NONSTANDARD ANALYSIS

DR. J. PONSTEIN

J. Ponstein

former professor at the University of Groningen

A Naive Way to the Infinitesimals
(an unorthodox treatment of Nonstandard Analysis)

Contents

Prologue

Jaap Ponstein has been a professor of Operations Research at the University of Groningen for more than 20 years. The beginning of his career in Groningen coincided with the beginning of the study of econometrics in the university, in which OR was positioned. His broad experience as a mathematician, both pure and applied, appeared to be instrumental in the development of OR in Groningen. He became an authority on optimization, duality, convexity and generalized differentiability. His book *"Approaches to the theory of optimization"* (Cambridge University Press, 1980) is a beautiful example of the precise and transparent way in which he succeeded in connecting various areas of optimization. He was the right person to act as promotor honoris causa on the occasion of the honorary degree of R.T. Rockafellar (University of Washington, Seattle, USA) on June 20, 1984.

Professor Ponstein was known as an excellent teacher. For students and co-workers he was the ideal advisor: inspiring, and always making time to discuss papers. Usually, his comments were critical, both on contents and on formulations, but always aimed at improvement and maximal clarity. Hans Nieuwenhuis, Caspar Schweigman, Gerard Sierksma and I, to name a few, learned a great deal from him. We still are impressed by his integrity, his commitment, and his scientific erudition.

In the eighties Imme van den Berg raised Jaap's interest in so-called *Nonstandard Analysis*. The possibility of dealing with infinitely small and infinitely large quantities as numbers, without losing mathematical rigor appealed to him. After his retirement (ultimo 1989) he made a profound study of this subject. The point of view of his research characterizes Jaap's scientific attitude: How to introduce infinitesimals and other "nonstandard" numbers naively and simplemindedly, but in a way such that the resulting theory is mathematically sound, and complete within obvious limits.

This book is the result of his study. It was finished in the summer of 1995. Unfortunately, he was not able to publish it: he died on November 22, 1995.

12

I am pleased and proud, that the Faculty of Economics at the University of Groningen has offered the opportunity to publish the book. It underlines the respect we feel for Jaap's merits.

Groningen, February 2001   Wim Klein Haneveld

After Jaap had presented his inaugural speech my mother said to me: "I had no idea what he was talking about, but he presented it in such a way that I couldn't help thinking: Any minute now it will all make sense". This was one of the special things about Jaap: his ability to inspire people with topics that appealed to him. It was the same when he explained to our small daughters, Marianne, Anne, Els and Ada, what a half and a quarter is. He cut an apple-pie first into two and then into four pieces. That same spirit of enthusiasm showed itself later on in Groningen when teaching econometrics students and when helping our neighbours' children with their "difficult maths".

To my mind the above illustrates the way in which Jaap began working on this book. Infinitely small numbers and their applications fascinated him. Jaap understood how to address the subject in an unconventional way, in a way that would make the subject more accessible to others. He had extensive discussions with Els, which word, and why, would be the best to describe the infinitely small numbers. As the manuscript neared completion, he discussed with Anne the possibility of publishing it.

That was the situation when Jaap became sick in September 1995. My daughters and I did not act on the manuscript for quite some time. We were unsure about what to do. Finally, I contacted Imme van den Berg who agreed to review the manuscript and bring it to the attention of the Faculty of Economics at the University of Groningen. The Faculty, represented by Wim Klein Haneveld, offered to publish the manuscript as a volume of the internal Research Reports series of the SOM Research School.

I would like to express my gratitude to Imme van den Berg for his rapid and thorough review of the manuscript, and to Wim Klein Haneveld for his careful handling the publication process. Last but not least, I would like to thank Suwarni Bambang Oetomo, who invested an enormous amount of effort into the task of transferring the manuscript from the word-processing program CW into the more modern TeX. It would not have been possible to publish this document without the financial assistance offered by the Faculty of Economics. My sincere thanks, also on behalf of our daughters, goes to all who have helped to bring about the publication of Jaap's manuscript.

Zeist, November 2000 Sileen Ponstein-Troelstra

14

16
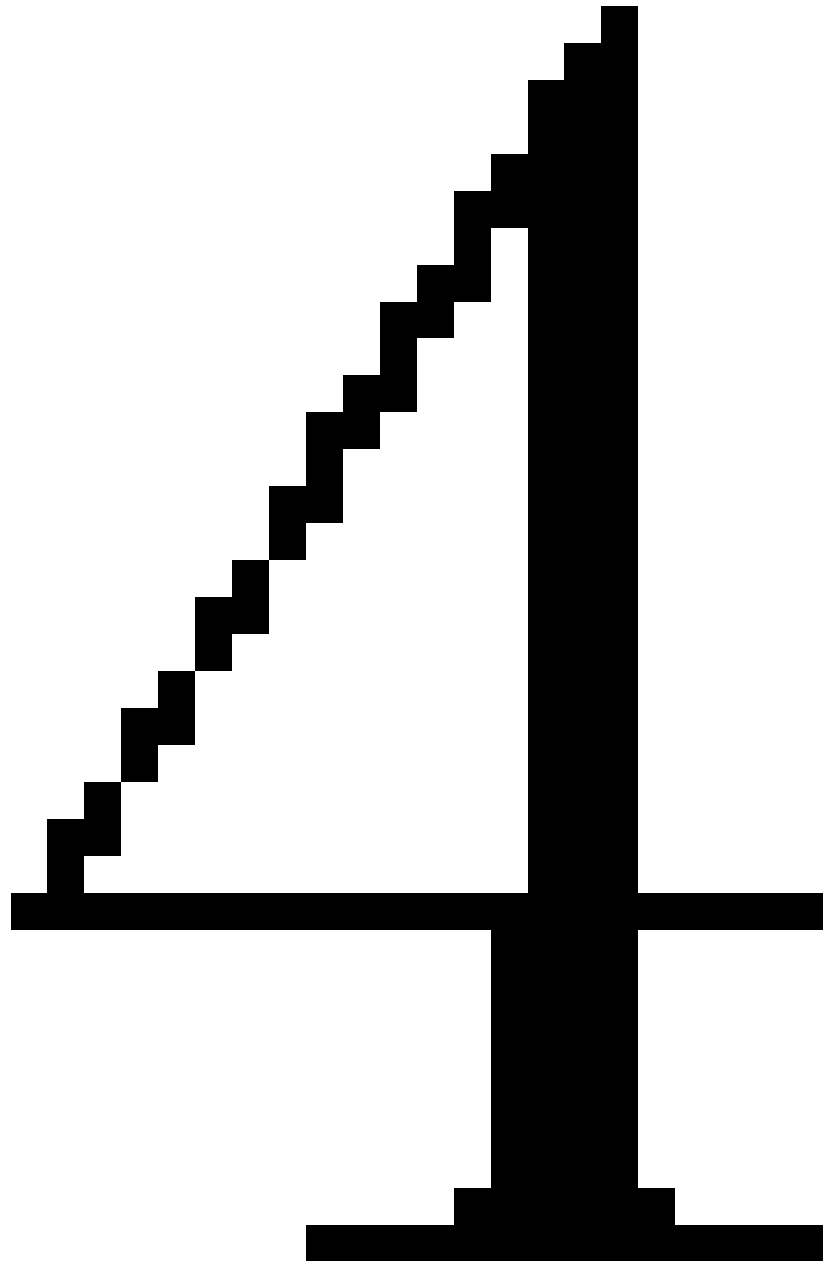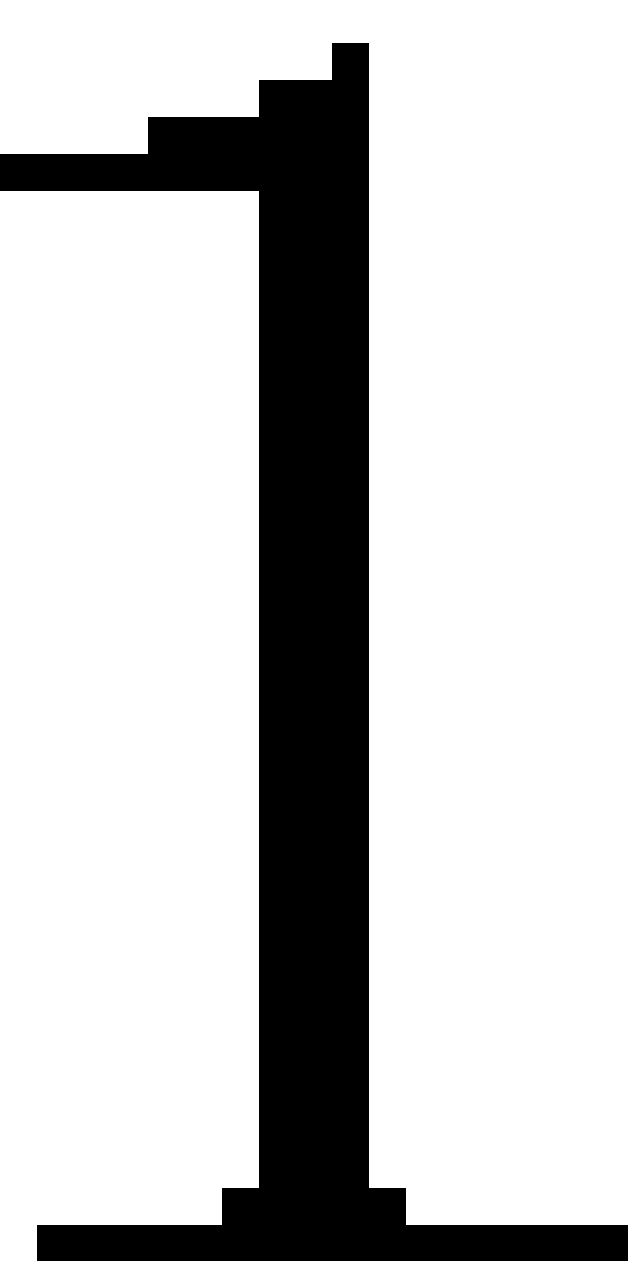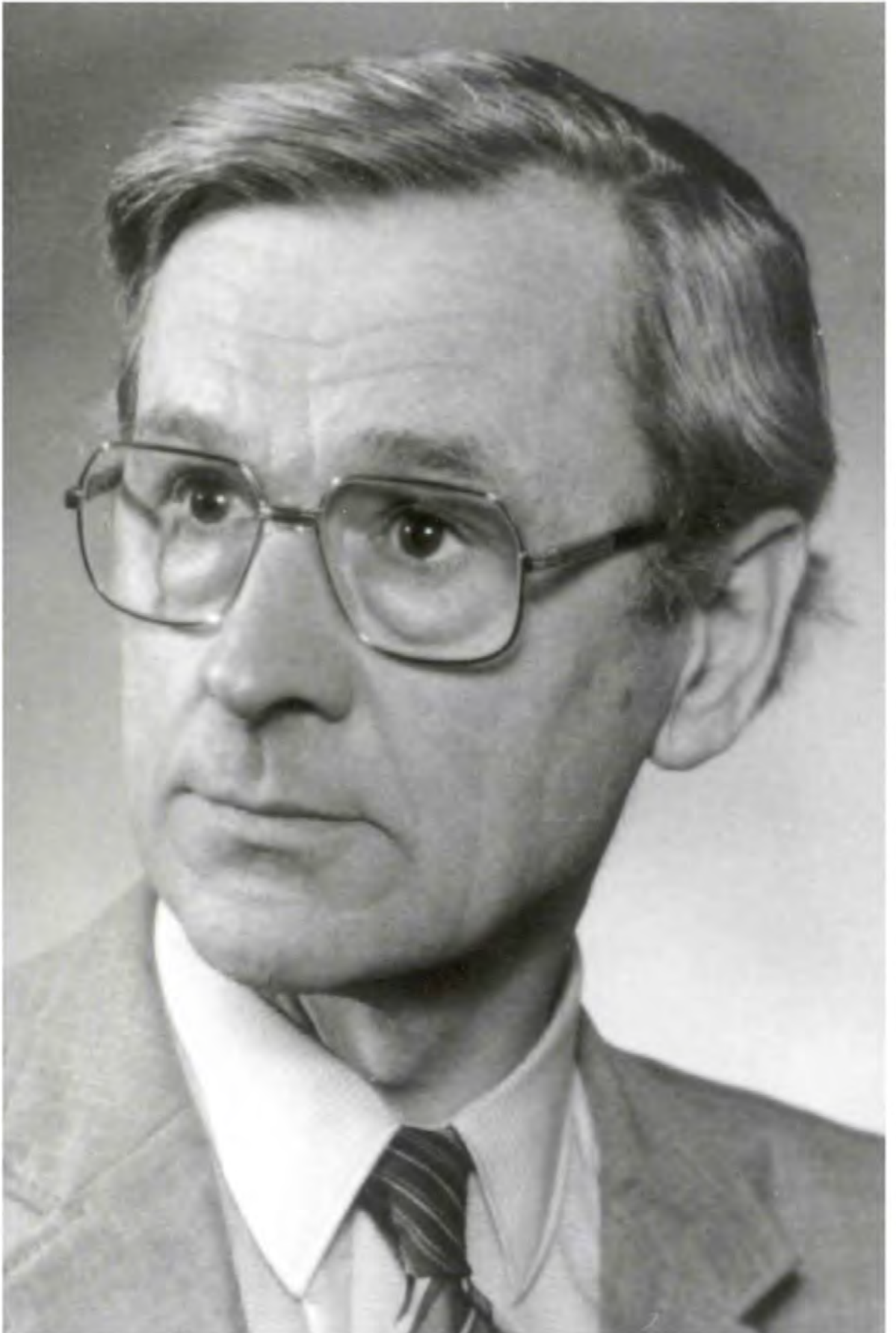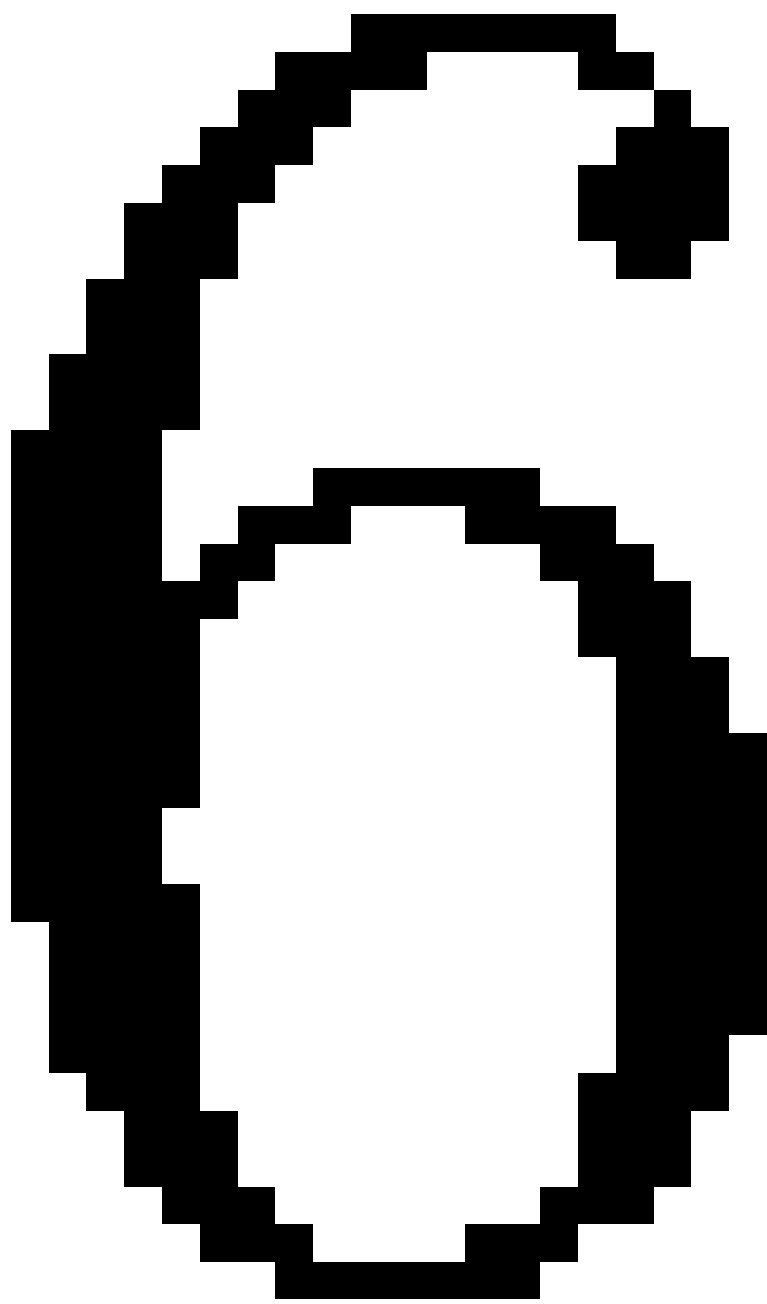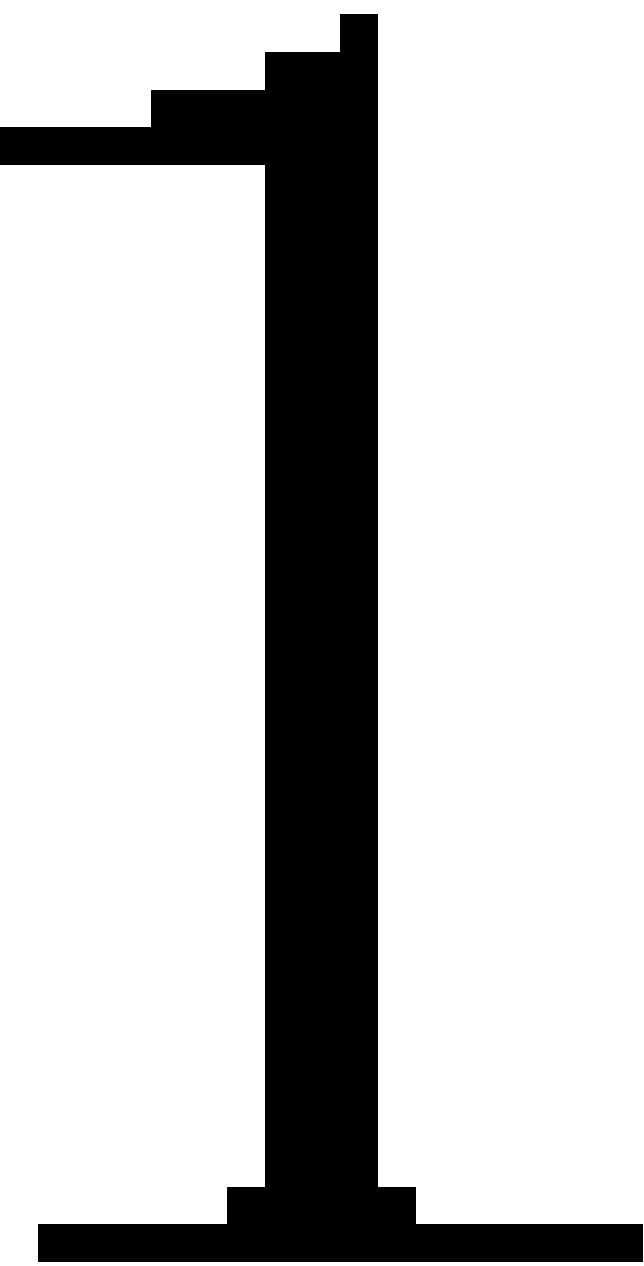
Preface

An *infinitesimal* is a 'number' that is smaller then *each* positive real number and is larger than *each* negative real number, so that in the real number system there is just one infinitesimal, i.e. zero. But most of the time only nonzero infinitesimals are of interest. This is related to the fact that when in the usual limit definition x is tending to c, most of the time only the values of x that are different from c are of interest. Hence the real number system has to be extended in some way or other in order to include all infinitesimals.

This book is concerned with an attempt to introduce the infinitesimals and the other 'nonstandard' numbers in a naive, simpleminded way. Nevertheless, the resulting theory is hoped to be mathematically sound, and to be complete within obvious limits. Very likely, however, even if 'nonstandard analysis' is presented naively, we cannot do without the axiom of choice (there is a restricted version of nonstandard analysis, less elegant and less powerful, that does not need it). This is a pity, because this axiom is not obvious to every mathematician, and is even rejected by constructivistic mathematicians, which is not unreasonable as it does not tell us *how* the relevant choice could be made (except in simple cases, but then the axiom is not needed).

The remaining basic assumptions that will be made would seem to be acceptable to many mathematicians, although they will be taken partly from *formalistic mathematics* – i.e. the usual logical principles, in particular the principle of the excluded third – as well as from *constructivistic mathematics* – i.e. that at the start of all of mathematics the natural numbers (in the classical sense of the term) are given to us. Not only the natural number, but also the set and the pair will be taken as primitive notions. The net effect of this is a version of mathematics that, except for truly nonstandard results, would seem to produce the same theorems as produced by classical mathematics.

One of the consequences of combining ideas from the two main schools of mathematical thinking is that the usual axioms of set theory, notably those due to Zermelo and Fraenkel, will be ignored. First of all, there will be elements that are

not sets, the natural numbers to begin with, only then sets will be formed from them in stages (or day by day), whereas when starting from the Zermelo-Fraenkel axioms each mathematical entity, in particular each natural number, is some set. From a formal point of view the latter has the advantage that there is just one primitive notion, but from a naive point of view it is not so obvious why numbers should be sets (in formalistic mathematics after the natural numbers come to life in the form of sets, this fact is concealed as soon as possible). Moreover, aren't we presupposing at least the order of the natural numbers already when writing down axioms by means of suitable symbols?

To a certain extent nonstandard analysis is superfluous! For if a theorem of classical mathematics has a nonstandard proof, it also has a classical proof (this follows from what in nonstandard analysis is known as the 'transfer' theorem). Often the nonstandard proof is intuitively more attractive, simpler and shorter, which is one of the reasons to be interested in nonstandard analysis at all. Another reason is that totally new mathematical models for all kinds of problems can be (and in the mean time have been) formulated when infinitesimals or other nonstandard numbers occur in such models. A trivial example is a problem involving a heap of sand containing *very* many grains of sand, but where the number of grains of sand must not be infinite. Then taking the inverse of some positive infinitesimal and rounding the result up or down produces a so-called infinitely large 'natural number' that is larger than *each* ordinary natural number, but is smaller than infinity. It can be manipulated in much the same way as the ordinary numbers, which cannot, of course, be said of infinity. As a consequence the mathematics of infinitely large sets is essentially simpler than that of infinite sets. A peculiarity, however, is that the 'selected' infinitesimal and hence the infinitely large natural number are not specified the way the number of elements of a set of, say, 25 elements is specified. On the other hand, if ω is that infinitely large natural number, it makes sense to consider another heap of sand with ω^2 grains of sand, that can be thought of as the result of combining ω heaps of sand each containing w grains of sand. But in what follows the analysis of practical models containing nonstandard numbers will not be stressed.

