AN INVITATION TO NONSTANDARD ANALYSIS

TOM LINDSTRØM

INTRODUCTION

Nonstandard Analysis - or the Theory of Infinitesimals as some prefer to call it - is now a little more than 25 years old (see Robinson (1961)). In its early days it was often presented as a surprising solution to the old and - it had seemed - impossible problem of providing infinitesimal methods in analysis with a logical foundation. It soon became clear, however, that the theory was much more than just a reformulation of the Calculus, when Bernstein and Robinson (1966) gave the first indication of its powers as a research tool by proving that all polynomially compact operators on Hilbert spaces have nontrivial invariant subspaces. Since then nonstandard techniques have been used to obtain new results in such diverse fields as Banach spaces, differential equations, probability theory, algebraic number theory, economics, and mathematical physics just to mention a few. Despite the wide variety of topics involved, these applications have enough themes in common that it is natural to regard them as examples of the same general method.

This paper is intended as an exposition of these recurrent themes and the theory uniting them. I have called it "An invititation to nonstandard analysis" because it is meant as an invitation - a friendly welcome requiring no other background than a smattering of general mathematical culture. My point of view is that of applied nonstandard analysis; I'm interested in the theory as a tool for studying and creating standard mathematical structures. As such, I feel that it is of greater interest to the analyst than to the logician, and this attitude is, I hope, reflected in the presentation; put paradoxically, I have tried to make the subject look the way it would had it been developed by analysts or topologists and not logicians. This is the explanation for certain unusual features such as my insistence on working with ultrapower models and my willingness to downplay the importance of first order languages.

Although the presentation may be a little unconventional, the choice of topics is not; there seems to be a fairly general agreement on what are the most important and powerful nonstandard techniques, and I have seen it as my main task to give a full and detailed account of these. The idea has been to bring the reader to the point where he can study more specialized nonstandard papers with only an occasional consultation of the literature, and where he can begin to think of applying nonstandard methods in his own field of interest. Unfortunately, this emphasis on methodology and basic techniques has made it impossible to include convincing examples of new results and at the same time keep the paper within reasonable bounds. But as the other contributions to the present volume contain applications which in depth and variety far exceed anything I could conceivably have put into an introduction of this kind, I do not think that these omissions are of much consequence.

The paper consists of four chapters, each divided into three sections. The first three chapters contain a systematic exposition of nonstandard techniques in different branches of analysis, while the fourth focuses on the underlying logical principles. Not all readers will want or need to read everything; those who are eager to get on to applications may wish to skip Chapter IV at the first reading and only concentrate on the most relevant parts of the other chapters. The chart in Figure 1 traces the dependences between the various sections in detail. Note in particular the sections in the dotted boxes; they are not really part of the systematic development of the theory, but contain examples and applications which add flesh and blood to the bare theoretical bones of the other sections. The paper ends with a comprehensive set of Notes with suggestions for further study.

Acknowledgements. It is a pleasure to thank Nigel Cutland for inviting me to give the lectures on which this paper is based; feedback from many members of the audience both improved the overall quality of the presentation and eliminated some rather embarassing mistakes. Special thanks are due to Keith Stroyan who left me with a heavily annotated copy of the first draft; many of his suggestions have been incorporated into the final version, while some of the more ambitious ones have been left out only for lack of time and space. Through the years a number of people have influenced my view of nonstandard analysis, but none more than Sergio

Cambridge University Press
978-0-521-35947-4 - Nonstandard Analysis and its Applications
Edited by Nigel Cutland
Excerpt
More information

Albeverio, Jens Erik Fenstad, and Raphael Høegh-Krohn, with whom I spent five years writing a book on the subject. I don't think I want to know how many of my best ideas are really theirs. Finally, I would like to thank the Nansen Fund for generous travel support.

