultrafilter, $\emptyset \notin \mathfrak{U}$. Thus, no equivalence classes are elements of $^*\emptyset$, and $^*\emptyset = \emptyset$. By a similar argument using the fact that $I \in \mathfrak{U}$, we find that $^*V = V$. Now we will show that $^*(E \cap F) = {^*E} \cap {^*F}$ for all $E, F \in \mathcal{P}(V)$. Because ultrafilters are closed over the taking of supersets, we see that $\{i \in I : f(i) \in E \cap F\} \in \mathfrak{U} \implies \{i \in I : f(i) \in E\} \in \mathfrak{U}$. Thus, $^*E \subset {^*(E \cap F)}$, and by the transitivity of the subset relation $^*E \cap {^*F} \subset {^*}\ (E \cap F)$. Additionally, if $\{i \in I : f(i) \in E\} \in \mathfrak{U}$ and $\{i \in I : f(i) \in F\} \in \mathfrak{U}$, then we can use the closure of ultrafilters over intersection to conclude that $\{i \in I : f(i) \in E \cap F\} \in \mathfrak{U}$. This demonstrates that $^*(E \cap F) \subset {^*E} \cap {^*F}$. Now that we have established the subset relation in both directions, we can see that the sets are equal. The only remaining fact which we need to demonstrate in order to establish that we have an isomorphism is that $^*(E^c) = (^*E)^c$, where ^{c} is taken to indicate complementation in V or $^*V$, depending on context. It is a special property of ultrafilters, not common to all filters, that allows us to state $\{i \in I : f(i) \in E^c\} \in \mathfrak{U} \iff \{i \in I : f(i) \in E\} \notin \mathfrak{U}$. Thus, the extension of E^c is precisely the complement of the extension of E. That is, if an equivalence class is not an element of $^*E$, then it is an element of $^*(E^c)$, and we have thus verified all conditions for our map to be an isomorphism between Boolean algebras.

Definition 10.7

Let $f, g \in (V^n)^I$ and $f_1, ..., f_n, g_1, ..., g_n \in V^I$ such that $f(i) = (f_1(i), ..., f_n(i)), g(i) = (g_1(i), ..., g_n(i))$. We say f and g are equivalent modulo $\mathfrak{U}$ if f_j is equivalent to g_j modulo $\mathfrak{U}$ for each $j = 1, 2, ..., n$.

Definition 10.8

The projection operator $\pi_j \in \mathcal{P}(V^{n-1})^{\mathcal{P}(V^n)}$ for $j = 1, 2, ..., n$ is given by

$$\pi_j(E) = \{(x_1, ..., x_{j-1}, x_{j+1}, ..., x_n) \in V^{n-1} : \exists x_j \in V[(x_1, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n) \in E]\}$$

That is, $\pi_j(E)$ contains all elements of V^{n-1} which agree with an element of E when the jth position of the n-tuple in E is discounted.

Theorem 10.9

If $^*V$ is an ultrapower of V and $E \subset V^n$, then $^*\pi_j(E) = \pi_j(^*E)$.

Proof

By definition, $^*\pi_j(E)$ is the set of equivalence classes modulo $\mathfrak{U}$ of functions from I to V^{n-1} for which the pre-image of $\pi_j(E)$ over the function is an element of $\mathfrak{U}$. Let f be an equivalence class of functions from I to V^{n-1}, with components $(f_1, f_2, ...f_{j-1}, f_{j+1}, ..., f_n)$. The pre-image of $\pi_j(E)$ over f is simply

$$\{i \in I : \exists x_i \in V[(f_1(i), f_2(i), ..., f_{j-1}(i), x_i, f_{j+1}(i), ..., f_n(i)) \in E]\}$$

Similarly, the set $\pi_j(^*E)$ is the set of equivalence classes of functions g such that there is a function $i \mapsto x_i$ where

$$\{i \in I : (g_1(i), ..., g_{j_1}(i), x_i, g_{j+1}(i), ..., g_n(i)) \in E\} \in \mathfrak{U}$$

Evidently, this is the same set we described for $^*\pi_j(E)$. Thus, $^*\pi_j(E) = \pi_j(^*E)$.