Chapter 1

Generalities

1.1 Infinitesimals and other nonstandard numbers: getting acquainted

An *infinitesimal* is a number that is smaller than *every* positive real number and is larger than *every* negative real number, or, equivalently, in absolute value it is smaller than $1/m$ for *all* $m \in \mathbb{N} = \{1, 2, 3, \ldots\}$. Zero is the only real number that at the same time is an infinitesimal, so that the nonzero infinitesimals do not occur in classical mathematics. Yet, they can be treated in much the same way as can the classical numbers. For example, each nonzero infinitesimal ε can be inverted and the result is the number $\omega = 1/\varepsilon$. It follows that $\mid w \mid > m$ for *all* $m \in \mathbb{N}$, for which reason ω is called (positive or negative) *hyperlarge* (or infinitely large). Hyperlarge numbers too do not occur in classical mathematics, but nevertheless can be treated like classical numbers. If, for example, ω is positive hyperlarge, we can compute $\sqrt{\omega}$, $\omega/2$, $\omega - 1$, $\omega + 1$, 2ω, ω^2, etc., and we have $(\omega - 1) + (\omega + 1) = 2\omega$, $(\omega - 1) \cdot (\omega + 1) = \omega^2 - 1$, etc. Also, for all $m \in \mathbb{N}$,

$$m < \sqrt{\omega} < \omega/2 < \omega - 1 < \omega < \omega + 1 < 2\omega < \omega^2$$

giving seven different hyperlarge numbers. The positive hyperlarge numbers must not be confused with infinity (∞), which should not be regarded a number at all, and which anyway does not satisfy these inequalities, except the first one.

Regrettably, there does not seem to exist a synonym for 'hyperlarge number' that would make a nice pair with 'infinitesimal', so let us introduce the synonym '*hypersmall* number' for the latter.

If ε is hypersmall, if δ too is hypersmall but nonzero, and if ω is positive hyperlarge, so that $-\omega$ is negative hyperlarge, we write,

$$\varepsilon \simeq 0, \ \delta \sim 0, \ \omega \sim \infty, \ -\omega \sim -\infty \text{ respectively.}$$

It would be wrong, of course, to deduce from $\omega \sim \infty$ that the difference between ω and ∞, or that between $-\omega$ and $-\infty$ would be hypersmall.

Given any $x \in \mathbb{R}$, $x \neq 0$, and any $\delta \simeq 0$, let $t = x + \delta$, then,

$$\varepsilon < \mid t \mid < \omega,$$

for all $\varepsilon \sim 0$ and all $\omega \sim \infty$. The number t is called *appreciable* (as it is not too small and not too large).

Three nonoverlapping sets of numbers (old or new) can now be presented:

a)   the set of all infinitesimals, to which zero belongs,
b)   the set of all appreciable numbers, to which all nonzero reals belong, and
c)   the set of all hyperlarge numbers, containing no classical numbers at all.

Together these three sets constitute the set of all numbers of 'real nonstandard analysis'. This set, which clearly is an extension of $\mathbb{R}$ is indicated by,

$$^*\mathbb{R}$$

and is called the $*$-*transform* of $\mathbb{R}$. The elements of $^*\mathbb{R}$ are called *hyperreal*. The use of the prefix 'hyper' here is not entirely defendable, as, say, 5, which obviously is an element of $^*\mathbb{R}$, is just an ordinary real.

Abbreviating hypersmall, appreciable, and hyperlarge to s, a and l, respectively, and assuming that x and y are positive numbers, for addition and multiplication the following holds,

$y \backslash^x$	s	a	l
s	s	a	l
a	a	a	l
l	l	l	l

addition

$y \backslash^x$	s	a	l
s	s	s	?
a	s	a	l
l	?	l	l

multiplication

where the quotation marks stand for s or a or l. Examples for the lower left quotation mark are $x \sim 0$ and $y = \sqrt{x^{-1}}$, or $1/x$, or $1/x^2$. For $x - y$ the results are the same as for $x + y$ (if still $x, y > 0$), except that if both x and y are appreciable, then $x - y$ is either hypersmall or appreciable, and that if both x and y are hyperlarge, then $x - y$ is either hyperlarge (positive or negative), or appreciable, or hypersmall, as is shown by the following examples: $y = x/2$, or $2x$, or $x - 1$, or $x + \varepsilon$, with $\varepsilon \simeq 0$.

If a number is not hyperlarge it is called *finite* or *limited*.

Remark: Elsewhere in the literature, *any* element of $^*\mathbb{R}$ is called finite.

Clearly, t is finite if and only if $t = x + \varepsilon$ for some $x \in \mathbb{R}$ and some $\varepsilon \simeq 0$. Given such a t, both x and ε are unique, for,

$$x + \varepsilon = y + \delta, \ x, y \in \mathbb{R}, \ \varepsilon, \delta \simeq 0$$

implies that $x - y = \delta - \varepsilon \simeq 0$, so that (as $x - y \in \mathbb{R}$), $x - y = 0$, hence $x = y$ and $\varepsilon = \delta$. By definition x is called the *standard part* of t, and this is written as,

$$x = \text{st}(t).$$

The standard part function st provides an important (mainly one-way) bridge between the finite numbers of nonstandard analysis and the classical numbers. Trivially, if t is itself a classical number, then $st(t) = t$.

1.2   Other ∗-transforms; generating new numbers

The ∗-transform not only can be obtained for $\mathbb{R}$ but also for $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$, and in fact any set X of classical mathematics (and for much more, see Section 1.5). Their ∗-transforms are indicated by $^*\mathbb{N}, {}^*\mathbb{Z}, {}^*\mathbb{Q}$, and $^*X$, respectively. Throwing all non-finite numbers out of $^*\mathbb{N}$ and $^*\mathbb{Z}$ we obtain again $\mathbb{N}$ and $\mathbb{Z}$, but something similar is not true for $^*\mathbb{Q}$ (for $^*\mathbb{R}$ we know this already), simply because $^*\mathbb{Q}$ (just as $^*\mathbb{R}$) contains finite non-classical numbers. Yet there is a striking difference between $^*\mathbb{Q}$ and $^*\mathbb{R}$ in this respect: the 'standard part theorem' discussed at the end of the preceding section does not hold for $^*\mathbb{Q}$, that is to say, there are finite elements t of $^*\mathbb{Q}$ that cannot be written as $t = x + \varepsilon$, with $x \in \mathbb{Q}$, $\varepsilon \in {}^*\mathbb{Q}$, $\varepsilon \sim 0$. For let c be any irrational number, say $c = \sqrt{2}$, and let $(r_1, r_2, \ldots)$ be some Cauchy sequence of rationals converging to c. Later on it will become clear that then the sequence $(r_1 - c, r_2 - c, \ldots)$ 'generates' an infinitesimal δ in $^*\mathbb{R}$ (because this sequence converges to zero). On the other hand $(r_1, r_2, \ldots)$ generates an element $r \in {}^*\mathbb{Q} \subset {}^*\mathbb{R}$, and r is finite (because the r_i are rational, and this sequence converges), but it has no standard part in Q, for otherwise $r = x + \varepsilon$ for some $x \in \mathbb{Q}$ and some $\varepsilon \in {}^*\mathbb{Q}$, $\varepsilon \simeq 0$. But $(r_1 - c, r_2 - c, \ldots)$ also generates the finite number $r - c \in {}^*\mathbb{R}$, so that $r - c = \delta \simeq 0$. It follows that $x - c = \delta - \varepsilon \simeq 0$, hence $x - c = 0$ (as $x - c$ is an ordinary real), which would mean that $c \in \mathbb{Q}$, a contradiction. On the other hand, in $^*\mathbb{R}$ we have that $\text{st}(r) = c$. (Carrying this argument further it turns out that there exists a $1 - 1$ mapping between $\mathbb{R}$ and the set of all finite elements of $^*\mathbb{Q}$ modulo the set of all rational infinitesimals, preserving addition and multiplication; i.e. the mapping is an isomorphism. In other words, $\mathbb{R}$ (not $^*\mathbb{R}$) can in a sense be produced by $^*\mathbb{Q}$.)

There are various ways to introduce the new numbers. Below this will be done by means of infinite sequences of classical numbers. In particular, the elements of $^*\mathbb{R}$ will be generated by means of infinite sequences of reals, and it will be necessary to consider *all* such sequences. (Recall that the elements of $\mathbb{R}$ can be generated by means of rather *special* infinite sequences of rationals, i.e. the Cauchy sequences.) More generally, given any classical set X the elements of its $*$-transform $^*X$ will be generated by means of infinite sequences of elements of X, and again *all* such sequences must be taken into account. Each such sequence 'generates' an element of $^*X$, and in case X is a set of numbers (or n-tuples of numbers) special sequences generate the elements of X itself. For example, $(1, 2, 3, \ldots)$ generates a hyperlarge element of $^*\mathbb{N}$, and $(3/2, 5/4, 9/8, \ldots)$ generates a finite element of $^*\mathbb{Q}$, that is equal to the sum of 1, generated by $(1, 1, 1, \ldots)$ and an infinitesimal, generated by $(1/2, 1/4, 1/8, \ldots)$. Different sequences may generate the same element of $^*X$. In fact, given any $x \in {}^*X$ there are many (uncountably many) different sequences that generate x (if X contains at least two elements). For example, changing finitely many terms of a generating sequence has no effect on the element generated. But there are many more variations on this theme. Wouldn't it be possible to restrict ourselves to a suitable subset of all sequences? Unless we are satisfied with some sort of mutilated nonstandard analysis, most likely the answer is 'no'. See Section 4.4.

Anyway, the nuisance of having to use generating sequences is only temporary. Once the new numbers have been introduced (as well as new functions, etc.) in most cases it is not necessary at all to know that they came about by means of infinite sequences. The situation is entirely analogous to that of introducing the real numbers: most of real analysis can be developed without the interference of Cauchy sequences. Most of the time an irrational such as $\sqrt{2}$ is treated as just a number, not as a sequence.

Although $^*\mathbb{N}$, $^*\mathbb{Z}$, $^*\mathbb{Q}$ and $^*\mathbb{R}$ are extensions of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$, respectively, in general $^*X$ is not always an extension of X. If, for example $X = \{\mathbb{N}\}$, then $^*X = \{^*\mathbb{N}\}$, and since $\mathbb{N} \neq {}^*\mathbb{N}$, X is not contained in $^*X$.

1.3   Bound and free variables; prenex normal form

At this point we must interrupt the main subject to say a few words about the formulation of mathematical statements and the difference between bound and free variables occurring in them. It will be convenient to use the *logical connectives* $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ (for NOT, AND, OR, IMPLIES, IS EQUIVALENT TO, respectively), the *universal quantifier* $\forall$ and the *existential quantifier* ('$\forall x : \ldots$' means: 'for all

x such that ...', and '$\exists x : ...$' means 'there exists an x such that ...'). Apart from operations, logical connectives and quantifiers, each mathematical statement contains a number of variables and constants, which may be (expressions of) numbers, or functions, or n-tuples of numbers or functions, or sets of numbers or functions, etc.

For example, in,

$$\exists x : [x \in \mathbb{R} \wedge x^2 + y = 0],$$

the constants are $\mathbb{R}$, 2, and 0 and the variables are x and y (unless y is fixed; then it may be regarded as an unspecified constant). Clearly, this statement changes into another meaningful statement if y is replaced by another variable or by some specified constant, and the same is true if one or more of the constants are replaced by other constants or by variables, at least within reasonable limits:

$$\exists x : [x \in \mathbb{R} \wedge x^2 + z = 0],$$

or,

$$\exists x : [x \in \mathbb{Q} \wedge x^2 + y = 0].$$

Obviously, the new statement may not be equivalent to the given one, but that is not important in the present discussion (in fact the last statement *is* different from the given one, and the last but one is in case z is a variable that is different from the variable y). On the other hand, (only) replacing x by a constant leads to nonsense. Moreover, replacing the symbol 'x' by, say, 'z' leads to an equivalent statement:

$$\exists z : [z \in \mathbb{R} \wedge z^2 + y = 0],$$

for it does not matter what symbol is used to indicate the relevant variable (again within reasonable limits). For these reasons, with respect to the given statement x is called a bound variable, and y is called a free variable. More generally, given any statement, if replacing any variable occurring in it by some constant leads to another meaningful statement that variable is called a *free variable* with respect to it, and any other variable occurring in it is called a *bound variable* (or a *dummy variable*) with respect to it. Any constant might be called *free* as well. It follows that the truth value of a statement depends on its free variables and its constants, but not on its bound variables.

Furthermore, it follows that the x in '$\forall x$' or in '$\exists x$' is a bound variable, but the converse does not seem to be true. For example, in a limit definition the clause

'for x tending to c' may be occur. Then x is a bound variable. And x may also be a bound variable in the definition of an integral, namely if 'dx' occurs in the well-known way. If limits and integrals are replaced by their definitions, however, x will be 'bound' to a quantifier. How about x in,

$$x \in \mathbb{R} \Rightarrow x^2 \geq 0?$$

Many people will interpret this as $\forall x : [x \in \mathbb{R} \Rightarrow x^2 \geq 0]$, and then x is a bound variable, yet x is a free variable simply because, say, $5 \in \mathbb{R} \Rightarrow 5^2 \geq 0$, is a meaningful (though uninteresting) statement. Note that many theorems take the form: 'if x satisfies ..., then ...', which usually is meant to be read as: 'for all x such that ..., it is true that ...'.

Some confusion may arise because of the possibility to indicate different bound variables by the same symbol (with free variables this is not possible), as in,

$$[\forall x : [x \in \mathbb{R} \Rightarrow x^2 \geq 0]] \wedge [\exists x : [x \in \mathbb{R} \wedge x^2 + 1 = 0]]$$

which is equivalent to,

$$[\forall x : [x \in \mathbb{R} \Rightarrow x^2 \geq 0]] \wedge [\exists y : [y \in \mathbb{R} \wedge y^2 + 1 = 0]].$$

Let us agree to avoid this, and in this example reject the first formulation.

Fortunately, any statement can be rewritten in its so-called *prenex normal form*. This is a unique rearrangement of the statement, whereby the quantifiers precede the logical connectives. Details regarding the existence and the uniqueness of the prenex normal form can be found in books on formal logic. For example, the following statement is in prenex normal form,

$$\forall x : \exists y : \forall z : P(c, d, \ldots, s, t, \ldots, x, y, z),$$

where $P(c, d, \ldots, s, t, \ldots, x, y, z)$ is a statement that contains no quantifiers at all, and apart from its free (!) variables x, y, and z, contains the constants $c, d, \ldots$ and other free variables $s, t, \ldots$, as well as logical connectives. It is clear that with three quantifiers there are eight possible normal forms if everything except the quantifiers is ignored: $\forall\forall\forall$, $\forall\forall\exists$, $\forall\exists\forall$, $\exists\forall\forall$, $\forall\exists\exists$, $\exists\forall\exists$, $\exists\exists\forall$, $\exists\exists\exists$.

It is better, however, to consider the following variations, where for clarity in P there is only one constant c and only one additional free variable s,

$$\forall x \in X : \exists y \in Y : \forall z \in Z : P(c, s, x, y, z),$$

where X, Y and $z \in Z$ are properly selected sets, or,

$$\forall x \in X,\ A(x) : \exists y \in Y,\ B(x, y) : \forall z \in Z,\ C(x, y, z) : P(c, s, x, y, z),$$

where $A(x)$, $B(x,y)$, and $C(x,y,x)$ are relatively simple conditions regarding the bound variables involved (such as $x > 0$), conditions that do not contain quantifiers. This is to be read as:

> *for all x in X satisfying $A(x)$ it is true that there exists a y in Y satisfying $B(x,y)$ such that for all z in Z satisfying $C(x,y,z)$ it is true that $P(c,s,x,y,z)$.*

The reason why set inclusions should be specified explicitly is to avoid certain errors in nonstandard analysis, errors that cannot occur in classical mathematics. Even if x is a subset of a larger set Y, it is better to replace '$x \subset Y$' by '$x \in \mathcal{P}(Y)$', where $\mathcal{P}(Y)$ is the power set of Y, i.e. the set of all subsets of Y. This will become clear when discussing the $*$-transform in Section 1.5, and the difference between internal and external constants in Section 1.6.

1.4   The purpose of nonstandard analysis

After the digression of the preceding section let us now contemplate the purpose of nonstandard analysis. Starting from $\mathbb{N}$, the sets $\mathbb{Z}, \mathbb{Q}$ and $\mathbb{R}$ (and $\mathbb{C}$, but below complex numbers will be ignored) have been introduced in classical mathematics in order to enrich mathematics with more tools and to refine existing tools. The introduction of negative numbers, of fractions, and of irrational numbers is felt as a strong necessity, and without it mathematics would only be a small portion of what it actually is. The introduction of $*\mathbb{N}$, $*\mathbb{Z}$, $*\mathbb{Q}$, and $*\mathbb{R}$, however, was not meant at all to enrich mathematics (at least not when it all started), but only to simplify doing mathematics. For as soon as notions like limit and continuity are involved, definitions in nonstandard analysis can be given a simpler form, and theorems can be proved in a simpler way. Often the simplifications are considerable. In one case the proof of a classical conjecture was found by means of nonstandard analysis, after which a classical proof was found as well. Moreover, both definitions and proofs receive a more natural appearance. This may even enhance the discovery of new facts.

In the mean time nonstandard analysis has also been applied in a more traditional way, namely to introduce new mathematical notions and models. Examples can be found in probability theory, asymptotic analysis, mathematical physics, economics, etc. In what follows the attention will primarily be focused, however, on simplifying mathematics, rather than on enriching it with new concepts.

As an example of a simpler definition, consider continuity. A function f from $\mathbb{R}$ tot $\mathbb{R}$ is continuous at $c \in \mathbb{R}$ if statement (1.1) holds,

$$\forall \varepsilon \in \mathbb{R}, \ \varepsilon > 0 : \exists \delta \in \mathbb{R}, \ \delta > 0 : \forall x \in \mathbb{R}, \mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon.$$

Now to f there corresponds a unique function $^*F$, called the $*$-transform of f, that is a function from $^*\mathbb{R}$ to $^*\mathbb{R}$, such that $^*f(x) = f(x)$ if $x \in \mathbb{R}$, and (1.1) is true if and only if (1.1), which is the $*$-transform of (1.1), is true,

$$\forall \varepsilon \in {}^*\mathbb{R}, \ \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \ \delta > 0 : \forall x \in {}^*\mathbb{R}, \ \mid x - c \mid < \delta : \mid {}^*f(x) - {}^*f(c) \mid < \varepsilon.$$

(More about the $*$-transform in the next section.)
Moreover, (1.1) is equivalent to the much simpler statement (1.1),

$$\forall \delta \in {}^*\mathbb{R}, \ \delta \simeq 0 : {}^*f(c + \delta) - {}^*f(c) \simeq 0.$$

Warning: The equivalence between (1.1) and (1.1) does in general not hold if c is replaced by a nonstandard number, or if f is replaced by a nonstandard function.

The essence of (1.1) is,

$$\delta \simeq 0 \Rightarrow {}^*f(c + \delta) - {}^*f(c) \simeq 0,$$

which is precisely what we want the definition of continuity to be: if $x - c = \delta$ is infinitely close to zero, then $f(x) - f(c)$ too should be infinitely close to zero. The only problem in classical mathematics is that 'infinitely close to' is not (and most likely will never be) a well defined notion. In nonstandard analysis, however, all that need be done is to replace 'infinitely close to' by '$\simeq 0$'.

Note that in all four definitions δ plays the same role (i.e. 'distance' from c), but that in (1.1) and (1.1) it is bound to $\exists$, whereas in Q it is bound to $\forall$. Also note that (1.1) and (1.1) each contain three quantifiers, but that (1.1) contains only one (that (1.1) contains no quantifiers at all is because it really is not complete as 'δ' is missing).

An illustration of a simpler proof is that of the intermediate value theorem: if $f : \mathbb{R} \to \mathbb{R}$ is continuous in the closed interval $[a, b]$, $a < b$, a and b both finite, and $f(a) < 0$, $f(b) > 0$, then $f(c) = 0$ for some $c \in [a, b]$.

A nonstandard proof of this theorem proceeds as follows. Let $m \in {}^*\mathbb{N}$ be hyperlarge. Divide $[a, b]$ in m equal subintervals, each of length $\delta = (b - a)/m$. Then $\delta \sim 0$. Let n be the smallest element of $^*\mathbb{N}$ such that $^*f(a + n\delta) > 0$, then $^*f(a + (n - 1)\delta) \leq 0$. Let $c = \text{st}(a + n\delta)$, then, by continuity,

$$^*f(a + n\delta) - {}^*f(c) = \varepsilon_1 \text{ and } {}^*f(c) - {}^*f(a + (n - 1)\delta) = \varepsilon_2,$$

for certain infinitesimals ε_1 and ε_2. Hence $-\varepsilon_1 < f(c) = {}^*f(c) \leq \varepsilon_2$. But $f(c) \in \mathbb{R}$, so $f(c) = 0$.

How come, dividing $[a, b]$ in m subintervals if m is not finite, and assuming that n, which also is not finite, exists? Yes, this is all right, because hyperlarge numbers behave like classical numbers.

The classical proof of the theorem is more involved, because it is based on the fact that a nonempty subset of $\mathbb{R}$ that is bounded above has a least upper bound (or supremum).

Exercise: Show this fact by means of nonstandard analysis.

1.5 More about the $*$-transform; transfer

So far a number of isolated instances of $*$-transforms have been presented (the $*$-transform of $\mathbb{R}$ and other sets, of functions from $\mathbb{R}$ to $\mathbb{R}$ and of statement (1.1) in the preceding section). Although it is too early to present a complete treatment of the $*$-transform, a number of interesting aspects of this notion may be discussed already now.

To each number, each set, each function, each operation (such as $+$ and $\cup$), each simple relation (such as $<$ and $\in$), each logical connective ($\neg$, $\vee$, $\wedge$, $\Rightarrow$, $\Leftrightarrow$), both quantifiers ($\forall$, $\exists$), each definition, and each statement of classical mathematics, there corresponds a unique $*$-transform in nonstandard mathematics. The notation is quite simple: just add an asterisk to the upper left of the symbol representing what is to be transformed. Sometimes the $*$-transform is identical to its inverse image, but often this is not so. In the former case the asterisk should, of course, be dropped, but even in the latter case this can sometimes be done without creating confusion. Below a number of typical examples is presented, but full details will only be given later on.

a)   *Numbers.* If $x \in \mathbb{R}$, then ${}^*x = x$.

b)   *Sets.* If X is a finite set of numbers, then ${}^*X = X$, and also (happily so) ${}^*\emptyset = \emptyset$, but if X is an infinite set of numbers, then X is strictly included in ${}^*X$ (in case X is an arbitrary abstract set and ${}^*X \neq X$, X need not be a subset of ${}^*X$.)

c)   *Pairs.* If $\langle x, y \rangle$ is a pair, then ${}^*\langle x, y \rangle = \langle x^*, y^* \rangle$, and similarly for n-tuples $\langle x_1, \ldots, x_n \rangle$.

d)   *Functions.* If $f : X \to Y$, then $^*f : {}^*X \to {}^*Y$, and $^*f(x) = f(x)$ if $x \in X$. Often the asterisk in $^*f$ may be dropped.

e)   *Operations.* As an example consider addition in $\mathbb{R}$. Its $*$-transform is $*$-addition in $^*\mathbb{R}$, and $x \,{}^*{+}\, y = x + y$ if $x, y \in \mathbb{R}$. The asterisk can safely be dropped.

f)   *Atomic relations.* These are relations in which neither logical connectives nor quantifiers play a part, but only such relations as $<$ or $\in$, etc. Consider first $<$ in $\mathbb{R}$, leading to $^* <$ in $^*\mathbb{R}$. Similarly as under e) we have that $x \,{}^*{<}\, y$ is equivalent to $x < y$ if $x, y \in \mathbb{R}$, and again the asterisk can safely be dropped. Next consider set inclusion. Let X be a subset of $\mathbb{R}$, then $\in X$ transforms to $^* \in {}^*X$. But ordinary set inclusion too is, of course, applicable to $^*X$, so that there would be two set inclusions for the $*$-transform $^*X$ of X. Fortunately, the two are identical, so that dropping the asterisk is a must.

g)   *The logical connectives,* and *both quantifiers.* For all of them the $*$-transform is identical to the inverse image, so that asterisks should be dropped.

h)   *Definitions.* For example continuity transforms to $*$-continuity and $^*f$, introduced in Section 1.4, is $*$-continuous at c if (1.1) is true.

i)   *Statements.* To some extent this covers, of course, case h). To find the $*$-transform of a statement (1.1), it should be formulated in such a way that each bound variable x occurs in some set inclusion of the form $x \in X$, *not* $x \subset Y$. Then the $*$-transform is obtained by replacing each constant and each free variable by its $*$-transform. As an example consider (1.1) and (1.1) defined in Section 1.4. Note that if in (1.1) '$\in \mathbb{R}$' would have been left out, something different would have been obtained. This is one of the reasons why at the end of Section 1.3 it was suggested to explicitly include each bound variable in some set inclusion. Why the inclusion $x \subset Y$ should be avoided will become clear in the next section.