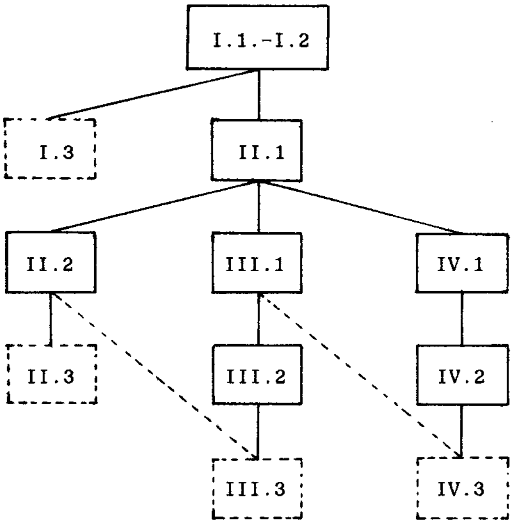

Figure 1

I. A SET OF HYPERREALS

Although nonstandard methods have been used in most parts of mathematics, I will start where it all began historically - with the construction of a number system $^*\mathbb{R}$ extending $\mathbb{R}$ and containing infinitely large and infinitely small elements.

I.1 CONSTRUCTION OF $^*\mathbb{R}$

To convince you that this construction is quite natural and not the least mysterious, let me compare it to something you are all familiar with - the construction of the reals from the rationals using Cauchy-sequences. Recall how this is done: If $\mathcal{C}$ is the set of all rational Cauchy-sequences, and $\equiv$ is the equivalence relation on $\mathcal{C}$ defined by

$$\{a_n\} \equiv \{b_n\} \quad \text{iff} \quad \lim_{n\to\infty} (a_n - b_n) = 0, \tag{1}$$

then the reals are just the set $\mathbb{R} = \mathcal{C}/\equiv$ of all equivalence classes. To define algebraic operations on $\mathbb{R}$, let $\langle a_n\rangle$ denote the equivalence class of the sequence $\{a_n\}$, and define addition and multiplication componentwise

$$\langle a_n\rangle + \langle b_n\rangle = \langle a_n + b_n\rangle; \quad \langle a_n\rangle \cdot \langle b_n\rangle = \langle a_n \cdot b_n\rangle. \tag{2}$$

The order on $\mathbb{R}$ is defined simply by letting $\langle a_n\rangle < \langle b_n\rangle$ if there is an $\varepsilon \in \mathbb{Q}_+$ such that $a_n < b_n - \varepsilon$ for all sufficiently large n. Finally, we can identify the rationals with a subset of $\mathbb{R}$ through the embedding

$$a \to \langle a,a,a,\ldots\rangle. \tag{3}$$

The construction of $^*\mathbb{R}$ follows exactly the same strategy. Beginning with the set $\mathcal{A}$ of all sequences of real numbers, I shall introduce an equivalence relation $\sim$ on $\mathcal{A}$ and define $^*\mathbb{R}$ as the set $\mathcal{A}/\sim$ of all equivalence classes. If as above $\langle a_n\rangle$ denotes the equivalence class of the sequence $\{a_n\}$, the algebraic operations are defined componentwise as in (2), and I shall also introduce an order on *ℝ which turns it into an ordered field. Finally, ℝ will be identified with a subset of *ℝ through the embedding $a \to \langle a,a,a,\ldots\rangle$.

Before I define the equivalence relation ~, it may be wise to say a few words about the philosophy behind the construction. When we create the reals from the rationals, we are interested in constructing limit points for all "naturally" convergent sequences. Since the limit is all we care about, it is convenient to identify as *many* sequences as possible; i.e. all those which converge to the same "point". No attention is paid to the rate of convergence; hence the two sequences $\{\frac{1}{n}\}$ and $\{\frac{1}{\sqrt{n}}\}$ are identified with the same number 0 although they converge at quite different rates. In creating *ℝ from ℝ, we want to construct a rich and well-organised algebraic structure which encodes not only the *limit* of a sequence but also its *mode of convergence*. To achieve this, we shall reverse the strategy above and identify as *few* sequences as possible.

This sounds silly; to "identify as few sequences as possible" must surely mean the trivial identification $\{a_n\} \sim \{b_n\}$ iff $\{a_n\} = \{b_n\}$. Well, it doesn't if you also want *ℝ to have all the nice algebraic properties of ℝ.