Definition 10.10

Let V be a set, $E \subset V^n$, $x \in V$. We write

$$E_{x,j} = \{(x_1, ..., x_{j-1}, x_{j+1}, ..., x^n) \in V^{n-1} : (x_1, ..., x_{j-1}, x, x_{j+1}, ..., x_n) \in E\}$$

That is, where $\pi_j(E)$ is the set of $(n-1)$tuples which agree with an element of E except in the jth position, $E_{x,j}$ is the set of $(n-1)$tuples which agree with an element of E except in the jth position, where x is the missing component. In the latter case it is specified that an element of E must be obtainable via the insertion of a particular $x \in V$, whereas in the former only the existence of such an element is necessary for membership in the set $\pi_j(E)$.

Theorem 10.11

Suppose $^*V$ is an ultrapower of V and take the natural injection $V \to ^*V$. If $E \subset V^n$ and $x \in V$, then $^*(E_{x,j}) = (^*E)_{x,j}$.

Proof

This proof has the same structure as the proof of (10.9). We see that $^*(E_{x,j})$ is the set of equivalence classes of functions from I to V^{n-1} such that the pre-image of $E_{x,j}$ over those functions is an element of the ultrafilter $\mathfrak{U}$. That is, an equivalence class f is an element of $^*(E_{x,j})$ if and only if the set of $i \in I$ such that $f(i)$ is an $(n-1)$tuple to which x can be added in the jth position to obtain an element of E is an element of $\mathfrak{U}$. Additionally, $(^*E)_{x,j}$ is the set of equivalence classes g from I to V^{n-1} such that inserting the function $i \mapsto x$ in the jth position – yielding the function $(g_1(i), ..., g_{j-1}(i), x, g_{j+1}(i), ..., g_n(i))$ from I to V^n – is an element of $^*E$: the pre-image of E over this function from I to V^n is an element of $\mathfrak{U}$. Clearly, the conditions for membership in these two sets are equivalent. This is simply the prosaic explanation of our proof of (10.9) in the case where x is fixed.

Remark

We can treat n-ary relations on V as elements of $\mathcal{P}(V^n)$. Of course, 'and' corresponds to intersection, 'or' to union, and 'not' to complementation. Existential quantifiers can be expressed through the projection operators, and universal quantifiers can be formed through combinations of projection operators and complementation. Thus, (10.6) and (10.9) imply that any statement about V formed from a finite set of finitary relations through first-order logic can be expressed as a statement about $^*V$ involving extensions of the given relations. By (10.11), this is also true for formulas involving free variables. This is what Nelson calls 'the transfer principle for ultrapowers'.

Theorem 10.12

Let $\xi \in ^*V$. The set of subsets E of V such that $\xi \in ^*E$ is an ultrafilter on V.

Proof

Call this set of subsets $t(\xi)$. Since $^*\emptyset = \emptyset$, as we have shown before, $\emptyset \notin t(\xi)$. Furthermore, $V \in t(\xi)$, since $\xi \in {}^*V$ as V is the codomain of ξ and the index set $I \in \mathfrak{U}$, where $^*V = V^I/\mathfrak{U}$. Suppose $E, F \in t(\xi)$. Since $\xi \in {}^*E$ and $\xi \in {}^*F$, it follows that $\xi \in {}^*E \cap {}^*F$. We have already shown that $^*(E \cap F) \subset {}^*E \cap {}^*F$, so $\xi \in {}^*(E \cap F)$ and thus $E \cap F \in t(\xi)$. So $t(\xi)$ is closed over intersection. Suppose now that $E \subset G$. Then $^*E \subset {}^*G$ by the fact that $\{i \in I : f(i) \in E\} \subset \{i \in I : f(i) \in G\}, \{i \in I : f(i) \in E\} \in \mathfrak{U}$, and ultrafilters are closed over the taking of supersets. Therefore, since $\xi \in {}^*E$, it follows that $\xi \in {}^*G$. So $t(\xi)$ is closed over the taking of supersets. Now we will show that the complement of E with respect to V is not in $t(\xi)$. Indeed, it is easy to see that $\{i \in I : f(i) \in E\} = \{i \in I : f(i) \in E^c\}^c$, so the properties of complementation we require from ultrafilters is inherited from the ultrafilter $\mathfrak{U}$. We have therefore shown that $t(\xi)$ is an ultrafilter on V.

Definition 10.13

Let V be a set, $\tilde{V}$ the set of ultrafilters on V. Suppose $^*V$ is an ultrapower of V. For $\xi \in {}^*V$, let $t(\xi)$ be the set of all subsets $E \subset V$ with $\xi \in {}^*E$. We say $^*V$ is an adequate ultrapower of V if the map t from $^*V$ to $\tilde{V}$ is surjective.