One of the basic principles of nonstandard analysis is that any given classical statement (1.1) is true if and only if its $*$-transform is true, which results from (1.1) by replacing all its constants and free variables by their $*$-transforms. Note that the bound variables are not replaced. The principle is applied both ways, from $\mathbb{R}$ to $^*\mathbb{R}$, or from $^*\mathbb{R}$ to $\mathbb{R}$. In either case one says that the deduction is done by '*transfer*'. Assuming that everything is in prenex normal form, two simple nontrivial cases are,

$$\forall x \in X : P(x, s) \text{ and } \exists x \in X : P(x, s),$$

where X is some set and $P(x, s)$ is some atomic substatement with x a free variable and s a constant or a free variable. The $*$-transforms are,

$$\forall x \in {}^*X : P(x, {}^*s) \text{ and } \exists x \in {}^*X : P(x, {}^*s), \text{ respectively.}$$

Clearly, for each of the two classical statements transfer is trivial in one direction, assuming that X is a subset of ${}^*X$, but not necessarily in the opposite direction. The following two implications are the nontrivial ones,

$$[\forall x \in X : P(x, s)] \quad \Rightarrow \quad [\forall x \in {}^*X : P(x, {}^*s)],$$

$$[\exists x \in {}^*X : P(x, {}^*x)] \quad \Rightarrow \quad [\exists x \in X : P(x, s)].$$

Note that the first implication starts from a classical statement and leads to its $*$-transform, whereas the second one starts from a $*$-transform and leads to the corresponding classical statement.

In by far the most practical situations applying transfer is fairly obvious. In what follows transfer is applied in a slightly complicated situation, where it is required to show the equivalence of statements (1.1) and (1.1), as well as that (1.1) can be simplified to Q, with (1.1), (1.1) and (1.1) as in Section 1.4. Trivially, by transfer, (1.1) and (1.1) are equivalent, so it remains to show the equivalence of (1.1) and (1.1).

a) Let (1.1) be true, and let $\varepsilon \in \mathbb{R}$, $\varepsilon < 0$, and $\delta \in {}^*\mathbb{R}$, $\delta \simeq 0$ be arbitrary. Then for some $\delta' \in \mathbb{R}$, $\delta' > 0$,

$$\forall x \in \mathbb{R}, \mid x - c \mid < \delta' : \mid f(x) - f(c) \mid < \varepsilon,$$

hence, by transfer, and because ${}^*c = c$, ${}^*\varepsilon = \varepsilon$, ${}^*\delta' = \delta'$, ${}^*f(c) = f(c)$,

$$\forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta' : \mid {}^*f(x) - {}^*f(c) \mid < \varepsilon.$$

Let $x = c + \delta$, then, because by definition of infinitesimal $\mid \delta \mid < \delta'$, $\mid {}^*f(c + \delta) - {}^*f(c) \mid < \varepsilon$. But since ε is arbitrary, this means that ${}^*f(c+\delta) - {}^*f(c) \simeq 0$, and since δ is arbitrary that (1.1) is true.

b) Conversely, let (1.1) be true, and let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, and $\delta \in {}^*\mathbb{R}$, $\delta \sim 0$, $\delta > 0$, be arbitrary. For each $x \in {}^*\mathbb{R}$, $\mid x - c \mid < \delta$ it follows that $x - c = \delta'$ for some $\delta' \simeq 0$, hence, by (1.1), ${}^*f(x) - {}^*f(c) \simeq 0$, and, by definition of infinitesimal, $\mid {}^*f(x) - {}^*f(c) \mid < \varepsilon$. Apparently,

$$\exists \delta'' \in {}^*\mathbb{R}, \delta'' > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta'' : \mid {}^*f(x) - {}^*f(c) \mid < \varepsilon,$$

(take, for example, $\delta'' = \delta$), hence, by transfer (in the opposite direction as under a)),

$$\exists \delta'' \in \mathbb{R}, \delta'' > 0 : \forall x \in \mathbb{R}, \mid x - c \mid < \delta'' : \mid f(x) - f(c) \mid < \varepsilon.$$

Since is arbitrary this proves (1.1). $\qquad \square$

1.6 Standard, internal, and external constants

Each ∗-transform is called standard, because it corresponds in a $1-1$ way to some classical constant, and in a number of cases is even identical to that constant. In particular any set *X is standard. For example, *$\mathbb{N}$, *$\mathbb{R}$, and *$\mathcal{P}(\mathbb{R})$, with $\mathcal{P}(\mathbb{R})$ the power set of $\mathbb{R}$, are all standard. Note that the term 'standard' must not be used within classical mathematics. The reason for this is that, for example, a function f from $\mathbb{R}$ to $\mathbb{R}$, regarded as a function of nonstandard analysis, may not be standard, and usually it isn't. On the other hand, *f is standard, but this is a function from *$\mathbb{R}$ to *$\mathbb{R}$.

Each element (*not* subset) of a standard set turns out to be a special kind of constant of nonstandard analysis, namely a so-called *internal* constant, so that among others, infinitesimals and hyperlarge numbers are internal, because they are elements of *$\mathbb{R}$. Also the classical reals are internal, as $\mathbb{R} \subset$ *$\mathbb{R}$. Since *$x = x$ if $x \in \mathbb{R}$, the reals are also standard (as ingredients of nonstandard analysis). More generally, each standard constant happens to be internal, but the converse is not true, as is exemplified by infinitesimals and hyperlarge numbers.

Not every constant of nonstandard analysis is internal. For example, neither $\mathbb{R}$, nor the set of all infinitesimals, nor the set of all hyperlarge numbers is internal. Any constant that is not internal is called *external*.

Whereas internal constants behave like classical constants, external constants do not. They have extraordinary properties. For example, although $\mathbb{R}$ is a subset of *$\mathbb{R}$ that is bounded above in *$\mathbb{R}$ by any positive hyperlarge number, $\mathbb{R}$ has no least upper bound in *$\mathbb{R}$. For if b would be such a bound, then $b \sim \infty$, and $b - 1$ too would be an upper bound, but $b - 1 < b$. In a similar way it can be shown that the (bounded) set of all infinitesimals has neither a least upper bound nor a largest lower bound; and that *$\mathbb{N}\backslash\mathbb{N}$, the set of all hyperlarge natural numbers, has no smallest element, whereas each nonempty subset of $\mathbb{N}$ has such an element. Therefore, working with external constants or external variables that are not recognized as such is rather dangerous. Fortunately, the occurrence of external variables can be avoided by explicitly using in classical statements set inclusions of the type $x \in X$ (*not* $x \subset Y$), because then their ∗-transforms contain the inclusions $x \in$ *X, and x is automatically internal. If an inclusion like $x \subset Y$ is involved, it should be replaced by $x \in \mathcal{P}(Y)$, because an arbitrary subset of *Y need not be internal, as we have already seen. In this way external variables can literally be kept out of nonstandard analysis.

For any constant or variable, the diagram below shows the three main possibilities,

standard and internal	nonstandard and internal
nonstandard and external	

'Internal' may be read as 'mildly non-standard', and 'external' as 'extremely non-standard'.

1.7   Infinitesimals in Greek geometry?

Maybe it was Antiphon, a Greek mathematician and contemporary of Socrates, who for the first time contemplated the existence of infinitesimals. According to Heath [1] he, Antiphon, stated that, in Heath's words:

> "If one inscribed any regular polygon, say a square, in a circle, then inscribed an octagon by constructing isosceles triangles in the four segments, then inscribed isosceles triangles in the remaining eight segments, and so on 'until the whole area of the circle was by this means exhausted, a polygon would thus be inscribed whose sides, in consequence of their smallness, would coincide with the circumference of the circle'."

There are at least two interpretations in modern terminology of this. One is that the end product of Antiphon's construction is a polygon with a hyperlarge number of sides, so that the length of each side is a positive infinitesimal. But this would imply that the end product would not coincide with the circumference of the circle, that is, not exactly. The other one is that the end product is the circumference of the circle itself. But this would imply that the end product no longer was a polygon. Either interpretation contains a contradiction, so it is difficult to say what really was in Antiphon's mind.

Anyway, Antiphon's idea was not accepted by his fellow mathematicians. Again in Heath's words:

> "The time had, in fact, not come for the acceptance of Antiphon's idea, and, perhaps as the result of the dialectic disputes to which the notion of the infinite gave rise, the Greek geometers shrank from the use of such expressions as infinitely great and infinitely small and substituted the idea of things *greater* or *less than any assigned magnitude*. Thus, as Hankel says, they never said that a circle *is* a polygon with an infinite number of infinitely small sides; they always stood still before the abyss of the infinite and never

ventured to overstep the bounds of dear conceptions. They never spoke of an infinitely close approximation or a limiting value of the sum of a series extending to an infinite number of terms."

Note that the two interpretations mentioned above are also present in this quotation ('infinitely close' and 'limiting value').

Nevertheless, the Greek geometers solved many problems involving limits. They managed to do so by means of the so-called method of *exhaustion*. Given the problem to determine, say, the area of some figure, it is the method to find a sequence of inscribed figures as well as a sequence of circumscribed figures, each of known area, such that the given figure is approximated better and better by the terms of either sequence. But this does not mean that they thought in terms of limits. From the areas of the terms of both sequences they derived (guessed?) the area of the given figure, and a rigorous proof was obtained by showing that the proposed area of the given figure always lied between the areas of corresponding terms of both sequences. All that we can criticize is that they took the existence of the desired area for granted. In fact they managed to determine many limits without ever presenting a definition of limit.

Perhaps in his 'Methods' Archimedes comes closer to the use of infinitesimals. For example (see [1], Supplement, p. 15), when showing that the area of a segment ABC of a given parabola is 4/3 of the area of the triangle ABC, if, with D the middle point of the chord AC, BD is parallel to the axis of the parabola, Archimedes begins with some sort of plausible reasoning, where he states that the segment is made up of line segments between the parabola and the chord of the segment, all parallel to the axis of the parabola. Apparently, in his mind all these line segments together make up the entire segment of the parabola. It is tempting to conclude that the line segments were treated by him as parallelograms of hypersmall but positive breadth. At any rate, Heath ([1], Supplement, p. 8) writes that the line segments are

> "... of course ... indefinitely narrow strips (areas) ...; but the breadth ... (dx, as we might call it) does not enter into the calculation because it is regarded the same in each of the two corresponding elements which are separately weighed against each other, and therefore divides out."

If this would be correct Archimedese would have continued his plausible reasoning by showing that the parallelograms could each be 'weighed' (letting the area of a parallelogram be its weight) against one of the parallelograms making up a certain figure F. But the area of F could easily be shown to be equal to 4/3 of the area of triangle ABC.

There is an alternative, however, similar to the second interpretation mentioned earlier when discussing Antiphon's idea, where not parallelograms but line segments are weighed against each other (letting the length of a line segment be its weight). In fact Archimedes neither mentioned something like breadth, nor discussed dividing something out at all. Instead, he considered line segments making up certain areas, not thin parallelograms. True, in this case the number of line segments is infinite, so a limit is involved, but when working with parallelograms each individual comparison of weights is not exact. And since (as Archimedes remarks himself) the reasoning is not to be regarded as a rigorous one, it is not clear which interpretation is the right one. Anyway, Archimedes later on presented a rigorous proof – based on the method of exhaustion – where he could use the ratio 4/3 that he found by plausible reasoning.

Let us close this discussion with Heath's remark that Archimedes' 'Method' is a rare instance where a Greek mathematician shows how his intuition has led him to the solution of some problem by means of plausible reasoning. Usually, in Greek mathematics any trace of the intuitive machinery used was completely cleared away.

Open question: Have infinitesimals been wandering through the minds of some Greek mathematicians, or didn't they?

1.8 Infinitesimals in the 17th to the 19th century

There can be no doubt that in the 1670's, some 1900 years after Archimedes lived, infinitesimals were conceived by Leibniz. Moreover, he formulated their main properties, and many contemporary mathematicians as well as mathematicians after him, among them Euler and Cauchy, were able to successfully work with them. But the theory of the infinitesimals lacked a rigorous basis, and during some 200 years all trials to improve this situation were in vein, so that at last one gave up, the more so because in the 1870's Weierstrass came up with a rigorous theory of limits and continuity, which became the basis of what now is known as classical analysis, and where there was and is no need to consider infinitesimals any more.

It is quite interesting to see how Euler [2] shows the well-known product formula for the sine function. He begins his proof with the equality,

$$2 \cdot \sinh x = (1 + x/n)^n - (1 - x/n)^n,$$

valid for – in Eulers's own words – 'infinitely large values' of n. Obviously, this is only true up to an infinitesimal. Then the right-hand side is treated as if n were

a classical natural number. This leads after a purely classical reasoning to,

$$(1 + x/n)^n - (1 - x/n)^n = (8x/n) \cdot \prod_{k=1}^{m} \sin^2(k\pi/n) \cdot \{1 + x^2/n^2 \tan(k\pi/n)\},$$

where $m = (n-1)/2$, taking n odd (the details of the reasoning do not matter here, and the case for n even is similar). So,

$$\sin h \; x = (4x/n) \cdot \prod_{k=1}^{m} \sin^2(k\pi/n) \cdot \{1 + x^2/n^2 \tan^2(k\pi/n)\}.$$

Taking $x \neq 0$, and dividing by x, and then taking $x = 0$, gives,

$$1 = (4/n) \cdot \prod_{k=1}^{m} \sin^2(k\pi/n),$$

and hence,

$$\sin h \; x = x \cdot \prod_{k=1}^{m} \{1 + x^2/n^2 \tan^2(k\pi/n)\}.$$

Now for k finite, $n^2 \tan^2(k\pi/n)$ is 'infinitely close' to $(k\pi)^2$, so (?)

$$\sin h \; x = x \cdot \prod_{k=1}^{\infty} \{1 + x^2/k^2\pi^2\},$$

and putting $x = iz$, this gives the desired result,

$$\sin z = z \cdot \prod_{k=1}^{\infty} \{1 - z^2/k^2\pi^2\}.$$

Obviously, at the question mark the argument goes a little too fast, and a number of steps must be included here (see e.g. Luxemburg [3]).

Another famous example is Cauchy's proof ([4], p. 131), that a convergent series of continuous functions has a continuous limit function. To many this theorem was not correct, because it would seem that all kinds of counter-examples could be given. One of them is the series with the partial sums,

$$s_n(x) = (4/\pi) \cdot \sum_{k=1}^{n} \frac{\sin(2k+1)x}{2k+1},$$

that is periodic modulo 2π and converges to,

$$f(x) = \begin{cases} -1 & \text{if} \quad -\pi < x < 0 \\ 0 & \text{if} \quad x = 0 \text{ or } x = \pi \\ +1 & \text{if} \quad 0 < x < \pi \end{cases}$$

as can be shown by classical Fourier analysis. Since the sine function is everywhere continuous and $s_n(x)$ converges to $f(x)$ for n tending to ∞, according to Cauchy's theorem f ought to be continuous, which it isn't. But sofar, everything takes place within $\mathbb{R}$, and Cauchy let everything happen in what we have indicated by $^*\mathbb{R}$.

For him continuity of f at c meant that,

$$\forall x \in {}^*\mathbb{R},\ x \simeq c : f(x) \simeq f(c),$$

where, however, $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ and f need not be a standard function, and $c \in {}^*\mathbb{R}$, not only $c \in \mathbb{R}$, which is why his continuity is not *continuity (in nonstandard analysis it is called S-continuity; recall definition (1.1) in Section 1.4, where $c \in \mathbb{R}$ and a standard function was involved, so that there S-continuity was the same as *continuity).

And by convergence of $s_n(x)$ to $f(c)$ he meant that,

$$\forall n \sim \infty : s_n(c) \simeq f(c),$$

where again everything is in $^*\mathbb{R}$. Note that the Weierstrassian definitions of limit and continuity appeared half a century after Cauchy's book, so Cauchy in a sense 'had to' work with definitions of the kind given here.

Now, by transfer,

$$^*s_n(x) = (4/\pi) \cdot \sum_{k=1}^{n} \frac{{}^*\sin(2k+1)x}{2k+1}, n \in {}^*\mathbb{N},\ x \in {}^*\mathbb{R},$$

and

$$^*f(x) = -1,\ \text{or } 0,\ \text{or } +1,\ x \in {}^*\mathbb{R},$$

since if the range of a classical function f is finite, the range of its transform is the same as that of f.

Let $m \sim \infty$ be fixed, and let $x = c = 1/(2m)$, and $dt = 1/m$, so that $x \sim 0$, $dt \sim 0$. Then,

$$^*s_m\left(\frac{1}{2m}\right) = (2/\pi) \cdot \sum_{k=1}^{m} \frac{{}^*\sin(2k+1)dt/2}{(2k+1)dt/2} \cdot dt.$$

If we had that $m \in \mathbb{N}$, then the sum to the right would be an approximation of the Riemann-integral,

$$J = \int_0^1 \frac{\sin t}{t} \cdot dt,$$

and it should therefore not come as a surprise that it can be shown that the standard part of the right-hand side is exactly equal to $2J/\pi$, and hence,

$$^*s_m\left(\frac{1}{2m}\right) - 2J/\pi \simeq 0.$$

But by direct calculation it follows that $2J/\pi \neq -1$, 0, and $+1$, and since in particular for $c = 1/(2m)$, $^*f(c) = -1$, or 0, or $+1$ (-1 is in fact impossible), it follows that $^*s_n(c)$ does not converge to $^*f(c)$. Also, since for all n, $^*s_n(0) = 0$, $^*s_m(x)$ is not continuous at $c = 0$, so that the 'counter-example' does not satisfy the assumptions of Cauchy's theorem, and this is why Cauchy maintained his theorem against all criticism, but without basing his proof (and much of his other work) on a rigorous theory of the infinitesimals and other nonstandard numbers. For many interesting details, see Lakatos [5].

1.9   Infinitesimals in the 20th century

When in the 1870's Weierstrass formulated the well-known $\varepsilon - \delta$ definitions of limit and continuity, definitions that completely ignore nonstandard numbers, the dispute regarding infinitesimals quickly settled in their disadvantage, but only temporarily, for in 1961 Robinson [6,7] presented a mathematically sound theory of the nonstandard numbers. These works embody the first fairly complete analysis of the nonstandard numbers. Not only are they based on work of fore-runners, but also on an amount of mathematical logic that hitherto was unusual in mathematics. Only a few references should suffice here, see [8–12].

Robinson starts from the axioms of set theory due to Zermelo and Fraenkel, and the axiom of choice (called together the *ZFC* axioms), derives $\mathbb{R}$ in a classical kind of way, and then extends $\mathbb{R}$ to $^*\mathbb{R}$ by applying a rather considerable amount of mathematical logic, as indicated before.

Another way to define $^*\mathbb{R}$ was already indicated by Hewitt [10] and worked out by Luxemburg [13]. Here the *ZFC* axioms are again the point of departure, but the more usual line of mathematical thinking is followed. (Except for the *ZF* axioms, this way is also followed in the next chapter.)

Still another way to introduce $^*\mathbb{R}$ was found by Nelson [14]. Nelson adds three more axioms to the *ZFC* axioms, as well as a new symbol, st (for 'standard') that is used as a kind of label to distinguish standard constants from nonstandard constants. This leads directly to the set of all standard as well as all nonstandard constants, without the intermediary step of first introducing $\mathbb{R}$; consequently

in internal set theory $^*\mathbb{R}$ is denoted by $\mathbb{R}$, and similarly, $^*\mathbb{N}$ is denoted by $\mathbb{N}$, etc. Actually, the point of view of internal set theory is that the $\mathbb{N}$ of classical mathematics is the same as the $\mathbb{N}$ of nonstandard analysis; and that all that happens is that unexpected elements of $\mathbb{N}$ are discovered, elements that had always been there. In other words, according to this point of view, 0, 1, 2, etc. do not at all fill up $\mathbb{N}$ (see Robert [15] and F. Diener et G. Reeb [16]). The additional axioms make sure that transfer is guaranteed (axiom of 'transfer'), that nonstandard numbers exist (axiom of 'idealization'), and that unique standard sets can be derived from given sets (axiom of 'standardization'). Even though internal set theory uses relatively little of mathematical logic, the new axioms require some study, and do not seem to be as obvious as, for example, the axioms of Greek geometry:

Transfer: $\forall^{\mathrm{st}} t_1 \ldots \forall^{\mathrm{st}} t_k : [\forall^{\mathrm{st}} x : P(x, t_1, \ldots, t_k) \Rightarrow \forall x : P(x, t_1, \ldots, t_k)]$.

Idealization: $[\forall^{\mathrm{st}\ \mathrm{fin}} x : \exists x : \forall y \in z : P(x, y)] \rightarrow [\exists x : \forall^{\mathrm{st}} y : P(x, y)]$.

Standardization: $\forall^{\mathrm{st}} x : \exists^{\mathrm{st}} y : \forall^{\mathrm{st}} z : [z \in y \Leftrightarrow z \in x \wedge P(z)]$.

Here $^{\mathrm{st}} u$ means that the variable u must be standard, and similarly the label fin means that the corresponding variable must be finite (but beware, in internal set theory any hyperlarge natural number is finite, only the combination of standard and finite amounts to the classical notion of finiteness). Note that whereas $^{\mathrm{st}} u$ means that the variable u is standard, $^*\wedge$ means the variable $^*\wedge$ is standard, because st is a label but $*$ is a mapping. $P(\ldots)$ denotes a given internal statement, except in the last axiom, where $P(\ldots)$ may even be external (see Section 1.6).

In naive nonstandard analysis these three additional axioms are not assumed but derived from the existence of the natural numbers and the axiom of choice. Transfer has already been discussed; and idealization is used to prove the existence of nonstandard elements in any internal set with an infinite number of elements. Perhaps standardization is the most intriguing of the three because it contains a statement $P(z)$ that may be external. Reformulated naively it means that,

$$\forall^* x : \exists^* y : \forall^* z : [^* z \in {}^* y \Leftrightarrow {}^* z \in {}^* x \wedge P(^* z)],$$

where x, y and z are, of course, classical. Since always $^* s \in {}^* S$ if and only if $s \in S$, it follows that,

$$y = \{z \in x : P(^* z)\},$$

or equivalently,

$$^* y = {}^* \{z \in x : P(^* x)\},$$

which in internal set theory are illegal set formations. Here are a few examples, where x and y are still classical, but z need not be classical.