I.1.1 Example

Let $\{a_n\} = \{1,0,1,0,1,\ldots\}$ and $\{b_n\} = \{0,1,0,1,0,\ldots\}$; then $\{a_n\}\cdot\{b_n\} = 0$, although $\{a_n\}$ and $\{b_n\}$ are both non-zero. Thus if we use the trivial identification, we get a structure with zero divisors.

The idea is to make the equivalence relation ~ just strong enough to avoid the problem of zero divisors. Before I can give the definition, I have to fix a finitely additive measure on ℕ with the following properties.

I.1.2 Definition

Throughout this chapter m denotes a (fixed) finitely additive measure on the set ℕ of positive integers such that:

(i) For all $A \subset ℕ$, m(A) is defined and is either 0 or 1.
(ii) $m(ℕ) = 1$, and $m(A) = 0$ for all finite A.

That m is a finitely additive measure means, of course, that $m(A \cup B) = m(A) + m(B)$ for all disjoint sets A and B. Note that m divides the subsets of $\mathbb{N}$ into two classes, the "big" ones with measure one and the "small" ones with measure zero, in such a way that all finite sets are "small". The existence of such measures is an exercise in Zorn's lemma (see the Appendix, Proposition A.1).

Observe that for any $A \subset \mathbb{N}$, either $m(A) = 1$ or $m(A^c) = 1$ but not both. Moreover, if $m(A) = 1$ and $m(B) = 1$, then $m(A \cap B) = 1$ since

$$m((A \cap B)^c) = m(A^c \cup B^c) \leq m(A^c) + m(B^c) = 0 + 0 = 0.$$

I.1.3 Definition

Let $\sim$ be the equivalence relation on the set $\mathcal{S}$ of all sequences of real numbers defined by

$$\{a_n\} \sim \{b_n\} \text{ iff } m\{n: a_n = b_n\} = 1,$$

i.e. if $\{a_n\}$ equals $\{b_n\}$ almost everywhere.

Having defined the equivalence relation $\sim$, I can now do as promised and let $^*\mathbb{R} = \mathcal{S}/\sim$ be my set of *nonstandard reals* or *hyperreals*. If $\langle a_n\rangle$ denotes the equivalence class of the sequence $\{a_n\}$, define addition and multiplication in $^*\mathbb{R}$ by

$$\langle a_n\rangle + \langle b_n\rangle = \langle a_n + b_n\rangle \quad ; \quad \langle a\rangle \cdot \langle b_n\rangle = \langle a_n \cdot b_n\rangle \tag{4}$$

and order it by

$$\langle a_n\rangle < \langle b_n\rangle \text{ iff } m\{m: a_n < b_n\} = 1. \tag{5}$$

I really ought to check that these definitions are independent of the representatives $\{a_n\},\{b_n\}$ of the equivalence classes $\langle a_n\rangle, \langle b_n\rangle$, but I shall gladly leave all book-keeping of this sort to you.

To see that the problem of zero divisors has disappeared, assume that $\langle a_n\rangle \cdot \langle b_n\rangle = \langle 0,0\ldots\rangle$, i.e. $m\{n: a_n \cdot b_n = 0\} = 1$. Since $\{n: a_n \cdot b_n = 0\} = \{n: a_n = 0\} \cup \{n: b_n = 0\}$, either $\{n: a_n = 0\}$ or $\{n: b_n = 0\}$ has measure one, and thus either $\langle a_n\rangle = \langle 0,0,\ldots\rangle$ or $\langle b_n\rangle = \langle 0,0,\ldots\rangle$. Note that the conditions on m are exactly right for this argument to work.

But *ℝ is much more than an algebraic structure without zero divisors; it is an ordered field with zero element $0 = \langle 0,0\ldots\rangle$ and unit $1 = \langle 1,1,\ldots\rangle$. As proving this in detail would just be boring, I'll restrict mystelf to the following typical example.

I.1.4 Example

If $a,b,c \in {}^*\mathbb{R}$ are such that $a > 0$ and $b < c$, how do we prove that $ab < ac$? Well, if $a = \langle a_n\rangle$, $b = \langle b_n\rangle$, and $c = \langle c_n\rangle$, then there are sets $A, B \subset \mathbb{N}$ of m-measure one such that $a_n > 0$ if $n \in A$ and $b_n < c_n$ if $n \in B$. Thus $a_n b_n < a_n c_n$ for all $n \in A \cap B$, and since $m(A\cap B) = 1$, this proves that $ab < ac$.