10.1 Remark

We know that t does not map to anything outside of $\tilde{V}$ by (10.12), so we know that $\tilde{V}$ is indeed a codomain of t. It is not true, however that every ultrapower $^*V$ generates a t which is onto $\tilde{V}$. The existence of such an ultrapower over each set is the surprising result of (10.14).

Lemma

If $\mathfrak{F}$ is a filter on a set I, then there is an ultrafilter $\mathfrak{U}$ on I such that $\mathfrak{F} \subset U$.

Proof

Suppose $\mathfrak{F}$ is not an ultrafilter. Then there must exist some $G \subset I$ such that neither G nor G^c is contained in $\mathfrak{F}$. Let $\mathfrak{G}$ be the set of all supersets of sets of the form $F \cap G$ for $F \in \mathfrak{F}$. Clearly, $\mathfrak{G}$ is a filter which is a proper superset of $\mathfrak{F}$. We can repeat the process to obtain a strictly increasing chain of filters, and their union must be a filter. The set $\mathcal{P}(I)$ is partially ordered by inclusion and thus so are all subsets of $\mathcal{P}(I)$. The sequence of filters obtained by this enlargement process is partially ordered, and of course every linearly ordered subset has an upper bound of $\mathcal{P}(I)$. Thus, by Zorn's lemma, the sequence has a maximal element. Whenever the filter we have is not an ultrafilter, an enlargement of that filter is also in that sequennce and thus no filter which is not an ultrafilter could be a maximal element. It follows that there is an ultrafilter in the sequence, and it is clearly a superset of $\mathfrak{F}$.

Theorem 10.14

There exists an adequate ultrapower of every set.

Proof

Let V be an arbitrary set. We take $\mathcal{P}_{fin}$ to denote the set of finite subsets of a given set. Put $I = \mathcal{P}_{fin}(\mathcal{P}(V))$. For each $E \in \mathcal{P}(V)$, let $\tilde{E} = \{i \in I : E \in i\}$. (Note that I here is not the index set of a function, but rather the finite power set of the power set of V.) Consider the intersection $\bigcap\limits_{j=1}^{n} \tilde{E}_j$, where $E_j \in \mathcal{P}(V)$. Naturally, the set $\{E_1, ..., E_n\}$ lies in this intersection. Since the intersection is always nonempty, there must be a filter on I which contains all sets $\tilde{E}$ for any $E \in \mathcal{P}(V)$. In particular, the set of subsets of I which contains all supersets and finite intersections of sets of the form $\tilde{E}$ is a filter, as it is nonempty, contains I, is closed over the taking of supersets, closed over intersection, and does not contain the empty set since the empty set is not a superset of any set except for itself and the intersection of any two sets of the form $\tilde{E}$ is nonempty by the above considerations. We can always enlarge filters to be ultrafilters, so we can assert the existence of a $\mathfrak{U}$ which is an ultrafilter on I and contains all sets of the form $\tilde{E}$. Now put $^*V = V^I/\mathfrak{U}$ and $\mathfrak{V}$ an ultrafilter on V. We claim that $^*V$ is an adequate ultrapower.

For each $i \in I$, we put $A_i = \bigcap\limits_{j \in i \cap \mathfrak{V}} j$. We assert that each A_i is nonempty. This follows from the fact that this is a finite intersection (all elements of I are finite sets of subsets of V) and the ultrafilter $\mathfrak{V}$ is closed over intersection, so by induction $A_i \in \mathfrak{V}$. Ultrafilters do not contain the empty set, and thus $A_i \neq \emptyset$. Let $f \in \mathfrak{V}^I$ be such that $f(i) \in A_i$ for each $i \in I$. Suppose ξ is the equivalence class modulo $\mathfrak{U}$ of f. We claim that $t(\xi) = \mathfrak{V}$, where $t(\xi)$ is the set of subsets $F \in \mathcal{P}(\mathcal{P}(V))$ such that $\xi \in {}^*F$. Let $E \in \mathfrak{V}$ be arbitrary. If $E \in i$, then $A_i \subset E$ by the construction of A_i; by assumption, E is one of the sets we are taking the intersection over. Thus, $f(i) \in E$ for all $i \in I$. Recall that $\tilde{E} = \{i \in I : E \in i\}$, and since