1) $P(z) \equiv z \in {}^*\mathbb{N} \wedge z$ is standard;
      then $x = \{1, 2, 3\}$ gives $y = x = 1, 2, 3\}$,
          $x = \mathbb{N}$ gives $y = x = \mathbb{N}$,
      and $x = \mathbb{R}$ also gives $y = \mathbb{N}$.

In fact $\mathbb{N}$ is the largest y that is possible for variable x.

2) $P(z) \equiv z \in {}^*\mathbb{N} \wedge z < n$, with $n \in {}^*\mathbb{N}$ given such that $n \sim \infty$;
      then the results are as under 1).

3) $P(z) \equiv z \in {}^*\mathbb{R} \wedge z \simeq 0$;
      then $y = \{0\}$ if $0 \in x$ and $y = \emptyset$ if $0 \notin x$.

For other details the reader should consult more adequate treatments of internal set theory.

In the mean time other versions of nonstandard analysis have been developed. In one of them external sets are 'legalized' by means of still other axioms, and another label, ext (for 'external').

By now many hundreds of publications have been devoted to nonstandard analysis: it is an established branch of mathematics.

No matter how infinitesimals are introduced, with or without the axioms of set theory, with or without extra axioms and new undefined symbols (st and ext), always the axiom of choice seems indispensible. If one tries to develop infinitesimal calculus without this axiom, it seems that one should be satisfied with a mutilated theory, as will be explained later on in Section 4.4. Here attempts by Chwistek [17,18] in this direction should be mentioned. In his 1926 paper Chwistek introduces new numbers by means of infinite sequences of classical numbers. These new numbers are called Progressionszahlen ('sequence numbers'), and equality for them is defined as follows. Let $N_i(\alpha_i)$ and $N_i(\beta_i)$ be two new numbers, then,

$$N_i(\alpha_i) = N_i(\beta_i) \text{ if and only if } \alpha_i = \beta_i \text{ for } i > n \text{ for some } n \in \mathbb{N}.$$

Something similar is done to define inequality, and an operation like addition is defined by,

$$N_i(\alpha_i) + N_i(\beta_i) = N(\alpha_i + \beta_i).$$

A classical function f is extended by means of,

$$f(N_i(\alpha_i)) = N_i(f(\alpha_i)).$$

The extended function happens to be quite similar to $^*\!f$, the $*$-transform of f. Even so not much new calculus is developed. An extension of $\mathbb{R}$ that includes all sequence numbers could be introduced, however.

In his 1948 book Chwistek spends less then ten pages on the subject, but nevertheless shows that he is well aware of the fact that 'infinitely small' numbers can be introduced, and he also introduces internal functions (called normal functions by him). Again there is no fully expanded calculus. Most likely, the deeper reason for this is that Chwistek defines (in)equality for his sequence numbers as indicated above. This definition has the advantage that the axiom of choice is not needed, but leads to rather serious problems, as will become clear in Section 4.4. It remains to remark that working with sequences is a technique used by Hewitt [10] and Luxemburg [13], and will be the technique of the next chapter, which is based on assumptions that from a naive, intuitive point of view are understandable, obvious, and acceptable, except perhaps the axiom of choice, and where everything that is not so obvious, such as transfer and all the rest, will be proved, rather than assumed.

1.10 Introducing infinitesimals by plausible reasoning; filters

Let f be a function from $\mathbb{R}$ to $\mathbb{R}$, and let c and b be real numbers. Then in today's notation the definition according to Leibniz's ideas that the limit of $f(x)$ for x tending to c is equal to b, is as follows,

$$\forall x,\ x - c \simeq 0 : f(x) - b \simeq 0.$$

As already mentioned, neither Leibniz nor anybody else at that time gave a mathematical sound definition of infinitesimal, and a dispute started, that temporarily stopped in the 1870's when Weierstrass was able to present a mathematically sound limit definition, that was completely void of infinitesimal thinking:

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall x,\ 0 <\mid x - c \mid< \delta :\mid f(x) - b \mid< \varepsilon,$$

where ε and δ too are real numbers. This definition leaves no room for misunderstanding, and is even intuitively clear: after that mr. E. specifies any $\varepsilon > 0$, mrs. D. is able to specify a $\delta > 0$, such that if x is within a distance δ of c (x need not be equal to c), then $f(x)$ is within a distance ε of b ($f(x)$ may be equal to b). Note that mrs. D. is able to come up with a $\delta > 0$ no matter how small $\varepsilon > 0$ turns out to be; in each case she has a response.

Leibniz's definition is more attractive: it contains only one quantifier, against three in Weierstrass's definition, and even though the latter is intuitively clear, it is not as easy to grasp, not even for beginning mathematics students (probably for many other students it never becomes entirely clear).

Let us now see how by means of plausible reasoning we can find a good definition of infinitesimal. First note that the $\varepsilon - \delta$ definition is of a dynamical nature. If mr. E. first presents $\varepsilon = \varepsilon_1$, after which mrs. D. presents $\delta = \delta_1$, he may present another $\varepsilon = \varepsilon_2$ that is so small that mrs. D. is forced to present another, i.e. smaller $\delta = \delta_2$ (excluding trivial cases where f is constant in a neighborhood of c). Repeating this again and again, infinite sequences $(\varepsilon_1, \varepsilon_2, \ldots)$ and $(\delta_1, \delta_2, \ldots)$ are the result, and mr. E. and mrs. D. are involved in some dynamical process. Obviously, $\varepsilon_1 > \varepsilon_2 > \ldots$, and $\delta_1 > \delta_2 > \ldots$ and it is clear that the ε's as well as the δ's tend to zero (again excluding trivial cases). Instead of considering the sequences $(\varepsilon_1, \varepsilon_2, \ldots)$ and $(\delta_1, \delta_2, \ldots)$, we can just as well consider the sequences $(f(x_1), f(x_2), \ldots)$ and $(x_1, x_2, \ldots)$ and require that if $x_i - c$ tends to 0 for i tending to infinity, but such that $x_i - c \neq 0$ for all i, then $f(x_i) - b$ too should tend to 0 for i tending to infinity. Let us abbreviate this to:

$$(x_i - c) \to 0 \Rightarrow (f(x_i) - b) \to 0,$$

where $(x_i - c)$ and $(f(x_i) - b)$ stand for the sequences $(x_1 - c, x_2 - c, \ldots)$ and $(f(x_1) - b, f(x_2) - b, \ldots)$, respectively. Now it so happens that the $\varepsilon - \delta$ definition is equivalent to,

$$\forall (x_i), \ (x_i - c) \to 0, \ x_i - c \neq 0 : (f(x_i) - b) \to 0.$$

Exercise: Show this.

This formulation contains only one quantifier, and in form comes quite close to Leibniz's definition. However, we have traded two quantifiers for two converging sequences. This is overcome by replacing sequences that converge to 0 by infinitesimals, or rather, by letting a sequence that converges to 0 generate an infinitesimal, which will have the effect that the dynamical process mentioned above is replaced by something static.

So, let each infinite sequence $(s_1, s_2, \ldots)$ of real numbers s_i that converges to 0 generate an infinitesimal, and let this number be indicated by $H(s_1, s_2, \ldots)$ or simply by $H(s_i)$, (with the H of Hyper).

As an example, let $s_i = 1/i$, then $H(s_i)$ is an infinitesimal, and since $1/i > 0$ for all i it should be positive. It is also reasonable to require that $H(0, 0, \ldots) = H(0) = 0$. Moreover, it should be possible to treat infinitesimals and other nonstandard

numbers as ordinary numbers. This requires that, for example, addition, subtraction and the greater-than relation be defined for them. The natural, but partly wrong, guesses (that are almost identical to Chwistek's definitions mentioned in Section 1.9) are,

$$\begin{aligned}
\text{addition:} \quad & H(s_i) + H(t_i) = H(s_i + t_i), \\
\text{subtraction:} \quad & H(s_i) - H(t_i) = H(s_i - t_i), \\
\text{greater than:} \quad & H(s_i) > H(t_i) \text{ if } s_i > t_i \text{ for all } i \in \mathbb{N}.
\end{aligned}$$

Note that so far everything (being positive, being zero, addition, subtraction, greater than) is defined *term* by *term*. Let us adopt this as a basic rule. Unfortunately, this leads to trouble. For let,

$$\begin{aligned}
(s_i) &= (1/1, 3/8, 1/4, 3/32, 1/16, 3/128, \ldots), \text{ and} \\
(t_i) &= (7/8, 7/16, 7/32, 7/64, 7/128, 7/256, \ldots).
\end{aligned}$$

Then both (s_i) and (t_i) converge to 0, and the terms of both (s_i) and (t_i) are all positive, and nicely decrease with increasing i. So $H(s_i) > 0$, $H(t_i) > 0$, $H(s_i) \sim 0$ and $H(t_i) \sim 0$, and there seems to be no reason at all to reject these sequences as generating sequences of certain infinitesimals. But now consider the difference of $H(s_i)$ and $H(t_i)$, which is generated by the sequence $(s_i - t_i)$, i.e. by,

$$(+1/8, -1/16, +1/32, -1/64, +1/128, -1/256, \ldots),$$

then, as everything should be defined term by term, the conclusion must be that $H(s_i - t_i)$ is different from zero, is not positive and is not negative, hence that $H(s_i - t_i)$ does not behave as an ordinary number. Where did we go wrong? The example given would not give trouble if the definition of greater-than were revised:

$$H(s_i) > H(t_i) \text{ if } s_i > t_i \text{ for all even } i,$$

hence if only the even indices were 'accepted' and the odd indices were 'rejected'. Then $H(s_i - t_i)$ would be negative. Not only would this revision be rather arbitrary, and could, for example 'even' have been replaced by 'odd', so that the odd indices would be accepted and the even indices would be rejected, it would not work either, as a new example could be concocted resulting in a hypernumber that was neither positive, nor negative, nor zero. The only way out would be that given any partitioning of $\mathbb{N}$ into two disjoint subsets of indices i, it would be allowed to accept one of these subsets (that is to say that all of its elements would be accepted) if it were infinite, and reject the other, and given any partitioning of any accepted subset into two disjoint subsubsets, it would again be allowed to

accept one of the latter if it were infinite and reject the other, such that any subset of a rejected subset is itself rejected, etc., etc., ad infinitum. That accepted subsets would have to be infinite is a reasonable requirement as infinitesimals should be generated by certain infinite sequences.

Even then the definition of greater-than would be rather arbitrary, but for all sequences (s_i), $H(s_i)$ would always be either positive, or negative, or zero. To see this, let $Q_0 = \{i : s_i > 0\}$. Then if Q_0 is accepted, hence $Q_1 = \{i : s_i \leq 0\}$ is rejected, $H(s_i)$ is positive. On the other hand, if Q_0 is rejected, so that Q_1 is accepted, then $H(s_i)$ is nonpositive. In the latter case, let $Q_{10} = \{i : s_i < 0\}$. Then if Q_{10} is accepted, hence $Q_{11} = \{i : s_i = 0\}$ is rejected, $H(s_i)$ is negative, and if Q_{10} is rejected, so that Q_{11} is accepted, then $H(s_i)$ is zero. Note that Q_0 and Q_1 are two complementary subsets of $\mathbb{N}$ and that Q_{10} and Q_{11} are two complementary subsets of Q_1. In this example it has tacitly been assumed that all the Q's involved are infinite sets. If this is not so fewer cases need be examined.

Note that a revision of the definitions for addition and subtraction would not be necessary, because although the attention could be restricted to some accepted subsets of indices i, there is no reason to do so.

Still there is some trouble. There is strong evidence that if we are fully free to accept or reject subsets of indices, but such that the given rules are obeyed ($\mathbb{N}$ is accepted; accepted sets must be infinite; of any two disjoint sets whose union is equal to an accepted set, precisely one must be accepted; subsets of rejected subsets must be rejected), we will never be able to give a complete specification of our choices or preferences. The meaning of this is that most likely we will never be able to write a computer program that contains *all* our preferences, and that, once the program has completely been finished and has been read in the relevant computer, it outputs either 'accepted' or 'rejected', after *any* subset Q of indices is presented as input to it.

Exercise: If this seems unbelievable, just try to give such a complete specification in the example above. See Section 1.16.

Although it seems not possible to specify all preferences in a constructive kind of way, it can be shown that a 'complete system of preferences' exists if the *axiom of choice* is invoked. The more technical term for such a system is *free ultrafilter*. But there are many such filters and when starting from one of them $H(s_i - t_i)$ may turn out to be positive, whereas when starting from another one $H(s_i - t_i)$ may turn out to be negative, because in the former case all odd indices and in the latter case all even indices would form an accepted set. *We shall have to live with this arbitrariness, however.* On the other hand, in practice only one free ultrafilter is needed. It is selected once and for all, and although there is much

arbitrariness in its selection, after the selection for each $H(s_i)$ it follows *uniquely* whether it is zero, or positive or negative, and similarly for all other choices with a finite number of alternatives that will present themselves. Ironically, we will never know the 'selected' filter completely, or rather we will only know it extremely incompletely. See Sections 1.15 and 1.16.

1.11 Basic assumptions of formalism

The two main schools of mathematical thinking are *formalism* and *constructivism*. They will be reviewed in the present and the next sections.

In formalism, which is the predominant school, basic assumptions are the *ZF* axioms of set theory due to Zermelo and Fraenkel (or equivalent variations thereof). In them each symbol that represents a variable, represents a set, and in each of them the symbol $\in$ occurs. Hence there is no reference at all to numbers, or geometrical concepts, or whatever. Instead of the word 'set' and the symbol '$\in$' a word and a symbol not yet existing could be used, because the axioms only have a formal meaning. Nevertheless they are intended to fix intuitive ideas regarding 'things contained in other things', but that is their semantic aspect. Even though it is often said that in formalism the notion of set is undefined, and that $\in$ is an undefined symbol, to a certain, but limited, extent they are defined implicitly by the axioms, because after all $x \in y$ must be a statement, i.e. something that has a truth value (but recall that logic too can be formalized).

Other basic assumptions are that from the axioms other truths can be derived by applying the well-known rules of logic, such as the syllogisms, the rules of substitution, and the principle of the excluded third (or middle).

In formalism logic prevails over mathematics, that is to say that mathematics is subject to all rules of logic. In constructivism the order is reversed, which has the consequence that each rule of logic has to be screened before it can be accepted as a rule for mathematical reasoning. This has resulted in the rejection of just one rule of logic, namely the principle of the excluded third.

Here are the first seven of the *ZF* axioms (see, for example, C.C. Chang and H.J. Keisler [19],

1. $\forall x, y : (x \equiv y \Leftrightarrow (\forall z : (z \in x \Leftrightarrow z \in y)))$.

2. $\exists x : \forall y : (\neg y \in x)$.

3. $\forall x, y : \exists z : \forall u : (u \in z \Leftrightarrow (u \equiv x \vee u \equiv y))$.

4. $\quad \forall x : \exists y : \forall z : (z \in y \Leftrightarrow (\exists w : (z \in w \wedge w \in x)))$.

5. $\quad \forall x : \exists y : \forall z : (z \in y \Leftrightarrow (\forall w : (w \in z \Rightarrow w \in x)))$.

6. $\quad \exists x : ((\exists y : y \in x) \wedge (\forall z : (z \in x \Rightarrow \exists w : (z \in w \wedge w \in x))))$.

7. $\quad \forall x : (\exists y : y \in x \Rightarrow (\exists z : (z \in x \wedge \neg(\exists w : w \in z \wedge w \in x))))$.

The formal character of these axioms may be emphasized by not mentioning their intuitive meaning, and by replacing $\in$ by any other suitable symbol. Whatever the interpretation of set and $\in$, it can be shown that the x whose existence is secured in Axiom 2, is unique: for assume that x and x' are such that,

$$\forall y : (\neg y \in x) \text{ and } \forall y : (\neg y \in x').$$

Then it has to be shown that $x \equiv x'$, or, by Axiom 1, that $\forall z : (z \in x \Leftrightarrow z \in x')$. But for any z it follows that $\neg z \in x$ and $\neg z \in x'$, hence for any z, both implications $z \in x \Rightarrow z \in x'$ and $z \in x' \rightarrow z \in x$ hold, so that the equivalence holds as well. Obviously, the intuitive meaning of the x of Axiom 2 is that of the empty set, and apparently the axioms dictate that there is a unique empty set, but for the proof the intuitive interpretation of neither set nor $\in$ is required, which illustrates the formalistic character of formalism.

Axiom 6, which is the 'infinity axiom', together with other axioms takes care of the existence of an infinite set, that, apart from the empty set $\emptyset$, contains the following sets as elements:

$$\{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \text{ etc.}$$

Here the set denoted by $\{\emptyset\}$, is defined by the requirements that $\emptyset \in \{\emptyset\}$ and that if $x \in \{\emptyset\}$ then $x \equiv \emptyset$, and similarly for the other sets.

Only now numbers appear; by definition,

$$0 = \emptyset, \ 1 = \{0\}, \ 2 = \{0, 1\}, \ 3 = \{0, 1, 2\} \text{ etc.}$$

In this way the natural numbers, including zero, appear as sets. Once in the possession of them the integers, the rationals, the reals, etc. can be defined in the well-known ways.

Most likely, Zermelo and Fraenkel's desire was to formulate a set of axioms that would lead to the natural numbers, but nothing more, i.e. to 0 and the elements of $\mathbb{N}$, but in fact the axioms happen to lead to 0 and $*\mathbb{N}$. After additional steps the latter leads to Nelson's internal set theory (see Section 1.9). If one insists on just $\mathbb{N}$, there is a way out: take the intersection of all sets that contain 0, 1, 2, etc., in other words, take the minimal infinite set whose existence follows from

Axiom 6. Altogether a fairly complicated way to get at the natural numbers and nothing more.

Moreover, one wonders whether in order to be able to even write down the *ZF* axioms the natural numbers are not at least implicitly presupposed. For this requires symbols to be written down one after the other in a linear order. There is a leftmost symbol, with just one right neighbor, for which the same is true, etc., until the rightmost symbol is reached (assuming sufficiently long lines). It would seem that here the natural numbers are implicitly used for ranking purposes. In addition, the required order not only is of a spatial nature, but of a timely nature as well, because the symbols must be written down after each other, presumably starting from the leftmost symbol, then proceeding with its right neighbor, etc. But here space and time are used as physical notions.

1.12 Basic assumptions of constructivism

Constructivism occurs in various forms, among which *intuitionism* and the theory of *recursive functions*, but only what they have in common will be briefly reviewed here. One of the starting points of constructivism is that before really starting mathematics the natural numbers, $1, 2, 3, \ldots$, are already given to us, simply because of our ability to *count*, and to do so *indefinitely*. This makes the 'infinity axiom' (Axiom 6 of the axioms listed in the previous section) superfluous, and has the agreeable consequence that the natural numbers are not sets. Nevertheless, there is no real difference between formalism and constructivism as far as the natural numbers in ordinary mathematics are concerned, since in formalistic mathematics the fact that the natural numbers are sets is largely ignored.

The same can be said about the rationals, but there is an essential difference as far as the reals are concerned, which is a consequence of the fact that the rationals form a countable set, but the reals do not. Yet the real numbers of both schools correspond to each other in a one-to-one kind of way. Differences appear if continua play a part. For example, according to both schools there exists a function f defined on the interval $[0, 1]$ of all rational numbers such that for all $x \in [0, 1]$, either $f(x) = 5$ or $f(x) = 7$, and such that both values are assumed somewhere. But as soon as this interval is replaced by the interval $[0, 1]$ of all real numbers, this is only true within formalism, because within constructivism such an f would not be accepted as a function, because it could not be defined constructively. In fact, within constructivism all functions defined on a finite real closed interval are continuous. For more details, see, for example, M.J. Beeson [20], or E. Bishop and D. Bridges [21].

More generally, a basic assumption (or perhaps restriction) in constructivism is that everything (definitions, proofs, etc.) must be of a constructive nature. For example, in formalism the following is a proper definition of p,

> p is the largest prime number such that $p-2$ too is a prime number, or if there is no such largest prime number, $p = 1$.

But in constructivism it is not, because it is not (yet) known whether or not the number of so-called twin primes is finite or not, hence (up to today) the definition is not constructive and must be abolished.

Another example is the 4-color problem. Let q be the minimum number of colors that is required to color any given map of countries, such that each country receives just one color, and neighboring countries receive different colors. Before 1977 q was not well defined within constructivism, because one did not know whether q was 4 or 5, but after 1977 q is also well defined in constructivism, because in 1977 Appel and Haken showed constructively that $q = 4$.

A consequence of constructivism is that it puts a restriction on the use of the logical principle of the excluded third. Consider some binary relation R, a statement,

$$\forall x \in \mathbb{R} : xRy, \tag{1.1}$$

and its converse,

$$\exists x \in \mathbb{R} : \neg xRy.$$

Then within formalism either (1.1) or (1.2), because of the principle of the excluded third. But within constructivism there are three possibilities: either (1.1) can be shown constructively, or (1.2) can be shown constructively, or neither (1.1) nor (1.2) can be shown constructively. The third possibility arises because a special kind of proof is required. It follows that within constructivism mathematics prevails over logic: the principles of logic must first be checked before they can be accepted as tools in mathematical reasoning. In fact, the principle of excluded third is the only logical principle that is rejected.

There is some flexibility in constructivistic thinking, though, as can be seen from the next example, taken from E.W. Beth [22], where it is required to show constructively that,

$$1^1 + 2^2 + \ldots + 99^{99}$$

is either dividable by 7 or not. Then all that is to be done is evaluate the indicated sum, divide it by 7 and see whether the remainder is zero or not. But the effort

involved is considerable, even if the computer is used (if the example is not impressive enough, replace 99^{99} by 999^{999}), but at this point the constructivist declares that the proposed procedure is a constructive one, because the division *could be* carried out!

The case is similar to the construction of a regular polygon with 65,537 ($= 1+2^{16}$) sides by means of ruler and compass (and paper and pencil) only, except that there seems to have been a mathematician who spent 10 years of his life checking this. He truly was a constructivist.

From the point of view of constructivism many statements of formalism are incorrect or even nonsensical. Yet, from the same point of view it seems impossible ever to derive a contradiction from any statement proved within formalism, because it would seem that the required proof would have to be based on the principle of the excluded third (see, for example, E.W. Beth [23]): constructivism nobly denies itself the indispensable weapon that it would need to defeat its enemy.

Although constructivism may seem a restricted kind of mathematics, it is sound mathematics and its achievements are remarkable. Nonstandard analysis would be impossible within it, however, because then the axiom of choice would be required, but, as we will see, this axiom is of a highly nonconstructive nature.

1.13 Selecting basic assumptions naively

From a naive, common sense point of view, both, formalism and constructivism, have agreeable as well as disagreeable characteristics. Fortunately, we are not bound to choose between the two. In this section a number of basic assumptions will be presented that combine the agreeable aspects of both schools, and that at the same time are easily understandable, and fairly obvious, perhaps with the exception of the axiom of choice, but this axiom is dictated because of the subject of this book. The net effect of this combining starting points of both schools will be that the ensuing mathematics is identical to the mathematics of the formalistic school. It is this combination that will serve as the basis of the theory of the next chapters.