As already indicated

$$a \to \langle a,a,a,\ldots\rangle \tag{6}$$

is an injective, order preserving homomorphism embedding ℝ in *ℝ, and I shall identify ℝ with its image under this map. Thus all real numbers are elements of *ℝ, but what do its other members look like? In particular, where do the infinitesimal and infinite numbers come from? Let us first agree on the terminology.

I.1.5 Definition

(a) An element $x \in {}^*\mathbb{R}$ is *infinitesimal* if $-a < x < a$ for all positive real numbers a.

(b) An element $x \in {}^*\mathbb{R}$ is *finite* if $-a < x < a$ for some positive real number a. An element in *ℝ which is not finite is called *infinite*.

Three examples of infinitesimals are 0, $\delta_1 = \langle \frac{1}{n}\rangle$, and $\delta_2 = \langle \frac{1}{\sqrt{n}}\rangle$. To check that, say, δ_1 is infinitesimal, note that for any positive $a \in \mathbb{R}$, the set $\{n: -a < \frac{1}{n} < a\}$ contains all but a finite number of n's and hence has measure one. Observe also that since $\delta_1 \neq \delta_2$, the two sequences $\{\frac{1}{n}\}$ and $\{\frac{1}{\sqrt{n}}\}$ converging to zero at different rates are represented by different infinitesimals. Finally note that zero is the only infinitesimal real number. Examples of infinite numbers, one positive and one negative, are $\langle n\rangle$ and $\langle -n^2\rangle$.

It is easy to check that the arithmetic rules one would expect really hold; e.g. the sum of two infinitesimals is infinitesimal, and so is the product of a finite number and an infinitesimal one. More interesting is the following observation which shows that the finite part of $^*\mathbb{R}$ has a very simple structure.

I.1.6 Proposition

Any finite $x \in {}^*\mathbb{R}$ *can be written uniquely as a sum* $x = a + \varepsilon$, *where* $a \in \mathbb{R}$ *and* ε *is infinitesimal.*

Proof. The uniqueness is obvious since if $x = a_1 + \varepsilon_1 = a_2 + \varepsilon_2$, then $a_1 - a_2 = \varepsilon_2 - \varepsilon_1$; but this quantity is both real and infinitesimal, so it must be zero.

For the existence, let $a = \sup\{b\in\mathbb{R}: b < x\}$; since x is finite, a exists. I must show that x-a is infinitesimal. Assume not, then there is a real number r such that $0 < r < |x-a|$ (absolute values in $^*\mathbb{R}$ are defined exactly as absolute values in $\mathbb{R}$). If $x-a > 0$, this implies that $a+r < x$, contradicting the choice of a. If $x-a < 0$, I get $x < a-r$, also contradicting the choice of a.◂

Let us write $x \approx y$ to mean x and y are *infinitely close*; i.e. x-y is infinitesimal.

I.1.7 Definition

For each finite $x \in {}^*\mathbb{R}$, the unique real number a such that $x \approx a$ is called the *standard part* of x and is denoted by $^\circ x$ or st(x). Conversely, for each $a \in \mathbb{R}$, the set of all $x \in {}^*\mathbb{R}$ such that $a = {}^\circ x$ is called the *monad* of a.

The next lemma shows that there is a reasonable relationship between the asymptotic behaviour of $\{a_n\}$ and the value of $\langle a_n \rangle$.

I.1.8 Lemma

If the sequence $\{a_n\}$ *has limit* a, *then* $a \approx \langle a_n \rangle$.