$E \in i \in I \implies f(i) \in E$, it follows that $\tilde{E} \subset \{i \in I : f(i) \in E\}$. Furthermore, since $\tilde{E} \in \mathfrak{U}$ by the construction of $\mathfrak{U}$, it follows that its superset $\{i \in I : f(i) \in E\} \in \mathfrak{U}$. Therefore, $\xi \in {}^*E$. By construction of t, it trivially follows that $E \in t(\xi)$. Of course, $E \in \mathfrak{V}$ is arbitrary, and thus $\mathfrak{V} \subset t(\xi)$. Both $\mathfrak{V}$ and $t(\xi)$ are ultrafilters, the former by construction and the latter by (10.12), so this implies that they are equal (see the lemma following Definition 10.2). Furthermore, as $\mathfrak{V}$ is an arbitrary ultrafiilter on V, it follows that t is surjective and thus ${}^*V = V^I/\mathfrak{U}$ is an adequate ultrapower of $\mathfrak{V}$.

At this point, we elect to skip over some of the finer points of the remainder of the proof and simply state the results which are important for the complete proof of (10.1). The general line of development is as follows. We establish an idealization principle for adequate ultrapowers analogous to the axiom of idealization. Analyzing the sequence of sets where ${}^0V$ is some set, ${}^{n+1}V$ an ultrapower of ${}^nV$ for $n = 0, 1, 2, ...$ we consider the sequence of natural injective mappings $j_n : {}^nV \to^{n+1} V$ which we discussed in the case of extensions above. We say an ultralimit is simply the direct limit of this sequence, and an adequate ultralimit is an ultralimit where, for each n, ${}^{n+1}V$ is an adequate ultrapower of ${}^nV$. By a pretty piece of transfinite induction and the properties of adequate ultralimits, we arrive at a point from which the model theorists can declare the proof complete. By producing a model of the axioms which are used to prove any result in IST, we can rewrite the proof without using the axioms of idealization, standardization, or transfer; that is, a proof within ZFC. So it is legitimate to use IST to obtain standard results.

It is admittedly quite difficult to use IST to prove anything, and it is difficult to conceive of a situation to which internal set theory is better suited than Robinson's nonstandard analysis. There is no reason to believe that IST ever provides more clarity than classical methods. Despite Nelson's protestations, the dangers of illegal set formation and illegal transfer, combined with the thoroughly counterintuitive structure of the theory, make it highly unlikely that it will ever be widely utilized. Though we have proved that it is theoretically legitimate to use IST to prove standard results, anyone interested in proving standard results would be quite better off with Weierstrass's analysis. It is not impossible, however, that it is possible to produce a nonstandard theory of analysis which truly encapsulates the simplicity of Leibniz's infinitesimals with the rigor of the twentieth-century logicians. The ultrafilter/ultrapower foundations used by Robinson and Nelson, reliant on reduction algorithms and properties of prenex normal forms, do not seem to provide compelling alternatives for working mathematicians.

11 References

1. Peter B. Andrews, *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*, Academic Press, Inc., Orlando, 1986.

2. Wilfrid Hodges, *Model Theory*, Cambridge University Press, New York, 1993.

3. Vladimir Kanovei, Saharon Shelah, A Definable Nonstandard Model of the Reals, *The Journal of Symbolic Logic*, **69**, 159-164 (2004).

4. H. Jerome Keisler, *Elementary Calculus: An Infinitesimal Approach*, 3rd ed., Dover, Boston, 2012.

5. Gregory Lawler, Comments on Edward Nelson's "Internal Set Theory: A New Approach to Nonstandard Analysis", *Bulletin of the American Mathematical Society*, **48**, 503-506 (2011).

6. Edward Nelson, Internal Set Theory: A New Approach to Nonstandard Analysis, *Bulletin of the American Mathematical Society*, **83**, 1165-1198 (1977).

7. ——, *Radically Elementary Probability Theory*, Princeton University Press, Princeton, 1987.

8. ——, The Syntax of Nonstandard Analysis, *Annals of Pure and Applied Logic*, **38**, 123-134 (1985).

9. Albert Nijenhuis, Strong Derivatives and Inverse Mappings, *The American Mathematical Monthly*, **81**, 969-980 (1974).

10. Abraham Robinson, *Non-standard Analysis*, 2nd ed., Princeton University Press, Princeton, 1996.