Basic assumptions regarding logic. Assume that mathematics is subject to the principles of logic, in particular to the principle of the excluded third, which holds that any statement is either true or false; there is no third possibility.

Basic assumptions regarding the natural numbers. Assume that when starting mathematics the natural numbers, including 0, are given to us, and that they can

be used to count. Here 0 is the natural number that is used if there is nothing to count.

In formalistic mathematics the natural numbers are regarded as sets, but below this will not be done. The natural number is the first kind of *variable* that will be considered. As usual a variable may assume certain values, each value is a *constant*.

From the present basic assumption it follows that it is legitimate to use the well-known *inductive proof*. An inductive proof has the following structure.

> Suppose, we are concerned with infinitely many mathematical statements that can be counted by means of the natural numbers 0, 1, 2, ... Suppose further that $P(0)$ is the first statement to be counted, that $P(1)$ is the second one, etc. Assume that $P(0)$ is a true statement. Also assume that given any natural number n if $P(n)$ would be true then $P(n')$ would be true, where n' is the natural number that comes immediately after n. Then $P(m)$ is true for any natural number m.

Indeed, as $P(0)$ is true, it follows that $P(1)$ is true (taking $n = 0$ and hence $n' = 1$), but since $P(1)$ is true it follows that $P(2)$ is true (taking $n = 1$ and hence $n' = 2$), etc. Hence, when counting the given statements they can at the same time be proven to be true. Since by assumption all statements can be counted, they are all true; which shows the validity of the inductive proof.

For interesting arguments in favour of the present basic assumption, the reader may consult A. Heyting [24].

Basic assumptions regarding sets. The main difference between the assumptions that follow below and the *ZF* axioms of set theory, is that in the latter there are sets that have to be filled in with elements, or with no elements at all, whereas below there are elements (for example, the natural numbers) to begin with that may be taken together to form sets. As a consequence, in formalism any constant is a set, whereas below there are numbers, sets, n-tuples (among them pairs and triples), and functions (among them sequences), that each have there own intuitively determined character, although it should be admitted that the intuitive meaning of the various types of constants is not strictly required when deducing theorems from the axioms.

Existence (or specification). Given any moment in time, assume the possibility to either label or not label each of the then available constants separately, and to consider the *set* of precisely all unlabeled constants. A suitable label would

be: 'no'. The unlabeled constants are called the *elements* of the set. As usual the relationship between a set and its elements is indicated by means of the symbol $\in$.

Labeling all available constants leads to the existence of the *empty set*, indicated by $\emptyset$.

A well-known way to form sets is by a mathematical statement $P(x)$ depending on a single free variable x; x is labeled if and only if $P(x)$ does not hold.

Equality (or extensionality). Two sets X and Y are equal if and only if they have the same elements, or, more formally, if $x \in X$ implies that $x \in Y$ and vice versa. It follows that the empty set is unique.

Assume that any set is a constant, and hence that variables can be introduced whose values are sets, as well as sets whose elements are themselves sets, etc., but with the limitation that given any set 'taking an element of' can be plied a finite number of times only. Let the maximum number of times this can be done be called the *level* of the set, so that the lowest possible level is 1. As a natural number is not itself a set it is called an *urelement* or *individual*. In general, an urelement is any constant that is not a set. One might say that urelements are at level 0.

Without much loss of generality it may be assumed that all sets are *regular*. By definition a set of level k is called regular if all its elements are urelements (so that $k = 1$), or if $k \geq 2$ and all its elements are regular of level $k - 1$. Hence $\{x, \{x, y\}\}$ is not regular, but $\{\{x\}, \{x, y\}\}$ is if x and y both are urelements, or both are regular.

It follows that *subsets* and the *power set* of a given set can be formed, as well as the *complement* of a subset, and the *difference*, the *union* and the *intersection* of two sets. The notations are as follows.

> Sets: $\{x : P(x)\}$ (i.e. the set of all x such that $P(x)$), or $\{x \in X : \ldots\}$ (i.e. the set of all x such that $x \in X$ and such that $\ldots$), or $\{a, b, c, d\}$ (i.e. the set with the elements a, b, c, and d), or $\{a_1, a_2, a_3, \ldots\}$ (i.e. the set with the elements a_n for each natural number n), etc.

The subset and superset relations: $\subseteq$ and $\supseteq$.
The proper subset and proper superset relations: $\subset$ and $\supset$.
The power set of a set X : $\mathcal{P}(X)$ or 2^X.
The complement of a subset X : X^c.
The difference, the union, and the intersection operations: $-$, $\cup$ and $\cap$.

From these assumptions it follows that a notion like the set of all sets is nonsensical. This notion plays a role in one of the many *paradoxes* that were found after the first, somewhat careless, set-up of the theory of sets. The argument runs as follows. If there were a set X of all sets, consider $\mathcal{P}(X)$, then $\mathcal{P}(X)$ would be an element of X, as X is the set of all sets, but $\mathcal{P}(X)$ cannot be an element of X, contradiction. This paradox arises because the interference of time is ignored: if today the set X of all sets that are known sofar is introduced, only tomorrow $\mathcal{P}(X)$ can be introduced. The existence of this and other paradoxes has been an impetus to the development of both formalism and constructivism.

As usual, the symbol $\mathbb{N}$ is reserved for the set,

$$\{n : n \text{ is a natural number different from } 0\},$$

or, simply,

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

Similarly,

$$\mathbb{N}^o = \{0, 1, 2, \ldots\}.$$

Basic assumptions regarding n-tuples. Apart from natural numbers and sets also pairs, triples, and in general n-tuples are taken as primitive notions. The *ordered pair* or simply pair of two constants x and y is denoted by $\langle x, y \rangle$. The basic property of pairs is that $\langle x, y \rangle = \langle x', y' \rangle$ if and only if $x = x'$ and $y = y'$. Often, in particular within formalism, $\langle x, y \rangle$ is defined as the following set,

$$P = \{\{x\}, \{x, y\}\},$$

a definition that is due to Kuratowski. But there are many variations on this theme. One of the alternatives is to let $\langle x, y \rangle$ be the set,

$$P' = \{\{\{x\}, \emptyset\}, \{\{y\}\}\}.$$

It is not difficult to show that both (1.7) and (1.8) satisfy the basic property of pairs, but clearly $P \neq P'$. Why prefer (1.7) over (1.8)? In fact definitions of $\langle x, y \rangle$ as sets do too much. If $\langle x, y \rangle$ would be equated to (1.7), for example, then $\{\langle x, x \rangle\} = \{\{\{x\}\}\}$. This property is entirely accidental. See M.D. Potter [25], from which alternative (1.8) has been taken, for interesting details. For these reasons we follow Bourbaki's decision to regard the pair as a primitive term.

The basic property of the *triple* $\langle x, y, z \rangle$ is, of course, that $\langle x, y, z \rangle = \langle x', y', z' \rangle$ if and only if $x = x'$ and $y = y'$ and $z = z'$. Instead of regarding the triple as a

primitive notion, it is often defined as the pair $\langle x, \langle y, z \rangle \rangle$, although the obvious alternative is $\langle \langle x, y \rangle, z \rangle$. Combining the first choice with Kuratowski's definition the result would be that

$$\langle x, y, z \rangle = \{\{x\}, \{x, \{\{y\}, \{y, z\}\}\}\},$$

but why not let,

$$\langle x, y, z \rangle = \{\{x\}, \{x, y\}, \{x, y, z\}\}?$$

Similar remarks can be made with respect to the n-tuple $\langle x_1, \ldots, x_n \rangle$, given any $n \in \mathbb{N}$, $n \geq 4$.

For any $n \geq 1$, x_j is the j-th *term* of the n-tuple $\langle x_1, \ldots, x_n \rangle$, $j = 1, \ldots, n$.

Basic assumptions regarding generating new constants. Given any constant, assume the possibility to let a, possibly new, constant be *generated* by this constant, subject to the following two rules:

1) *Identification.* If the generated constant is not to be new it must be identified explicitly with a constant already known.

2) *Equality.* Equality of constants must be defined in such a way that all constants, new and old, satisfy the well-known rules of equivalence: $x = x$, $x = y \Rightarrow y = x$, and $x = y \wedge y = z \Rightarrow x = z$.

The character of a new constant is determined by these two rules, as well as by its intuitive interpretation. Anyway, a new constant is regarded as an urelement.

The axiom of choice. Assume that the axiom of choice holds true, i.e. that given any infinite set, whose elements are nonempty sets, it is possible to select an element from each of these elements.

As will be illustrated in Section 1.16, this axiom is of a highly nonconstructive character, but it will only be used to establish the existence of free ultrafilters.

1.14 Basic definitions

A) *Functions.* Given two nonempty sets X and Y, a *function* (or *mapping* or *map*) $f : X \rightarrow Y$ is generated by a set G with elements of the form $\langle x, y \rangle$, $x \in X$, $y \in Y$, such that for all $x \in X$ there is a unique $y \in Y$ with $\langle x, y \rangle \in G$. G is called the *graph* of the function f. Functions are regarded as new constants so that they are urelements, and two functions are equal

if and only if their generating sets are equal. The intuitive meaning of a function is that of an assignment: a function assigns a certain y to a given x. As usual the relationship between x and y is written as $y = f(x)$, or as $x \mapsto y$, or as $x \mapsto f(x)$. In case X is a subset of $\mathbb{N}$, and $n \in X$, $f(n)$ is often written as f_n. Also the definitions of *function value*, *domain*, *range*, *injective* (or *one-to-one*), *surjective* (or *onto*), *and bijective* (or *one-to-one onto*) functions, as well as the *inverse* of a bijective function are as usual.

B) *Sequences.* If, for $n \in \mathbb{N}$, $X = \{1, \ldots, n\}$ (or $X = \{0, 1, \ldots, n\}$), a function $f : X \to Y$ is called a *finite sequence*; and if $X = \mathbb{N}$ (or $X = \mathbb{N}^o$) then f is called an *infinite sequence*. The f_j for $j \in X$ are called the *terms* of the sequence. The usual notation of sequences is:

$$(f_1, \ldots, f_n) \text{ or } (f_1, f_2, \ldots),$$

and similarly in case $0 \in X$.

Note that although intuitively there may not be much difference between $\langle x_1, \ldots, x_n \rangle$ and $(x_1, \ldots, x_n)$, formally there is.

Infinite sequences will play a crucial part in what is going to follow.

C) *Kinds of sets.* Given a set X, if there exists a bijection from X onto $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, then X is called a *finite* set, otherwise an *infinite* set. Alternative formulations in these cases are that X contains a *finite* number of elements or an *infinite* number of elements, respectively. If X is an infinite set, and if there exists a bijection from X to $\mathbb{N}$, then X is called a *countably infinite* or *denumerably infinite* set, or simply *countable* or *denumerable*.

The *Cartesian* or *direct product* of two sets S and T is the set of pairs $\{\langle s, t \rangle : s \in S, t \in T\}$. The product is indicated by $S \times T$. A similar definition can be given for n sets by means of n-tuples, $n \geq 3$.

D) *Sets representing n-tuples or functions.* In what follows it will be convenient to relate n-tuples and functions in a one-to-one way to certain sets. The choice of these sets is completely arbitrary, as long as this requirement is satisfied. (If n-tuples and functions would have been defined as sets themselves, this representation would be superfluous!). Given the n-tuple $\langle x_1, \ldots, x_n \rangle$, let the set be,

$$\{\{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, \ldots, x_n\}\},$$

and given the function f let the set be its graph G, i.e. the set that generates f. Then there exists bijections T_n and F such that,

$$\langle x_1, \ldots, x_n \rangle = T_n(\{\{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, \ldots, x_n\}\}),$$

$$\{\{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, \ldots, x_n\}\} = T_n^{-1}(\langle x_1, \ldots, x_n \rangle),$$
$$f = F(G), \text{ and } G = F^{-1}(f).$$

1.15 Filters

In Section 1.10, where infinitesimals were introduced by plausible reasoning, the conclusion was reached that a set of 'acceptable' subsets of $\mathbb{N}$ was needed, satisfying the following rules:

1) $\mathbb{N}$ is accepted,
2) accepted sets are infinite,
3) of any two disjoint sets whose union is equal to an accepted set, precisely one is accepted, and,
4) subsets of rejected subsets are rejected.

Such a set was called a free ultrafilter (over $\mathbb{N}$). Hence U is a free ultrafilter if:

1) $\mathbb{N} \in U$,
2) if $Q \in U$, then Q is infinite,
3) if $Q = Q_1 \cup Q_2 \in U$, $Q_1 \cap Q_2 = \emptyset$, then either $Q_1 \in U$ or $Q_2 \in U$, but not both, and,
4) if $Q \notin U$ and $S \subseteq Q$, then $S \notin U$.

These requirements are equivalent to the following more usual ones. U is a *free ultrafilter* (over $\mathbb{N}$) if:

1a) $\mathbb{N} \in U$,
2a) if $Q \in U$ and if $R \supseteq Q$, then $R \in U$,
3a) if $Q \in U$ and if $R \in U$, then $Q \cap R \in U$,
4a) if $Q \in U$, then Q is infinite, and,
5a) if $Q \subseteq \mathbb{N}$, then either $Q \in U$ or $Q^c = \mathbb{N} - Q \in U$, but not both.

Proof of the equivalence:

A. Let 1) to 4) hold. Then 1a) and 4a) hold trivially and 5a) follows immediately from 1) and 3). If $Q \in U$ and $R \supseteq Q$, then $Q^c \notin U$ and $R^c \subseteq Q^c$, hence $R^c \notin U$ and $R \in U$, which proves 2a). Finally, if $Q, R \in U$, then $Q^c \notin U$, hence $Q^c \cap R \notin U$, hence $Q \cap R \in U$, which proves 3a).

B. The proof of the converse is left as an exercise. $\square$

The terminology used rightly suggests that there are more general filters (over $\mathbb{N}$). Here are their definitions.

Any set F of subsets of $\mathbb{N}$ is called a *filter* (over $\mathbb{N}$) if:

1b) $\mathbb{N} \in F$,
2b) $\emptyset \notin F$,
3b)   if $Q \in F$ and if $\mathbb{N} \supseteq R \supseteq Q$, then $R \in F$,
4b)   if $Q \in F$ and if $R \in F$, then $Q \cap R \in F$.

Note that 2b) and 4b) imply if $Q \in F$ then $Q^c \notin F$, but this is not to say that either $Q \in F$ or $Q^c \in F$.

A filter is called *free* if all of its elements are infinite.

Remark: Another definition is: a filter is called free if the intersection of all of its elements is empty. The two definitions are equivalent for an ultrafilter, but not for any filter.

Finally, a filter is called an *ultrafilter* if for any $Q \subseteq \mathbb{N}$, either $Q \in U$ or $Q^c \in U$, but not both.

Nonfree filters are not very interesting for our subject, but free filters and free ultrafilters are. An example of a free filter (that is not an ultrafilter) is the set $F^o = \{Q : [\exists k \in \mathbb{N} : Q \supseteq Q_k]\}$ with $Q_k = \{i : i \geq k\}$. This filter, which is called the *Fréchet* filter, can be used for the definition of converging sequences, and hence for that of infinitesimals as suggested by Chwistek (see Section 1.9). For nonstandard analysis the really important filters are the free ultrafilters, because with a filter like F^o only an incomplete kind of nonstandard analysis can be developed, as will be shown later on in Section 4.4.

Filters over a set M different from $\mathbb{N}$ can also be introduced. Then all that is needed is to replace $\mathbb{N}$ in the definitions above by M. Below this will only happen in the next section, where M is an infinite subset of $\mathbb{N}$.

Exercise: Show that if $Q = R_1 \cup R_2 \cup \ldots \cup R_n$, for some $n \in \mathbb{N}$, $R_i \cap R_j = \emptyset$ if $i \neq j$, and if F is an ultrafilter, then $R_i \in F$ for precisely one $i \in \{1, 2, \ldots, n\}$.

Exercise: Show that if F is an ultrafilter, and if $M \in F$, then an ultrafilter G over M is *induced* by F and M, if we let $G = \{Q : Q \subseteq M, Q \in F\}$. Also show that $G = \{Q \cap M : Q \in F\}$.

Exercise: Show that if F is a free ultrafilter, if $Q \in F$ and if $q_i \in Q$, $i = 1, 2, \ldots, n$, $n \in \mathbb{N}$, then $Q - \{q_1, \ldots, q_n\} \in F$.

Most likely, no completely specified examples of free ultrafilters can ever be given. Evidence for this statement is given in the next section. But if the axiom of choice is invoked their existence can nevertheless be shown.

Theorem 1.15.1 *Free ultrafilters over any infinite set exist.*

Proof: The proof is given in the appendix. $\qquad\square$

It is only in this proof that the axiom of choice is required in the version of nonstandard analysis that is presented in this book, assuming, of course, that when nonstandard analysis is used to prove a classical theorem, the axiom is not necessary in a classical proof of that theorem.

Since there are many free ultrafilters, it is necessary to select one. The selection is completely arbitrary, or rather the selection must of necessity be extremely arbitrary: apart from a tiny little bit we will not know which free ultrafilter we are really dealing with, so that, in fact, the term 'selection' is not quite adequate. This is a consequence of the highly nonconstructive character of the axiom of choice.

From now on assume that U is some free ultrafilter (over $\mathbb{N}$) and that it is fixed once for all.

Since U is not known, it is not known either whether the infinitesimal $s = H(+1, -1/2, +1/3, -1/4, \ldots)$, that is generated by the infinite sequence $(+1, -1/2, +1/3, -1/4, \ldots)$ is positive or negative. For if $Q = \{1, 3, 5, \ldots\} \in U$ then $s > 0$, and if $Q^c = \{2, 4, 6, \ldots\} \in U$ then $s < 0$, but we do not know whether or not Q is in U. Surprisingly enough, this and similar dichotomies will not hurt at all, simply because the final answers are independent from the underlying filter.

1.16 About the nature of free ultrafilters

As already mentioned there are many free ultrafilters U (over $\mathbb{N}$). Actually their number is uncountable. Although this fact is not very important for what is going to follow, so that the next argument may be skipped, a proof of this statement is given here, because it may give some more insight into the nature of free ultrafilters, in particular as far as nonconstructability is concerned.

First of all the next theorem is needed.

Theorem 1.16.1 *Let M be any infinite subset of $\mathbb{N}$, then there is a free ultra-filter U over $\mathbb{N}$ such that $M \in U$.*

Proof: Let $D = \mathbb{N} - M$, then if $Q \subseteq \mathbb{N}$ then $Q = Q' \cup R$ for unique $Q' \subseteq M$ and $R \subseteq D$. Let U' be a free ultrafilter over M, and let U be the set of all Q such that $Q' \in U'$. Then U is a free ultrafilter over $\mathbb{N}$, and $M \in U$. The verification that U indeed is a free ultrafilter is left as an exercise. $\qquad\square$

From this theorem it follows that if a free ultrafilter U over $\mathbb{N}$ is wanted, we can require beforehand that either $Q_0 = \{1, 3, 5, \ldots\} \in U$ or $Q_1 = \{2, 4, 6, \ldots\} \in U$, for simply let $M = Q_0$ or $M = Q_1$. This is also true for the other Q's that will be introduced. If we require beforehand that $Q_0 \in U$, then we can also require beforehand that either $Q_{00} = \{1, 5, 9, \ldots\} \in U$ or $Q_{01} = \{3, 7, 11, \ldots\} \in U$, and if we require beforehand that $Q_1 \in U$, then we can also require beforehand that either $Q_{10} = \{2, 6, 10, \ldots\} \in U$ or $Q_{11} = \{4, 8, 12, \ldots\} \in U$. These four cases can be split up into eight new cases in total, which in turn can be split up into sixteen new cases, and so on. Taking the n-th step, 2^n new Q's are added, each with n indices, each of which is 0 or 1. Apparently, we can require beforehand that, e.g. $Q_0, Q_{00}, Q_{000}, Q_{0000}, \ldots$ are all filter elements, but we can just as well replace this sequence of selections by $Q_0, Q_{01}, Q_{010}, Q_{0101}, \ldots$, and so on. Clearly, if, say $Q_{abcd} \in U$, the next Q is Q_{abcbe} with $e = 0$ or $e = 1$. It follows that each infinite sequence of 0's and 1's defines an infinite sequence of selections, and vice versa. Since different infinite sequences of 0's and 1's thus lead to different filters, and since the set of all these sequences is uncountable, it follows that there are uncountably many free ultrafilters over $\mathbb{N}$, as claimed. $\qquad\square$

Note that selecting just one infinite sequence of zeros and ones would not specify U completely. Far from it, because if the sequence is, say, $(0, 1, 0, 1, \ldots)$, so that $Q_0, Q_{01}, Q_{010}, Q_{0101}, \ldots$ all are in U, then each of these Q's can be split up into two, four, $\ldots$ arbitrary infinite subsets and at each split a choice has to be made as far as the membership of U is concerned.

The conclusion must be that free ultrafilters cannot be completely constructed and that this is due to the fact that the axiom of choice cannot be dispensed with. Nevertheless, as already indicated in the previous section, classical mathematics developed by means of nonstandard analysis can be kept free from this axiom, if desired.

Chapter 2

Basic theory

2.1 Reviewing the introduction of $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$

A)   *Integers.* Each integer is generated by a pair of natural numbers. If $\langle m, n \rangle$ is such a pair, the integer generated by it is indicated by $Z(m, n)$. $\mathbb{Z}$ is the set of all $Z(m, n)$.
Equality. $Z(m, n) = Z(p, q)$ if and only if $m \geq n$, $p \geq q$ and $m - n = p - q$, or $m < n$, $p < q$ and $n - m = q - p$.
Exercise: Show that this equality relation satisfies the rules of equivalence.
Identification. For all $m \in \mathbb{N}$, let $Z(m, 0) = m$.
The more usual notation instead of $Z(m, n)$ is, of course, say, q in case $m \geq n$ and then $q = m - n$, or $-q$ in case $m \leq n$ and then $q = n - m$.

B)   *Rationals.* Each rational is generated by a pair $\langle m, n \rangle$ of integers with $n \neq 0$. Let $Q(m, n)$ be the rational generated by (m, n). $\mathbb{Q}$ is the set of all $Q(m, n)$.
Equality. $Q(m, n) = Q(p, q)$ if and only if $mq = np$.
Exercise: Verify again the rules of equivalence.
Identification. For all $k \in \mathbb{Z}$, let $Q(k, 0) = k$.
Again the usual notation is different: $\mid m \mid / \mid n \mid$ or $- \mid m \mid / \mid n \mid$ instead of $Q(m, n)$, depending on whether $mn \geq 0$ or $mn \leq 0$, preferably such that m and n have no common divisors.

C)   *Reals.* Each real is generated by a *Cauchy sequence* of rationals, i.e. an infinite sequence $(r_1, r_2, \ldots)$ of rationals r_n, such that,

$$\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n, p \in \mathbb{N},\ n, p > k :\mid r_n - r_p \mid < 1/m.$$

Let the real generated by $(r_1, r_2, \ldots)$ be indicated by $R(r_1, r_2, \ldots)$. $\mathbb{R}$ is the set of all $R(r_1, r_2, \ldots)$.

Equality. $R(r_1, r_2, \ldots) = R(s_1, s_2, \ldots)$ if and only if $(r_1, r_2, \ldots)$ and $(s_1, s_2, \ldots)$ are concurrent Cauchy sequences of rationals, i.e. if

$$\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n \in \mathbb{N}, \, n > k :| r_n - s_n |< 1/m.$$

The verification of the rules of equivalence is now somewhat more involved.