Proof. All we have to show is that $a-\varepsilon < \langle a_n \rangle < a+\varepsilon$ for any given $\varepsilon \in \mathbb{R}_+$. But since $\{a_n\}$ converges to a, the set $\{n: a-\varepsilon < a_n < a+\varepsilon\}$ contains all but a finite number of n's and hence has measure one.◂

Let me briefly summarise the contents of this section. We have constructed a set *ℝ of nonstandard reals or *hyperreals* which is an ordered field extension of ℝ and contains infinitely small and infinitely large numbers. A simple but useful picture to have in mind is the one shown in Figure 2; it depicts *ℝ as an ordered structure consisting

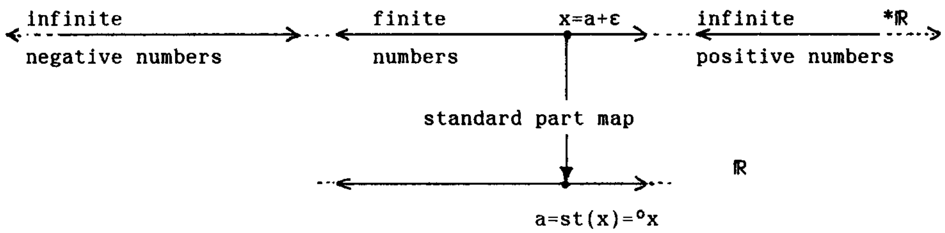

Figure 2

of three parts; the infinite negative numbers, the finite numbers, and the infinite positive numbers. According to Proposition I.1.6, the finite part looks exactly like ℝ except that each point in ℝ has been blown up to become a copy of the set of infinitesimals.

Although the constructions of ℝ and *ℝ are so very similar, there is an important difference between the two sets; the dependence on the measure m makes *ℝ "less canonical" than ℝ. Indeed, if you look back at Example I.1.1, you will see that in *ℝ one of the two sequences {0,1,0,1,0,...}, {1,0,1,0,1,...} is identified with 0 and the other one with 1; and which is which depends on the measure m. If we stick to our philosophy above and consider ℝ and *ℝ as structures constructed to reflect the asymptotic behaviour of sequences, this is not too disconcerting; the difference between the two sets is just that in creating ℝ from the rational Cauchy-sequences we throw out the sequences that do not have a decent asymptotic behaviour at the very beginning, while in creating $^*\mathbb{R}$ we keep them and treat them in an arbitrary but coherent way instead. Mathematically, this point of view is supported by the fact that hyperreals arising from different measures m have the same interesting analytic properties (although they can only be shown to be isomorphic under extra set-theoretic assumptions such as the continuum hypothesis). In Chapter III, I will show that there is occasionally a need for richer sets of hyperreals constructed not from the set of all sequences $\mathbb{R}^{\mathbb{N}}$ but from a larger set $\mathbb{R}^{A}$, where A is uncountable, and I will continue the present discussion then.

I.2 INTERNAL SETS AND FUNCTIONS

One of the first things you do when you have introduced a new mathematical structure is to look for the classes of "nice" subsets and functions (such as open sets and continuous functions in topology, measurable sets and functions in measure theory). In nonstandard analysis the "nice" sets and functions are called internal, and they arise in the following way.

I.2.1 Definition

(a) A sequence $\{A_n\}$ of subsets of $\mathbb{R}$ defines a subset $\langle A_n\rangle$ of $^*\mathbb{R}$ by

$$\langle x_n\rangle \in \langle A_n\rangle \quad \text{iff} \quad m\{n: x_n \in A_n\} = 1,$$

and a subset of $^*\mathbb{R}$ which can be obtained in this way is called *internal*.

(b) A sequence $\{f_n\}$ of functions $f_n: \mathbb{R} \to \mathbb{R}$ defines a function $\langle f_n\rangle: {}^*\mathbb{R} \to {}^*\mathbb{R}$ by

$$\langle f_n\rangle(\langle x_n\rangle) = \langle f_n(x_n)\rangle,$$

and any function on $^*\mathbb{R}$ which can be obtained in this way is called *internal*.

I.2.2 Example

(a) If $a = \langle a_n\rangle$ and $b = \langle b_n\rangle$ are two elements of $^*\mathbb{R}$, then the interval $[a,b] = \{x\in{}^*\mathbb{R}: a \le x \le b\}$ is internal as it is obtained as $\langle [a_n,b_n]\rangle$.

(b) If $c = \langle c_n\rangle$ is in $^*\mathbb{R}$, the function $\sin(cx)$ is an internal function defined by $\sin(cx) = \langle \sin(c_n x_n)\rangle$.