Identification. For all $r \in \mathbb{Q}$, let $R(r, r, \ldots) = r$.

Once again the usual notation is different, although this time there is no simple notational rule as there was for integers and rationals. Examples are $\sqrt{2}$, π, e, etc., etc.

In all three cases, A) to C), operations such as $+$ and simple relations such as $\leq$ can easily be defined for the new numbers by means of the corresponding operations and simple relations that are known for the terms of the sequences generating them. In some, but certainly not all, cases the definitions can be given 'term by term', such as $+$ and $\leq$ for the reals:

$$R(r_1, r_2, \ldots) + R(s_1, s_2, \ldots) = R(r_1 + s_1, r_2 + s_2, \ldots),$$

and,

$$R(r_1, r_2, \ldots) \leq R(s_1, s_2, \ldots) \text{ if } r_1 \leq s_1, \, r_2 \leq s_2, \ldots$$

It would be wrong, however, to define $<$ and $>$ for reals in this way! A simple counterexample is,

$$?? \, 0 = R(0, 0, \ldots) < R(1, 1/2, 1/3, \ldots) = 0, \text{ or } 0 < 0 \, ??$$

Exercise: Present the right definitions. More generally, present the definitions for $+$, $-$, $*$, $/$, $| \cdot |$ for integers, rationals and reals, where appropriate.

The success of all this is, of course, that apart from the fact that any two numbers of some kind have a sum and a product of the same kind, in addition any two integers have a difference that is an integer, any two rationals (one of which is nonzero) have a difference and a quotient that are rationals, any two reals (one of which is nonzero) have a difference and a quotient that are reals, and any Cauchy sequence of reals has a limit that is a real (the proof of the last statement is not entirely trivial).

From their definitions it follows that any integer, any rational and any real is an urelement.

2.2 Introducing internal constants; definition of equality

The collection of all constants of nonstandard analysis can be determined in two steps, starting from the collection of all classical constants. First of all generate constants by means of infinite sequences of classical constants and add them to the collection if they are new, i.e. if they are not identified with a classical constant. The generation of these constants will be explained in detail below. Then add new constants (sets, pairs, triples, ..., functions) to the collection that can be defined by means of what is in the collection so far in the same way as this is done in classical mathematics (for example, add the set of all infinitesimals).

Every constant of nonstandard analysis is either *internal* or *external*. The internal constants should be regarded as the decent members of the society of nonstandard analysis, and the external constants as its outcasts, because the behaviour of the former is like that of similar classical constants, but that of the latter may be strange, unexpected and counterintuitive, which is why in internal set theory (see Section 1.9) by set automatically is meant internal set, and an external set is not a set at all. Here we do not go that far: any set is either internal or external, and in the latter case still is a set. An example of an external set is $\mathbb{N}$.

Every internal constant is either *standard* or not. Standard constants are characterized by a close relationship to classical constants, and some of them are even identified with the latter. Every internal constant that is not standard and every external constant is *nonstandard* (see the diagram in Section 1.6). The internal constants are introduced first, then follow the standard constants and the external constants.

Remark: The terms 'standard', 'internal' and 'external' have been taken from Nelson's internal set theory.

Recall that a free ultrafilter U over $\mathbb{N}$ has been 'selected' once for all, but that so far the definition of new constants did not involve U, i.e. so far only *classical constants* have been considered.

Now let each infinite sequence $(s_1, s_2, \ldots)$ of classical constants s_i generate an *internal constant* or *hyperconstant* $H(s_1, s_2, \ldots)$, which will often be abbreviated to $H(s_i)$. After the development of some theory, most of the time internal constants can be used without that a generating sequence of classical constants is needed. This is entirely analogous to the generation of the reals $R(r_1, r_2, \ldots)$ by means of Cauchy sequences of rationals r_i. Most of the time one can use, say, $\sqrt{2}$ without that some Cauchy sequence tending to $\sqrt{2}$ is needed.

If all s_i are numbers, also $H(s_i)$ will be *regarded* as a number (as it may not be a set in the present theory, it is not so easy to say *what* a number really is, all that can be done is tell who is a number). Similarly, if all S_i are sets, also $H(S_i)$ will be a set; if all p_i are pairs, also $H(p_i)$ will be a pair (and similarly for n-tuples), and if all f_i are functions, also $H(f_i)$ will be a function. Hence *hypersets* will be sets, *hyperpairs* will be pairs (and similarly for n-tuples), and *hyperfunctions* will be functions, in the classical sense of these terms, but *hypernumbers* will not necessarily be classical numbers.

The introduction of $H(s_1, s_2, \ldots)$ requires that the rules of identification and equality must be specified. The latter is the simpler of the two:

Definition of equality of internal constants:

$$H(s_1, s_2, \ldots) = H(t_1, t_2, \ldots) \text{ if and only if } \{i : s_i = t_i\} \in U.$$

This definition applies no matter whether the s_i and t_i are urelements or not. It implies that it is pointless to let the s_i be of two different kinds, say, numbers and pairs of numbers. For example, suppose that $s_{2i-1} \in \mathbb{R}$, and $s_{2i} = \langle p_i, q_i \rangle$, $p_i, q_i \in \mathbb{R}$. Then, since either $\{2i-1 : i \in \mathbb{N}\} \in U$ or $\{2i : i \in \mathbb{N}\} \in U$, we can just as well assume that *all* s_i are reals, or that *all* s_i are pairs of reals, respectively. Making the changes such that this is true will change the presentation, but not the value of $H(s_i)$. This follows immediately from the definition of equality given. The same argument can be applied if there were not two but three or more kinds of elements, as long as the number of kinds is finite. A case with an infinite number of kinds is where s_i is an i-tuple of reals. Another case is where s_i is a set of level i. In order to avoid difficulties such cases are to be avoided themselves. This is quite acceptable, because in practice they would seem to be rather fancy.

2.3   Identification of internal constants

Identification of internal numbers

A)   If all s_i are numbers, then let,

$$H(s_1, s_2, \ldots) = s \text{ if and only if } \{i : s_i = s\} \in U.$$

Hence $H(3, 3, 3, \ldots)$ is identified with 3, but $H(1, 2, 3, \ldots)$ is new as it is not identified with any classical number.

Identification of internal sets

The identification of internal sets is based on the following result.

Theorem 2.3.1 *Given infinite sequences $(S_1, S_2, \ldots)$ and $(T_1, T_2, \ldots)$ of sets S_i and T_i,*

$$\{H(s_i) : s_i \in S_i\} = \{H(t_i) : t_i \in T_i\} \text{ if and only if } \{i : S_i = T_i\} \in U.$$

Proof: Let $S = \{H(s_i) : s_i \in S_i\}$, $T = \{H(t_i) : t_i \in T_i\}$ and $Q = \{i : S_i = T_i\}$.

If $Q \in U$, take $H(s_i) \in S$ arbitrarily, then for all i, $s_i \in S_i$, and for $i \in Q$, $s_i \in T_i$, hence $H(s_i) \in T$, because of the definition of equality, which means that $S \subseteq T$ and similarly it follows that $T \subseteq S$, so that $S = T$.

Conversely, if $Q \notin U$ then $\{i : S_i \neq T_i\} \in U$, hence either

$$R = \{i : s_i \notin T_i \text{ for some } s_i \in S_i\} \in U \text{ or,}$$
$$\{i : t_i \notin S_i \text{ for some } t_i \in T_i\} \in U,$$

or both. Suppose $R \in U$ (the other case is similar). If $i \in R$ take $s_i \in S_i$ but such that $s_i \notin T_i$, and if $i \notin R$ take $s_i \in S_i$ arbitrary. Then $\{i : s_i \notin T_i\} \supseteq R$, so that $\{i : s_i \notin T_i\} \in U$ and $H(s_i) \notin T$, because otherwise

$$\emptyset = \{i : s_i \notin T_i\} \cap \{i : s_i \in T_i\} \in U,$$

a contradiction. Hence, indeed, $H(s_i) \notin T$ and $S \neq T$. $\qquad\square$

Comparing this result with the definition of equality, it follows that $H(S_i) = H(T_i)$ if and only if $S = T$, where S_i, T_i, S and T are as in the theorem and its proof. This suggests the following definition.

B)   If all S_i are sets, let,

$$H(S_1, S_2, \ldots) = \{H(s_i) : s_i \in S_i\},$$

which obviously is a (classical) set, although its elements need not be classical.

Proof of the rules of equivalence for equality for internal numbers and internal sets

(1) $H(s_i) = H(s_i)$ because $\{i : s_i = s_i\} = \mathbb{N} \in U$.

(2)   If $H(s_i) = H(t_i)$ then $H(t_i) = H(s_i)$, because $\{i : s_i = t_i\} = \{i : t_i = s_i\}$.

(3) $H(r_i) = H(s_i)$ and $H(s_i) = H(t_i)$, so that $P = \{i : r_i = s_i\} \in U$ and $Q = \{i : s_i = t_i\} \in U$, then $R = \{i : r_i = t_i\} \in U$, because $R \supseteq P \cap Q$, hence $H(r_i) = H(t_i)$.   □

The next result takes care of some simple cases.

Theorem 2.3.2 *If for all i, $S_i = \{s_i\}$, then,*

$$H(\{s_1\}, \{s_2\}, \ldots) = \{H(s_1, s_2, \ldots)\}, \ \ or \ H(\{s_i\}) = \{H(s_i)\}.$$

Also, if for all i, $S_i = \{s_i, s_i'\}$, then,

$$H(\{s_i, s_i'\}) = \{H(s_i), H(s_i')\},$$

and in general, if for some $n \in \mathbb{N}$ and all i, $S_i = \{s_{i1}, \ldots, s_{in}\}$, then,

$$H(\{s_{i1}, \ldots, s_{in}\}) = \{H(s_{i1}), \ldots, H(s_{in})\}.$$

Proof: Left as an exercise.   □

In shorthand, this result may be written as: $H(\{\ldots\}) = \{H(\ldots)\}$, i.e. the 'operator' H may be interchanged with set formation if all the terms S_i of the generating sequence are sets with n elements, $n \in \mathbb{N}$. If all S_i are infinite, the theorem is wrong. As a counterexample, let $S_i = \{s_{i1}, s_{i2}, \ldots\} = \{1, 2, \ldots\} = \mathbb{N}$ for all i. Then $H(S_i)$ contains $H(1, 2, 3, \ldots) = H(i)$, which is not contained in $\{H(s_{i1}), H(s_{i2}), \ldots\} = \{H(1), H(2), \ldots\} = \{1, 2, \ldots\} = \mathbb{N}$.

Exercise: What if $S_i = \{1, \ldots, i\}$?

*Identification of internal **n**-tuples*

Next consider $H(p_i)$ where all $p_i = \langle x_i, y_i \rangle$ are pairs. It would be nice if $H(\langle x_i, y_i \rangle)$ too would be a pair, but is it? Now recall from Section 1.14 that,

$$\langle x_i, y_i \rangle = T_2(\{\{x_i\}, \{x_i, y_i\}\}),$$

so that applying Theorem 2.3.2 twice it follows that,

$$H(T_2^{-1}(\langle x_i, y_i \rangle)) = H(\{\{x_i\}, \{x_i, y_i\}\}) = \{H(\{x_i\}), H(\{x_i, y_i\})\} =$$
$$\{\{H(x_i)\}, \{H(x_i), H(y_i)\}\} = T_2^{-1}(\langle H(x_i), H(y_i) \rangle).$$

Therefore,

$$T_2(H(T_2^{-1}(\langle x_i, y_i \rangle))) = \langle H(x_i), H(y_i) \rangle,$$

from which it is clear that if T_2 *were* replaced by the identity map, $H(\langle x_i, y_i \rangle)$ *would be* the pair $\langle H(x_i), H(y_i) \rangle$. Fortunately,

$$H(\langle x_i, y_i \rangle) = H(\langle x_i', y_i' \rangle) \text{ if and only if } \langle H(x_i), H(y_i) \rangle = \langle H(x_i'), H(y_i') \rangle,$$

because,

a) $H(\langle x_i, y_i \rangle) = H(\langle x_i', y_i' \rangle)$ if and only if $\{i : \langle x_i, y_i \rangle = \langle x_i', y_i' \rangle\} \in U$, and

b) $\langle H(x_i), H(y_i) \rangle = \langle H(x_i'), H(y_i') \rangle$ if and only if,
        $T_2(H(T_2^{-1}(\langle x_i, y_i \rangle))) = T_2(H(T_2^{-1}(\langle x_i', y_i' \rangle)))$, hence if and only if, $\{i : T_2^{-1}(\langle x_i, y_i \rangle) = T_2^{-1}(\langle x_i', y_i' \rangle)\} \in U$, hence if and only if, $\{i : \langle x_i, y_i \rangle = \langle x_i', y_i' \rangle\} \in U$.

For this reason, $H(\langle x_i, y_i \rangle)$ is identified with $\langle H(x_i), H(y_i) \rangle$. This means that the H operator may be interchanged with pairing, and that the *internal pair* or *hyperpair* $H(\langle x_i, y_i \rangle)$ is a (classical) pair, albeit with terms that need not be classical.

The same reasoning can be followed for *internal* or *hyper n-tuples*.

Identification of internal functions

Turning to functions, the next question is whether, given an infinite sequence $(f_1, f_2, \ldots)$ of functions f_i, $H(f_i)$ can be identified with a function, and if so which one? Let $f_i : X_i \to Y_i$, for certain sets X_i and Y_i. Then f_i is generated by the set $G_i = \{\langle x_i, f_i(x_i) \rangle : x_i \in X_i\}$, with $f_i(y_i) \in Y_i$. Also $f_i = F(G_i)$, see again Section 1.14. Hence,

$$H(F^{-1}(f_i)) = H(G_i) = H(\{\langle x_i, f_i(x_i) \rangle : x_i \in X_i\}).$$

Again interchanging H with set formation, this becomes,

$$\{H(\langle x_i, f_i(x_i) \rangle) : H(x_i) \in H(X_i)\},$$

and interchanging H with pairing,

$$\{\langle H(x_i), H(f_i(x_i)) \rangle : H(x_i) \in H(X_i)\},$$

which set clearly generates a function g from $H(X_i)$ to $H(Y_i)$, because $H(f_i(x_i)) \in H(Y_i)$, and it follows that $g = F(H(F^{-1}(f_i)))$. Fortunately, $H(f_i) = H(f_i')$ if and only if $g = g'$, with $g' = F(H(F(f_i')))$, because,

a) $H(f_i) = H(f_i')$ if and only if $\{i : f_i = f_i'\} \in U$, and

b) $g = g'$ if and only if
       $H(F^{-1}(f_i)) = H(F^{-1}(f_i'))$, hence if and only if,
       $\{i : F^{-1}(f_i) = F^{-1}(f_i')\} \in U$, hence if and only if,
       $\{i : f_i = f_i'\} \in U$.

Therefore, $H(f_i)$ is identified with g, so that the *internal function* or *hyperfunction* $H(f_i)$ is a classical function from the internal set $H(X_i)$ to the internal set $H(Y_i)$, and,

$$H(f_i)(H(x_i)) = H(f_i(x_i)).$$

Summarizing, we have:

C)   For n-tuples: let $H(\langle x_{i1}, \ldots, x_{in} \rangle) = \langle H(x_{i1}), \ldots, H(x_{in}) \rangle$.

D)   For functions: given $f_i : X_i \to Y_i$, let $H(f_i) : H(X_i) \to H(Y_i)$, with

$$H(f_i)(H(x_i)) = H(f_i(x_i)).$$

Proof of the rules of equivalence for equality in the case of n-tuples and functions: In the case of n-tuples, this follows from the identification for n-tuples and the fact that, if $n = 2$,

$$H(\langle x_i, y_i \rangle) = H(\langle x_i', y_i' \rangle) \text{ if and only if } \langle H(x_i), H(y_i) \rangle = \langle H(x_i'), H(y_i') \rangle,$$

and in the case of functions, this follows from the identification for pairs and functions and the fact that,

$$H(f_i) = H(f_i') \text{ if and only if } g = g'.$$

The details are left as an exercise.   □

If for each i the function g_i maps W_i to X_i and the function f_i maps X_i to Y_i, so that the composition $f_i \circ g_i$ is well-defined, then the corresponding *internal composition* $H(f_i \circ g_i)$ is the (classical) composition of $H(f_i)$ and $H(g_i)$, because,

$$H(f_i)(H(g_i)(H(w_i))) = H(f_i)(H(g_i(w_i)) = H(f_i(g_i(w_i))), \ w_i \in W_i.$$

Identification of internal sequences

Just as sequences are special functions, internal sequences are special internal functions. Suppose, given some $n \in \mathbb{N}$, that each f_i is a sequence $(f_i(1), \ldots, f_i(n))$ with n terms. (For more clarity, their arguments are not indicated by subscripts, but are between parentheses.) Then $H(f_j)$ too is a sequence with n terms, because

the domain of $H(f_i)$ is equal to the domain of each f_i, i.e. $\{1,\ldots,n\}$. Why? The j-th term of this sequence is,

$$H(f_i)(j) = H(f_i)(H(j)) = H(f_i(j)).$$

Hence in this case the *internal sequence* is a classical sequence with n internal terms. But if n depends on i this is not true in general. For if the i-th sequence is $(f_i(1),\ldots,f_i(n_i))$, then the argument or index of $H(f_i)$ takes the form $H(j_i) = H(j_1,j_2,\ldots)$ and hence is an internal natural number that need not be a classical natural number. Still it is meaningful to speak of the $H(j_i)$-th term of the internal sequence, and this term is equal to $H(f_i)(H(j_i)) = H(f_i(j_i))$, $j_i = 1,\ldots,n_i$.

Example: Let $n_i = i$ and $f_i(j_i) = i + j_i$, then the term with index $H(j_i) = H(j_1,j_2,\ldots)$, $j_i = 1,\ldots,i$, is $H(1 + j_1, 2 + j_2,\ldots)$. Taking all $j_i = 1$ this gives that $H(2,3,\ldots)$ is the first term. The n-th term is found by taking $j_i = n$ for all $i \geq n$, and the result is $H(?,\ldots,?,2n,2n+1,\ldots)$, where the $n-1$ question marks may be replaced by any numbers, as their values are irrelevant to the value of $H(?,\ldots,?,2n,2n+1,\ldots)$. The obvious choice leads to $H(n+1,\ldots,2n-1,2n,2n+1,\ldots)$. And the term corresponding to index $H(i) = H(1,2,3,\ldots)$ is $H(2i) = H(2,4,6,\ldots)$.

Suppose now that all f_i are infinite sequences. Then the internal sequence is not a classical sequence; its argument is $H(j_i)$ and the $H(j_i)$-th term is equal to $H(f_i)(H(j_i)) = H(f_i(j_i))$, $j_i = 1,2,\ldots$.

Example: Let $f_i(j_i) = i^2 + j_i^3$, then the term with index $H(j_i) = H(j_1,j_2,\ldots)$ is $H(1^2 + j_1^3, 2^2 + j_2^3,\ldots)$, so that the first term is $H(2,5,10,\ldots)$ and the second one is $H(9,12,17,\ldots)$. And the term corresponding to index $H(i) = H(1,2,3,\ldots)$ is $H(i^2 + i^3) = H(2,12,36,\ldots)$.

2.4 Standard constants; basic results for internal constants

Any constant that is equal to $H(s) = H(s,s,\ldots)$ for some classical constant s is called *standard*. This does not mean that it needs to be a constant of 'standard', that is to say classical, mathematics. It only means that it is closely related to some classical constant, although in a number of cases it is indeed equal to such a constant. Obviously, there is a bijection between the set of all classical constants and the set of all standard constants, whereby s is mapped to $H(s)$. In order to emphasize this special relationship, instead of $H(s) = H(s,s,\ldots)$ the

usual notation will be $^*\!s$, and $^*\!s$ is called the $*$-*transform* of s. So, in case $H(s)$ is identified with s, it follows that $^*\!s = s$, so that, for example, $^*5 = 5$, hence 5 is a standard constant, but a function $f : \mathbb{R} \to \mathbb{R}$ is not, because $^*\!f : {}^*\mathbb{R} \to \mathbb{R}^*$ and $^*\mathbb{R} \supset \mathbb{R}$. The strict inclusion here follows from the fact that, for example, $H(1, 2, 3, \ldots) \notin \mathbb{R}$ and $H(1, 1/2, 1/3, \ldots) \notin \mathbb{R}$, as can easily be shown by an indirect argument and using the identification rules. Nevertheless, if $x \in \mathbb{R}$, then $^*\!f(x) = f(x)$, so that $^*\!f$ is an extension of f.

Exercise: Show this.

Obviously, a constant is internal if it is an element of some internal set. Even the following is true.

Theorem 2.4.1 *A constant is internal if and only if it is an element of some standard set.*

Proof: The if-part is obvious. Conversely, given any internal constant $H(s_i) = H(s_1, s_2, \ldots)$, let $S = \{s : s = s_i \text{ for some } i\}$, then $H(s_i) \in {}^*\!S$. $\qquad \square$

Theorem 2.4.2 *Let S be a classical set such that $^*\!s = s$ for each $s \in S$. Then $^*\!S = S$ if and only if S is finite. Otherwise $S \subset {}^*\!S$.*

Proof: Since $^*\!s = H(s) = s$ if $s \in S$, $^*\!S = \{H(S) = H(s_i) : s_i \in S\} \supseteq S$. If S is finite then any $H(s_i) = s$ for some $s \in S$, hence $^*\!S \subseteq S$, and if S is infinite then there exist $s_1, s_2, \ldots$, all different, so that $H(s_i) \notin S$. $\qquad \square$

As an example let S be a set of numbers.

Corollary 2.4.1 *Let S be a classical set, such that each of its elements is a classical set of which each element is equal to its $*$-transform. Then $^*\!S = S$ if and only if S is finite and all its elements are finite. A similar result holds if S is of any level (see Section 1.13 for a definition).*

Proof: Left as an exercise. $\qquad \square$

In general, it may happen that S is not contained in $^*\!S$. Examples are where $S = \{\mathbb{R}\}$, because then $^*\!S = \{^*\mathbb{R}\}$, so that $\mathbb{R}$ is not an element of $^*\!S$, or where S is a set of functions from $\mathbb{R}$ to $\mathbb{R}$.

Now let f be some function from a set of numbers to another set of numbers. When is $^*\!f = f$? If this is true then the domains of both functions must be equal

and the theorem tells us when this is the case. Conversely, if both domains are equal and hence are finite, then if X is their common domain it follows that $^*f(x) = f(x)$ for all $x \in X$, so that if f is a function from X onto some set Y, then Y must be finite, $^*Y = Y$ and $^*f$ too is a function from X onto Y. This can be generalized as follows.

Corollary 2.4.2 *Let f be a function from X onto Y, and assume that X is finite, that $^*x = x$ for all $x \in X$ and $^*y = y$ for all $y \in Y$. Then $^*X = X$, $^*Y = Y$ and $^*f = f$.* □

Corollary 2.4.3 *Let g be a function from W to X, let f be a function from X onto Y, and assume that X is finite, that $^*x = x$ for all $x \in X$ and that $^*y = y$ for all $y \in Y$. Then, with w_i, $w \in W$,*

$$^*(f{\circ}g)(H(w_i)) = (f{\circ}^*g)(H(w_i)) \text{ and } ^*(f{\circ}g)(^*w) = (f{\circ}g)(w),$$

even if W is not finite.

Proof: It follows that $^*X = X$, $^*Y = Y$ and that $^*f = f$. Since $^*(f{\circ}g) = ^*f{\circ}^*g$, the first equality follows immediately, and since $^*g(^*w) =^* (g(w)) = g(w)$ also the second one follows quickly. □

A number of useful results has been summarized below.

a)   (Empty set) $^*\emptyset = \emptyset$.

b)   (Relations for sets) Given nonempty sets S_i, S, T_i and T,

        $H(S_i) = H(T_i)$ if and only if $\{i : S_i = T_i\} \in U$,
        $H(S_i) \neq H(T_i)$ if and only if $\{i : S_i \neq T_i\} \in U$,
        $H(s_i) \in H(S_i)$ if and only if $\{i : s_i \in S_i\} \in U$,
        $H(S_i) \subseteq H(T_i)$ if and only if $\{i : S_i \subseteq T_i\} \in U$;

    in particular,

        $^*S = {}^*T$ if and only if $S = T$,
        $^*S \neq {}^*T$ if and only if $S \neq T$,
        $^*s \in {}^*S$ if and only if $s \neq S$,
        $S \subseteq {}^*T$ and only if $S \subseteq T$,

    and similarly for $\subset$, $\supseteq$ and $\supset$.

c)   (Operations on sets) Given sets S_i, S, T_i and T,

        $H(S_i \cup T_i) = H(S_i) \cup H(T_i)$,
        $H(S_i \cap T_i) = H(S_i) \cap H(T_i)$,
        $H(S_i - T_i) = H(S_i) - H(T_i)$,

if $S_i \subseteq T_i$ then $H(S_i^c) = (H(S_i))^c$, where taking the complement is with respect to T_i and $H(T_i)$, respectively;

in particular,

$$*(S \cup T) = *S \cup *T,$$
$$*(S \cap T) = *S \cap *T,$$
$$*(S - T) = *S - *T,$$

if $S \subseteq T$ then $*(S^c) = (*S)^c$, where taking the complement is with respect to T and $*T$, respectively.

d)   (Pairs)

$$H(\langle s_i, t_i \rangle) = \langle H(s_i), H(t_i) \rangle;$$

in particular,

$$*\langle s, t \rangle = \langle *s, *t \rangle,$$

and similar equalities hold for n-tuples, $n = 3, 4, \ldots$.

e)   (Functions)

$$H(f_i(x_i)) = H(f_i)(H(x_i));$$

in particular,

(a) $*(f(x)) = *f(*x).$

f)   (Composite functions)

$$H((f_i \circ g_i)(w_i)) = (H(f_i) \circ H(g_i))(H(w_i));$$

in particular,

$$*[(f \circ g)(w)] = (*f \circ *g)(*w).$$

Most of these relationships are easily shown or even follow directly from definitions. As far as a) is concerned, apply the definition of $*S$ with $S = \emptyset$. Then it appears that no $H(s_i)$ can be found, so that $*S$ must empty, hence, $*\emptyset = \emptyset$. And the proof to show under b) that $*s \in *S$ if and only if $s \in S$, follows from $H(s_i) \in H(S_i)$ if and only if $\{i : s_i \in S_i\} \in U$, because this gives that $*s \in *S$ if and only if $\{i : s \in S\} \in U$ and this is true if and only if $s \in S$.   □

Remark: Obviously, not only, say, $H(f_i)(H(x_i))$ and $*f(*x)$, but also mixtures like $*f(H(x_i))$ and $H(f_i)(*x)$ are well-defined.

Remark: It is not true that if $t \in *S$, then $t = *s$ for some $s \in S$. As a counterexample, let $S = \mathbb{N}$, and let $t = H(1, 2, 3, \ldots)$.

Remark: It has been made clear that the inclusion $*S \supseteq S$ does not always hold, but it can be 'saved' by introducing $^\sigma S$, defined by

$$^\sigma S = \{*s : s \in S\},$$

because then trivially $^\sigma S \subseteq *S$. Moreover, there is a 'natural' bijection φ from S onto $^\sigma S$, defined by $\varphi(s) = *s$. $^\sigma S$ is sometimes called the *standard copy* of S, but often it is not a standard set.

Remark: In another presentation of nonstandard analysis, $^*S$ is called the *non-standard version* of S, but $^*S$ is still called standard, which leads to the confusing conclusion that the nonstandard version of a classical set is standard. In still another presentation of nonstandard analysis $^*S$ is called the *natural extension* of $^*S$, but as we have seen, S not always is an extension of $^*S$.

So even the wording of nonstandard analysis is sometimes nonstandard.

2.5   External constants

Having defined internal and standard constants, let us now consider external constants. As already said before, an *external* constant is a constant that is not internal.

Theorem 2.5.1 *If S is a classical infinite set of numbers, then S is external.*

Proof: First notice that if T is any set of classical numbers, then $^*T \supseteq T$, because if $t \in T$ then $^*t = H(t) = t$. This implies that if $T \subseteq S$, then $S \cap {}^*T = T$, for if $s \in S \cap {}^*T$ then $s = {}^*s \in {}^*T$, so that $s \in T$, hence $S \cap {}^*T \subseteq T$. Since also $T \subseteq S \cap {}^*T$, it follows that $S \cap {}^*T = T$.

Now if S would be internal, then,

$$S = H(S_i) = \{H(s_i) : s_i \in S_i\} \text{ for suitable sets } S_i.$$

Since S is infinite, it contains a countably infinite set,

$$T = \{t_n : n \in \mathbb{N}\}.$$

Let $T_i = S_i \cap T$, then,

$$H(T_i) = H(S_i \cap T) = H(S_i) \cap H(T) = S \cap {}^*T = T,$$

so that T too would be internal. Let $Q = \{i : T_i \text{ is infinite}\}$, then $Q \notin U$, as otherwise it is possible to select $t_i' \in T_i$, if $i \in Q$, all different, so that, taking t_i' arbitrarily if $i \notin Q$, $t' = H(t_i') \notin T$, because $\{i : t_i' = t'\}$ contains at most one element and hence is not in U. But at the same time $t' \in T$, because $H(T_i) = T$, a contradiction. Therefore,

$$Q^c = \{i : T_i \text{ is finite }\} \in U.$$

Let, for $i \in Q^c$, $t_{k(i)}$ be the element of T_i with largest index (if T_4 would be $\{t_6, t_7, t_{13}\}$, then $t_{k(4)}$ would be 13), and let $t_{k(i)}$ be arbitrary if $i \notin Q^c$, then

$$H(t_{k(1)}, t_{k(2)}, \ldots) \in T, \text{ so that } H(t_{k(1)}, t_{k(2)}, \ldots) = t_k$$

for some $k \in \mathbb{N}$, hence $\{i : t_{k(i)} = t_k\} \in U$, or

$$R = \{i : T_i \text{ contains at most } k \text{ elements}\} \in U,$$

because of the definition of $k(i)$, so that $t_{k+1} \notin T_i$ for all $i \in R$, hence,

$$t_{k+1} \notin H(T_i) = T,$$

whereas $t_{k+1} \in T$, a contradiction. $\square$

The theorem implies that, although we do not yet know much about $^*\mathbb{N}$, $^*\mathbb{R}$, etc., we can say already now that $\mathbb{N}$ is an external subset of $^*\mathbb{N}$, that $\mathbb{R}$ is an external subset of $^*\mathbb{R}$, etc.

Another consequence of the theorem is related to power sets. Let, given any classical set A, $\mathcal{P}(A)$ be the *power* set of A, defined by,

$$\mathcal{P}(A) = \{S : S \subseteq A\},$$

or, in words, $\mathcal{P}(A)$ is the set of all subsets of A (including A itself and $\emptyset$). Instead of $\mathcal{P}(A)$ also the notation 2^A is used.

Exercise: Show that if A is a set of numbers, then $^*(\mathcal{P}(A)) = \mathcal{P}(A)$ if A is finite.

Now let A be a set of numbers and let us compare $^*(\mathcal{P}(A))$ with $\mathcal{P}(^*A)$. By definition,

$$^*(\mathcal{P}(A)) = \{H(S_i) : S_i \in \mathcal{P}(A)\} = \{H(S_i) : S_i \subseteq A\},$$

and

$$\mathcal{P}(^*A) = \{T : T \subseteq {}^*A\}.$$

If T is any element of $^*(\mathcal{P}(A))$, then $T = H(S_i)$ for certain $S_i \subseteq A$, hence

$$T = \{H(s_i) : s_i \in S_i\} \subseteq {}^*A,$$

so that $T \in \mathcal{P}(^*A)$. It follows that $^*(\mathcal{P}(A)) \subseteq \mathcal{P}(^*A)$, i.e. applying the $*$-transform at a lower level (A's level) gives no less than at a higher level ($\mathcal{P}(A)$'s level). When is this inclusion strict? If A is finite, then $\mathcal{P}(A)$ is finite too, hence $^*(\mathcal{P}(A)) = \mathcal{P}(A) = \mathcal{P}(^*A)$, so that the inclusion is not strict. If A is infinite, then (by the last theorem) A is an external subset T of $^*A$, i.e. $\mathcal{P}(^*A)$ contains an element that is not internal, whereas any element of $^*(\mathcal{P}(A))$ is internal, and it follows that the inclusion is strict. In other words:

Theorem 2.5.2 *If A is a set of numbers, the inclusion $^*(\mathcal{P}(A)) \subseteq \mathcal{P}(^*A)$ always holds and is strict if and only if A is an infinite set, and if so, $\mathcal{P}(^*A)$ contains an element that is an external subset of A.* $\square$

Since in many cases A will be infinite (e.g. if $A = \mathbb{N}$ or $A = \mathbb{R}$), one should be careful when dealing with $\mathcal{P}(^*A)$ and prefer to work with $^*(\mathcal{P}(A))$ instead, which is standard, so that all its elements are internal. This is the deeper reason behind the rule advocated from the beginning of this book to formulate statements by means of $\in$, not by means of $\subset$ or $\subseteq$. When applying transfer later on it will almost be a must to stick to this rule.

2.6 The $*$-transform of operations and expressions

By *operations* are meant functions like taking the absolute value, addition, subtraction, multiplication, division, taking the complement, union, intersection, etc. And by *expressions* are meant what results when operations or other functions, or compositions thereof are applied to suitable constants or variables. The simplest expressions contain only one function, in particular one operation and no other functions, for example $\mid x \mid$, or $x + y$, or $S \cup T$; more complicated expressions are, for example, $\{f(x) + \mid y \mid\} \cdot z^2$, or $(S \cup T) \cap (V \cup W)$. (In certain computer languages expressions all of whose constants or variables are numbers are called arithmetic expressions.)

Since, when introducing the internal version $H(\&_1, \&_2, \ldots)$ of any classical operation $\&$, only one operation is given, i.e. $\&$ itself, all $\&_i$ must be taken equal to $\&$, so that only the standard operation $^*\&$ can be introduced. Since operations are functions, the definition of $^*\&$ is straightforward. For example, standard addition in $^*\mathbb{R}$ becomes $^*+$, defined by,

$$H(s_i)^* + H(t_i) = H(s_i + t_i),$$

given any $H(s_i)$ and $H(t_i)$ in $^*\mathbb{R}$. Since $^*+$ is the only addition for hyperreals, the asterisk in $^*+$ may be dropped, because from the context it will always be clear whether classical addition is meant or its $*$-transform $^*+$. It should be kept in mind, however, that often the domains of $+$ and $^*+$ are different, so that strictly speaking the two operations are different. So the definition becomes,

$$H(s_i) + H(t_i) = H(s_i + t_i).$$

Whereas, by definition, internal addition is always standard, an internal sum need not, of course, be standard.

Standard subtraction, multiplication, division, taking the absolute value etc. are defined in exactly the same way, except that the divisor of a division must be nonzero. This extra condition can be met in the following two ways. Let $H(t_i)$ be the divisor, then $H(t_i) \neq 0$ if and only if $\{i : t_i \neq 0\} \in U$. Now either each t_i that is zero is changed to an arbitrary nonzero number, say, 1, which has no effect on the value of $H(t_i)$, and the definition becomes,

$$H(s_i)/H(t_i) = H(s_i/t_i),$$

or the t_i that are zero are left unchanged, and the definition becomes,

$$H(s_i)/H(t_i) = H(r_i), \text{ with } r_i = s_i/t_i \text{ if } t_i \neq 0 \text{ and } r_i \text{ arbitrary otherwise.}$$

With regard to other operations similar refinements can be formulated, if necessary.

Exercise: Consider the composition of two functions.

Another group of operations is formed by the set operations of forming the union or the intersection or the difference of sets, or the complement of a set, hence $\cup$, $\cap$, $-$, and ^{c}. Again only their standard forms can be introduced. But whereas, say, addition has no meaning for the hyperreals, unless it is explicitly introduced for them, the classical versions of $\cup$, $\cap$, $-$, and ^{c} also have a meaning for internal sets, so that, for example $^*\cup$ must be distinguished from $\cup$, leading to,

$$H(S_i)^*\cup H(T_i) \text{ and } H(S_i) \cup H(T_i),$$

where, by definition,

$$H(S_i)^*\cup H(T_i) = H(S_i \cup T_i), \text{ for hypersets } H(S_i) \text{ and } H(T_i).$$

Fortunately, $^*\cup = \cup$, and the same holds for $\cap$, $-$, and ^{c}.

It would be wrong, however, to assume that any operation on sets would be equal to its $*$-transform. The notorious exception is power set formation. For, by definition,

$$(^*\mathcal{P})(H(S_i)) = H(\mathcal{P}(S_i)) = H(\{T_i : T_i \subseteq S_i\}) = \{H(T_i) : H(T_i) \subseteq H(S_i)\},$$

and,

$$\mathcal{P}(H(S_i)) = \{T : T \subseteq H(S_i)\},$$

but whereas $H(T_i)$ is internal, T might be external. So the equality $^*\& = \&$ is only guaranteed for $\& = \cup$, $\cap$, $-$, or ^{c}, and $\mathcal{P}$ should be avoided even for internal sets.

The internal forms of more complicated expressions such as $\{f(x)+ \mid y \mid\} \cdot z^2$ and $(S \cup T) \cap (V \cup W)$, become

$$\{H(f_i)(H(x_i))+ \mid H(y_i) \mid\} \cdot (H(z_i))^2 \text{ and } (H(S_i) \cup H(T_i)) \cap (H(V_i) \cup H(W_i)),$$

respectively. Taking everything standard gives,

$$\{^*f(^*x)+ \mid {}^*y \mid\} \cdot (^*z)^2 = {}^*[\{f(x)+ \mid y \mid\} \cdot (z)^2] \text{ and}$$
$$(^*S \cup {}^*T) \cap (^*V \cup {}^*W) = {}^*[(S \cup T) \cap (V \cup W)].$$

2.7 The $*$-transform of relations and statements; Łoš' theorem; the internal definition principle

The simplest mathematical relations are the *atomic relations*, by which are meant relations containing neither logical connectives nor quantifiers, hence relations such as $=$, $<$, $\in$, etc. They can be regarded as functions to the set $B = \{$TRUE, FALSE$\}$, where TRUE and FALSE are the Boolean constants, which are urelements. By definition, *TRUE $\equiv$ TRUE and *FALSE $\equiv$ FALSE, hence $^*B = B$. (In order to avoid confusion equivalence will be indicated by means of $\equiv$.) An atomic relation with n arguments is called n-ary. Many atomic relations are binary. *Atomic statements* result when atomic relations are applied to suitable arguments, which must be expressions. The smaller-than relation in $\mathbb{R}$ thus leads to the atomic statements $s < t$ (in which case R is binary), $s < t < u$ (now R is ternary), etc. with $s, t, u \in \mathbb{R}$. So if a relation is regarded as a function, the corresponding statement should be regarded as a function value. As with operations, only the standard form of internal relations is introduced, but whereas some of them are new in nonstandard analysis (e.g. $<$) others are not, namely when the classical form is also meaningful for internal constants (e.g. $=$, $\in$, $\subset$).

From the correspondence between relations and functions it follows, that if R is a binary relation such as $<$ or $\subset$, $^*R$ is defined by,

$$H(s_i) \, {}^*R \, H(t_i) \equiv H(s_i R t_i),$$

for suitable expressions s_i and t_i. Here by definition,

$$H(s_i R t_i) \equiv \{i : s_i R t_i\} \in U,$$

that is to say,

$$H(s_i R t_i) \equiv \text{TRUE if and only if } \{i : s_i R t_i \equiv \text{TRUE}\} \in U.$$

Hence $^*R$ too is a binary relation. In fact, this is a very special case of Łoš' theorem, to be considered below. In particular, letting $s_i = s$ and $t_i = t$ for all i,

$$^*[sRt] \equiv {}^*s\, {}^*R\, {}^*t \equiv sRt,$$

as $\{i : sRt\} \in U \equiv sRt$, since sRt does not depend on i. This means that $^*s\, {}^*R\, {}^*t$ is equivalent to the classical statement that is obtained by removing all asterisks, which is a very simple case of *transfer*. For n-ary atomic relations the definitions are analogous.

In case R is one of the relations $=$, $\neq$, $\in$, $\subset$, $\subseteq$, $\supset$, $\supseteq$, R has also a meaning for internal expressions. Fortunately, these relations are equivalent to their standard forms.

Proof: Consider $\in$. By definition,

$$H(s_i)\, {}^*\!\in H(S_i) \equiv H(s_i \in S_i) \equiv \{i : s_i \in S_i\} \in U \equiv H(s_i) \in H(S_i).$$

The other cases are left as exercises. $\qquad\square$

In case R has no meaning within nonstandard analysis, it will usually cause no confusion when the asterisk in $^*R$ is dropped. Below this is done anyway when explicit examples are given, such as the next two.

Examples:

1) $\{H(f_i)(H(x_i)) + |H(y_i)|\} \cdot (H(z_i))^2 < (H(s_i) + H(t_i)) \cdot H(r_i) \equiv$
   $H[\{f_i(x_i) + \mid y_i \mid\} \cdot z_i^2 < (s_i + t_i) \cdot r_i]$
   and this is also equivalent to,
   $\{i : \{f_i(x_i) + \mid y_i \mid\} \cdot z_i^2 < (s_i + t_i) \cdot r_i\} \in U.$

Taking everything standard the results are,

2) $\{^*f(^*x) + \mid {}^*y \mid\} \cdot (^*z)^2 < (^*s + {}^*t) \cdot {}^*r \equiv {}^*[\{f(x) + \mid y \mid\} \cdot z^2 < (s + t) \cdot r]$
   which is equivalent to,
   $\{f(x) + \mid y \mid\}] \cdot z^2 < (s + t) \cdot r,$
   because,
   $\{i : \{f(x) + \mid y \mid\}] \cdot z^2 < (s + t) \cdot r\} \in U$ if and only if $\{f(x) + \mid y \mid\}] \cdot z^2 < (s + t) \cdot r,$
   as the latter statement does not depend on i.

In the first example to the left there are several internal variables (the constant 2 is even standard), whereas to the right there is only a single internal statement; and in the second example to the left there are several standard variables, whereas to the right there is only a single standard statement, that, moreover, is

equivalent to the corresponding classical statement. As already indicated before this equivalence is an example of *transfer*, but a very simple one because neither logical connectives nor quantifiers occur. Note that in cases like,

$$\{{}^*f({}^*x) + \mid {}^*y \mid\} \cdot ({}^*z)^2 = {}^*[\{f(x) + \mid y \mid \cdot(z)^2] \text{ and}$$
$$({}^*S \cup {}^*T) \cap ({}^*V \cup {}^*W) = {}^*[(S \cup T) \cap (V \cup W)],$$

considered in the preceding section, it is not always allowed to drop the asterisk to the right, because if *s is some classical constant s may be different from s. So at this point there is a divergence between expressions and statements.

An *arbitrary statement* is composed of a finite number of atomic relations, logical connectives ($\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$), quantifiers ($\forall$ and $\exists$), constants, free variables and bound variables. First of all, statements with at least one logical connective, but without quantifiers will be considered. They will be written as,

$$R(P(s, s', s'', \ldots), Q(t, t', t'', \ldots), S(u, u', u'' \ldots), \ldots),$$

where P, Q, S, $\ldots$ are atomic relations, and s, s', s'', $\ldots$, t, t', t'', $\ldots$, u, u', u'', $\ldots$ are expressions of constants and free variables. In the beginning of this section it was found more convenient to write the atomic statement derived from an atomic binary relation R as sRs'. This is now written as $R(s, s')$. In order to simplify the notation $(s, s', s'', \ldots)$, $(t, t', t'', \ldots)$, $(u, u', u'', \ldots)$ will be abbreviated to (s), (t), (u), respectively. Regarding $R(P(s), Q(t), S(u), \ldots)$ as a function value, the corresponding (Boolean) function is, of course, R, which is called a *relation*. Its domain is $B \times B \times B \times \ldots$, and its range is B, where, as before, $B = \{\text{TRUE}, \text{FALSE}\}$.

As an example, let

$$R(P(s), Q(t), S(u)) \equiv (P(s) \wedge Q(t)) \Rightarrow \neg(S(u)),$$

which, emphasizing that the logical connectives are functions, can be written as,

$$((\Rightarrow \circ (\wedge, \neg)) \circ (P, Q, S)) (s, t, u),$$

where by definition $R \equiv (\Rightarrow \circ (\wedge, \neg))$, $(P, Q, S) (s, t, u) \equiv (P(s), Q(t), S(u))$ and

$$(\Rightarrow \circ (\wedge, \neg))(x_1, x_2, x_3) \equiv (x_1 \wedge x_2) \Rightarrow (\neg x_3).$$

*Remark:*In formal logic expressions, statements, and statements without free variables are often called *terms*, *formulae* (or *predicates*), and *sentences*, respectively.

As before, internal relations will only be standard, so that internal statements have the form,

$$H(R(P(s_i),\ Q(t_i),\ S(u_i),\dots))) \equiv {}^*R({}^*P(H(s_i)),\ {}^*Q(H(t_i)),\ {}^*S(H(u_i)),\dots).$$

Continuing the last example, this gives,

$$H(R(P(s_i),\ Q(t_i),\ S(u_i))) \equiv {}^*R({}^*P(H(s_i)),\ {}^*Q(H(t_i)),\ {}^*S(H(u_i))) \equiv$$
$$({}^*P(H(s_i))\ {}^*\wedge\ {}^*Q(H(t_i)))\ {}^* \Rightarrow\ {}^*\neg({}^*S(H(u_i))) \equiv$$
$$(({}^* \Rightarrow \circ({}^*\wedge, {}^*\neg)) \circ ({}^*P, {}^*Q, {}^*S))(H(s_i), H(t_i), H(u_i)).$$

But, whereas the domains of P, Q and S are rather arbitrary their ranges are B, as is that of R, so that the range of $g \equiv (P, Q, S)$ is $X = B \times B \times B$ and as ${}^*B = B$ and ${}^*(B \times B \times B) = B \times B \times B$, by Corollary 2.4.2 with $f \equiv (\Rightarrow \circ(\wedge, \neg))$ and $Y = B$, it follows that,

$$ {}^*R \equiv R \equiv {}^*((\Rightarrow \circ\wedge, \neg)) \equiv ({}^* \Rightarrow \circ({}^*\wedge, {}^*\neg)) \equiv (\Rightarrow \circ(\wedge, \neg)).$$

Hence,

$$H(R(P(s_i),\ Q(t_i),\ S(u_i))) \equiv R({}^*P(H(s_i)),\ {}^*Q(H(t_i)),\ {}^*S(H(u_i))),$$

where the asterisks in ${}^*P$, ${}^*Q$, and ${}^*S$ could be dropped, as it is clear that standard forms are meant. This is a less trivial case of Łoš' theorem.

Taking $s_i = s$, $t_i = t$, $u_i = u$ for all i, Corollary 2.4.3 gives that,

$$ {}^*[R(P(s), Q(t), S(u))] \equiv H(R(P(s), Q(t), S(u))) \equiv R(P(s), Q(t), S(u))$$

that is,

$$ {}^*[R(P(s), Q(t), S(u))] \equiv R(P(s), Q(t), S(u)),$$

hence the standard form of the sample statement is equivalent to its classical form, which expresses a less trivial case of *transfer*.

Exactly the same kind of argument can be repeated for any other statement without quantifiers.

Finally, let arbitrary statements with at least one quantifier be given. They are assumed to be in prenex normal form (see Section 1.3), hence with all logical connectives to the right of the quantifiers. Also it is assumed that each bound variable occurs to the left of the $\in$ relation. From this it follows that in any internal statement each bound variable automatically is internal, even if it is not

explicitly written as such. To see this observe that if to the right of this $\in$ relation there is a constant or a free variable, then the bound variable must be internal, because that constant or that free variable is assumed to be internal, and if there is a bound variable, then the latter must be internal, as can be seen by repeating the argument. As an example see the least upper bound theorem below.

First let just one quantifier be included:

$$\exists x \in X : R(P(x, s), \ldots),$$

or,

$$\forall x \in X : R(P(x, s), \ldots),$$

where X is some set and $R(P(x, s), \ldots)$ is an arbitrary statement without quantifiers. Then, given sets X_i,

$$H(\exists x_i \in X_i : R(P(x_i, s_i), \ldots)) \equiv \exists H(x_i) \in H(X_i) : R(^*P(H(x_i), H(s_i)), \ldots),$$

and similarly for $\forall$. Note that to the left x_i and to the right $H(x_i)$ may be replaced by x. To prove this equivalence observe that the statement to the left is equivalent to,

1) $Q = \{i : [\exists x_i \in X_i : R(P(x_i, s_i), \ldots)]\} \in U$,
       and that the statement to the right is equivalent to,
       $R(^*P(H(x_i'), H(s_i)), \ldots)$,
       for certain x_i', hence, as no quantifiers are involved, to,
       $H(R(P(x_i', s_i), \ldots))$,
       hence to,
2) $\{i : R(P(x_i', s_i), \ldots)\} \in U$.

Now if statement 1) is true, for $i \in Q$ take some $x_i' \in X_i$ such that $R(P(x_i, s_i), \ldots)$, and if $i \notin Q$ take x_i' arbitrary. Then,

$$\{i : [\exists x_i \in X_i : R(P(x_i, s_i), \ldots)]\} \subseteq \{i : R(P(x_i', s_i), \ldots)\},$$

and statement 2) follows. Conversely, if statement 2) is true, then, since,

$$\{i : [\exists x_i \in X_i : R(P(x_i, s_i), \ldots)]\} \supseteq \{i : R(P(x_i', s_i), \ldots)\},$$

statement 1) follows. This completes the proof of the equivalence. $\qquad\square$

A similar result for $\forall$ can be shown as follows.

$$\neg[H(\forall x_i \in X_i : R(P(x_i, s_i), \ldots))] \equiv H(\exists x_i \in X_i : \neg[R(P(x_i, s_i), \ldots))] \equiv$$
$$\exists H(x_i) \in H(X_i) : \neg[R(^*P(H(x_i), H(s_i)), \ldots)] \equiv$$
$$\neg[\forall H(x_i) \in H(X_i) : R(^*P(H(x_i), H(s_i)), \ldots)].$$

□

Secondly, let two quantifiers be involved, as in,

$$\exists x \in X : \forall y \in Y : R(P(x, y, s), \ldots),$$

then a similar result can be proved by using that,

$$H(\forall y_i \in Y_i : R(P(x_i, y_i, s_i), \ldots)) \equiv$$
$$\forall H(y_i) \in H(Y_i) : R(^*P(H(x_i), H(y_i), H(s_i)), \ldots),$$

since here only one quantifier is involved.

Applying induction, it follows that a similar result can be shown for a statement containing arbitrarily many quantifiers. Again the asterisk in $^*P$ may be dropped if no confusion can arise. Hence the following result – which is quite fundamental – has been proved. In its formulation R is no longer regarded as a function of substatements $P(s)$, $Q(t)$, $S(u), \ldots$, and of sets X, X', $X'', \ldots$, required in the quantifications, but simply as a function of constants and free variables X, X', $X'', \ldots, s, s', s'', \ldots$, so that $P, Q, S, \ldots$ have altogether disappeared.

Theorem 2.7.1 *(Łoś' theorem.)*
Let any classical statement,

$$R(X, X', X'', \ldots; s, s', s'', \ldots),$$

with a finite number of constants or free variables X, X', $X'', \ldots, s, s', s'', \ldots$, and a finite number of logical connectives and quantifiers be given. Certain standard constants such as 0 need not be mentioned explicitly. X, X', $X'', \ldots$ are the sets required to formulate the quantifications properly; that is to say that X must occur in $\exists x \in X$ or in $\forall x \in X$, for some suitable bound variable x, and similarly for X', $X'', \ldots$, and that conversely each quantification is taken care of this way. Then,

$$H[R(X_i, X'_i, X''_i, \ldots; s_i, s'_i, s''_i, \ldots)] \equiv$$
$$R(H(X_i), H(X'_i), H(X''_i), \ldots; H(s_i), H(s'_i), H(s''_i), \ldots).$$

□

In other words, (regarding for the time being each constant as a free variable) given any classical statement, to each of its free variables q (so that here $q \in \{X, X', X'', \ldots, s, s', s'', \ldots\}$) add the index i, which defines infinite sequences (q_i), and which also defines an infinite sequence of classical statements, which sequence in turn defines an internal statement (the one to the left). Then the latter is equivalent to the statement (the one to the right) that results from the given classical statement by replacing each free variable q by the internal free variable $H(q_i)$ that is defined by the infinite sequence (q_i). It may happen, of course, that for certain q this leads to $^*q$, namely when $q_i = q$ for all i. Or, in still other words, given any classical statement, replace each of its free variables q by its internal version $H(q_i)$ (that might be its standard version $^*q$), then the resulting statement (the one to the right) is equivalent to the statement (the one to the left) that results from it by removing all H's and all asterisks, and putting one single H in front.

The formulation of the theorem implies that each bound variable must occur in some set inclusion. A more careless formulation would be allowed if it would be required that each bound variable is internal (which is automatically true in the formulation given).

So far, bound variables were only combined with quantifications, but they can occur in other formulations as well. A very usual one is the definition of a set, such as, $\{x \in X : P(x, X, s)\}$, where $P(x, X, s)$ is some statement. Then obviously x is a bound variable that is not combined with $\exists$ or $\forall$, although one says 'the set of all x in X such that $\ldots$' Anyway, the result is a set, hence a constant or a variable, not a statement. Yet, the theorem can be applied here and leads to the following corollary.

Corollary 2.7.1 *(The internal definition principle.)*
Let any statement be given, say $P(x, X, s)$, where X is some set and $x \in X$ makes sense. Let X and s be internal, then so is the set

$$T = \{x \in X : P(x, X, s)\}.$$

Proof: Since X is internal, also x is internal, so for suitable X_i and s_i,

$$T = \{H(x_i) \in H(X_i) : P(H(x_i), H(X_i), H(s_i))\}.$$

Let,

$$T_i = \{x_i \in X_i : P(x_i, X_i, s_i)\}.$$

Then $H(x_i) \in T$ if and only if $H(x_i) \in H(X_i) \wedge P(H(x_i), H(X_i), H(s_i))$, hence by Łoš' theorem, if and only if $H(x_i \in X_i \wedge P(x_i, X_i, s_i))$, hence if and only if $H(x_i \in T_i)$, hence if and only if $H(x_i) \in H(T_i)$, so that

$$H(x_i) \in T \equiv H(x_i) \in H(T_i),$$

i.e. $T = H(T_i)$, hence T is internal. $\qquad\square$

2.8 Transfer; the standard definition principle

In this section a few consequences of Łoš' theorem will be presented that belong to the main tools of nonstandard analysis.

Theorem 2.8.1 *(Transfer, first formulation.)*
Let $R(X, X', X'', \ldots; s, s', s'', \ldots)$ be as before. Then,

$$R(X, X', X'', \ldots; s, s', s'', \ldots) \equiv$$
$$R(^*X, ^*X', ^*X'', \ldots; ^*s, ^*s', ^*s'', \ldots).$$

Proof: In Łoš' theorem take $X_i = X$, $X_i' = X'$, $X_i'' = X'', \ldots$, $s_i = s$, $s_i' = s'$, $s_i'' = s'', \ldots$, for all i, then,

$$^*[R(X, X', X'', \ldots; s, s', s'', \ldots)] \equiv$$
$$R(^*X, ^*X', ^*X'', \ldots; ^*s, ^*s', ^*s'', \ldots),$$

but,

$$^*[R(X, X', X'', \ldots; s, s', s'', \ldots)] \equiv$$
$$R(X, X', X'', \ldots; s, s', s'', \ldots).$$

$\qquad\square$

One fact is disguised in this formulation of transfer, namely that in the statement to the right the bound variables need not be standard. A simple example may clarify this:

$$\exists x \in X : P(x, s) \equiv$$
$$\exists H(x_i) \in {}^*X : P(H(x_i), {}^*s) \equiv \exists x \in {}^*X : P(x, {}^*s),$$

where $P(x, s)$ is some substatement. So transfer expresses the fact that any classical statement is equivalent to the nonstandard statement that results from it by replacing everything by its $*$-transform except the bound variables. What will happen if the bound variables too are replaced by their $*$-transforms? In the example this gives,

$$\exists^* x \in {}^*X : P({}^*x, {}^*s),$$

but since ${}^*x \in {}^*X \equiv x \in X$ and $P({}^*x, {}^*s) \equiv P(x, s)$, it follows that this is equivalent to $\exists x \in X : P(x, s)$, and a similar equivalence holds for any classical statement. This leads to transfer in another formulation.

Theorem 2.8.2 *(Transfer, second formulation.)*
Given any internal statement, replacing everything, including every bound variable, by its standard version is equivalent to replacing everything except every bound variable by its standard version. $\qquad\square$

The advantage of transfer in its first formulation is that it is the bridge between classical mathematics and nonstandard mathematics; that of transfer in its second formulation that one remains within nonstandard mathematics, so that one may forget about classical mathematics, that via the $*$-transform is mapped in a one-to-one kind of way to a certain part of nonstandard mathematics. In the second formulation it comes close to what it is in Nelson's internal set theory, that strictly speaking ignores classical mathematics completely, and that within nonstandard mathematics defines the difference between standard and internal (and external, which in essence is an irrelevant notion in this theory, however).

Let us now see what really is the essential part of transfer. First of all it is clear that it is only of some value if bound variables are present. Simple examples of transfer in its second formulation are,

$$\exists^* x \in {}^*X : P({}^*x, {}^*s) \equiv \exists x \in {}^*X : P(x, {}^*s),$$

and

$$\forall^* x \in {}^*X : P({}^*x, {}^*s) \equiv \forall x \in {}^*X : P(x, {}^*s),$$

where $H(x_i)$ has been replaced by x. Keep in mind that $x \in {}^*X$ implies that x is not necessarily standard, but internal. Clearly, in one direction these two equivalences are trivial (at least if $X \subseteq {}^*X$), and what is really important are the following two implications,

$$\exists x \in {}^*X : P(x, {}^*s) \Rightarrow \exists x \in {}^*X : P({}^*x, {}^*s),$$

and

$$\forall^* x \in {}^* X : P(^* x, {}^* s) \Rightarrow \forall x \in {}^* X : P(x, {}^* s),$$

or, in words, if there exists an internal x such that something is true, then there even exists a standard $^*x$ such that that something is true; and if something is true for all standard $^*x$, then that something is even true for all internal x. The second implication leads from classical mathematics towards nonstandard mathematics, and the first one leads in the opposite direction back from nonstandard mathematics to classical mathematics.

Obviously, in case $^*X = X$, also the two implications are trivial, and transfer is of no use. Examples where this is not the case have already been given in Section 1.4. In that section it also became clear what really is the purpose of transfer: in a number of important cases the nonstandard form of a classical statement can be given a much simpler form (see Section 1.4, where statements (1.1) and (1.1) are equivalent to statement (1.1)). Inevitably, this simpler form requires the use of certain internal constants that are not standard (nonzero infinitesimals in the example of Section 1.4).

Example: As a nontrivial example let us consider the least upper bound theorem for $\mathbb{R}$, which in its classical form reads,

$$\forall X \in \mathcal{P}(\mathbb{R}) : \{X \neq \emptyset \wedge [\exists b \in \mathbb{R} : \forall x \in X : x \leq b]\} \Rightarrow$$
$$\exists \beta \in \mathbb{R} : [\forall x \in X : x \leq \beta] \wedge [\forall \varepsilon \in \mathbb{R}, \, \varepsilon > 0 : \exists x \in X : x > \beta - \varepsilon].$$

By transfer in its first formulation this is equivalent to,

$$\forall X \in^* (\mathcal{P}(\mathbb{R})) :$$
$$\{X \neq \emptyset \wedge [\exists b \in {}^*\mathbb{R} : \forall x \in X : x \leq b]\} \Rightarrow$$
$$\exists \beta \in {}^*\mathbb{R} : [\forall x \in X : x \leq \beta] \wedge [\forall \varepsilon \in {}^*\mathbb{R}, \, \varepsilon > {}^*0 : \exists x \in X : x > \beta - \varepsilon],$$

which by transfer in its second formulation is equivalent to,

$$\forall^* X \in^* (\mathcal{P}(\mathbb{R})) :$$
$$\{^*X \neq \emptyset \wedge [\exists^* b \in {}^*\mathbb{R} : \forall^* x \in {}^*X : {}^*x \leq {}^*b]\} \Rightarrow$$
$$\exists^* \beta \in {}^*\mathbb{R} : [\forall^* x \in {}^*X : {}^*x \leq {}^*\beta] \wedge [\forall^* \varepsilon \in {}^*\mathbb{R}, \, {}^*\varepsilon > 0 : \exists^* x \in {}^*X : {}^*x > {}^*\beta - {}^*\varepsilon],$$

which is equivalent to the classical version of the theorem. Note that in this example different bound variables have been indicated by the same symbol, which is against the rule advocated before in Section 1.3, but this time seems appropriate. Also note that internal bound variables have not explicitly be indicated a such, simply because they are automatically internal.

Exercise: Apply Łoš' theorem to the least upper bound theorem, but with internal free variables.

Just as transfer is a direct specialization of Łoš' theorem, the next result is a direct specialization of the internal definition principle.

Theorem 2.8.3 *(The standard definition principle.)*
Let $^*X$ be a standard set, $x \in {}^*X$ make sense, $^*s$ be standard, and $P(x, {}^*X, {}^*s)$ be any statement. Then also,

$$\{x \in {}^*X : P(x, {}^*X, {}^*s)\}$$

is standard.

Proof: The proof is a simplification of the proof of the internal definition principle, and is left as an exercise. □

2.9 The ∗-transform of attributes

So far the ∗-transform was concerned with expressions and statements. In this section the ∗-transforms of a number of attributes, such as finiteness are considered.

A) *finite or hyperfinite sets.
A set $S = H(S_i)$ is called **finite* or *hyperfinite* if all S_i are finite sets, or equivalently, if $\{i : S_i$ is finite$\} \in U$. This does not mean that S is a finite set. As a counterexample, let $S_i = \{1, 2, \ldots, i\}$, then the smallest element of S is $m = 1 = H(1, 1, 1, \ldots)$, the largest is $M = H(1, 2, 3, \ldots) = H(i)$, and those in between are $H(s_i)$ with $1 \le s_i \le i$. It follows that S is an infinite set, for if not then all of them would be at most $n = H(n, n, n, \ldots)$ for some $n \in N$. In particular, $M \le n$, which is not true. Nevertheless, a hyperfinite set can be treated as if it were finite. For example, just as finite sets of classical numbers, hyperfinite sets of internal numbers have a smallest as well as a largest element, hence are bounded.

Exercise: Show this.

Another example is where $\omega \sim \infty$ and f is an internal function from $^*\mathbb{N}$ to $\{1, 2\}$, such that $f(1) = 1$, $f(\omega) = 2$. Then there exists a largest $k \in {}^*\mathbb{N}$, $1 \le k \le \omega$, such that $f(k) = 1$, hence such that $f(j) = 2$ if $k + 1 \le j \le \omega$. The hyperfinite

set here is $\{j : 1 \leq j \leq H(n_i)\} = H(S_i)$ with $\omega = H(n_i)$ and $S_i = \{1, \ldots, n_i\}$. Such a result is trivial for finite sets, and by transfer carries over to the present case. For let F be the set of all classical $f : \mathbb{N} \to \{1, 2\}$. Then that trivial result reads, more formally,

$$\forall n \in \mathbb{N} : \forall f \in F, f(1) = 1, \, f(n) = 2 : \exists k \in \mathbb{N}, \, 1 \leq k \leq n :$$
$$[f(k) = 1 \wedge \forall j \in \mathbb{N}, \, k + 1 \leq j \leq n : f(j) = 2],$$

hence, by transfer,

$$\forall n \in {}^*\mathbb{N} : \forall f \in {}^*F, \, f(1) = 1, f(n) = 2 : \exists k \in {}^*\mathbb{N}, \, 1 \leq k \leq n :$$
$$[f(k) = 1 \wedge \forall j \in {}^*\mathbb{N}, \, k + 1 \leq j \leq n : f(j) = 2],$$

in particular if $n = \omega \sim \infty$.

B) *finite or hyperfinite numbers.
A number $x = H(x_i)$ is called *finite or hyperfinite if all x_i are finite. But there is no other choice, as all classical numbers are finite. In other words each internal number is hyperfinite. Yet an internal number might be larger than any natural number; just consider $H(i)$, or $H(i - 2)$. Nevertheless they can be treated as classical numbers: $H(i) - 1$, $(H(i))^2$, $H(i) - H(i - 2)$, etc. make sense.

C) *real or hyperreal.
A number $H(x_i)$ is called *real or hyperreal if all x_i are real, that is to say if $H(x_i) \in {}^*\mathbb{R}$. Obviously, such a number need not be real, but it can be treated as a classical real number.

D) *continuity or hypercontinuity.
A function $H(f_i)$ from ${}^*\mathbb{R}$ to ${}^*\mathbb{R}$ is called *continuous or hypercontinuous at $H(c_i) \in {}^*\mathbb{R}$ if for all i, f_i is continuous at c_i.

E) *countable or hypercountable.
A set $H(S_i)$ is called *countable or hypercountable if all S_i are countable. Hence ${}^*\mathbb{N}$ is hypercountable, although it is not countable, as will be shown below.

It should now be clear what is the purpose of this section, and that there are many variations of the theme.

Exercise: Define hyperfinite and hyperinfinite sequences.

Remark: In Nelson's internal set theory the prefix 'hyper' is not used: there 'finite' means hyperfinite, 'countable' means hypercountable, etc. On the other hand 'standard finite' means finite.

2.10 *$\mathbb{N}$, *$\mathbb{Z}$, *$\mathbb{Q}$, *$\mathbb{R}$: main definitions and properties

The sets *$\mathbb{N}$, *$\mathbb{Z}$, *$\mathbb{Q}$, and *$\mathbb{R}$ have already been used in various examples, but will now be treated in a more orderly fashion. Their definitions, the rules of equality and identification, and the definitions of the arithmetic operations and the inequalities are all obvious from the foregoing theory:

$$
\begin{aligned}
{}^*\mathbb{N} \; &= \; H(n_i) : n_i \in \mathbb{N}, \text{ and similarly for } {}^*\mathbb{Z}, {}^*\mathbb{Q} \text{ and } {}^*\mathbb{R},\\
H(x_i) \; &= \; H(y_i) \text{ if and only if } \{i : x_i = y_i\} \in U,\\
H(x_i) \; &= \; x \text{ if and only if } \{i : x_i = x\} \in U, \text{ so that } H(x) = {}^*x = x,\\
H(x_i) + H(y_i) \; &= \; H(x_i + y_i), x_i, y_i \in \mathbb{N},\\
H(x_i) \cdot H(y_i) \; &= \; H(x_i . y_i), x_i, y_i \in \mathbb{N},\\
\mid H(x_i) \mid \; &= \; H(\mid x_i \mid), x_i \in \mathbb{Z},\\
H(x_i) - H(y_i) \; &= \; H(x_i - y_i), x_i, y_i \in \mathbb{Z},\\
1/H(x_i) \; &= \; H(1/x_i), \text{ if all } x_i \neq 0 \text{ and } x_i \in \mathbb{Q},\\
H(x_i) < H(y_i) \; &\equiv \; H(x_i < y_i), \; x_i, y_i \in \mathbb{N}, \text{ and similarly for } >, \leq, \geq .
\end{aligned}
$$

Trivially, here $\mathbb{N}$ may be replaced by $\mathbb{Z}$, $\mathbb{Z}$ by $\mathbb{Q}$, and $\mathbb{Q}$ by $\mathbb{R}$. As far as inversion is concerned, it is of course sufficient that $\{i : x_i \neq 0\} \in U$, as then the x_i that are 0 can be changed to, say, 1, without that the value of $H(x_i)$ is changed. Note that $x < y$ is a statement, so $x < y \in B$, where $B = \{\text{TRUE}, \text{FALSE}\}$, hence $H(x < y^*) \in {}^*B = B$.

Definitions: Let x be a hypernumber.
x is called *positive hyperlarge* if $x > m$ for all $m \in \mathbb{N}$. Notation: $x \sim \infty$.
x is called *negative hyperlarge* if $-x > m$ for all $m \in \mathbb{N}$. Notation: $x \sim -\infty$.
Instead of hyperlarge the term *infinitely large* may be used, but this does not mean that x would be equal to ∞, which is not regarded a number at all.
x is called *finite* or *limited* if it is nor hyperlarge.
x is called *hypersmall*, or is called an *infinitesimal* if $\mid x \mid < 1/m$ for all $m \in \mathbb{N}$. Notation: $x \simeq 0$, or in case $x \neq 0$, $x \sim 0$.
x is called *appreciable* if x is limited but not hypersmall.

Theorem 2.10.1 *Infinitely large numbers and nonzero infinitesimals exist!*

Proof: For example,

$$
H(1, 2, 3, \ldots) \sim \infty, \; H(-1, -2, -3, \ldots) \sim -\infty,
$$
$$
H(1, 1/2, 1/3, \ldots) \sim 0 \, , H(-1, -1/2, -1/3, \ldots) \sim 0.
$$

□

Clearly $x = H(x_i) = H(+1, -1/2, +1/3, -1/4, \ldots) \sim 0$, but is it positive or negative? This depends: if $\{i : x_i > 0\} \in U$ then $x > 0$ and otherwise $x < 0$. Many dichotomies of this kind exist in nonstandard analysis, but they will not cause any trouble, because it will not take very long before generating sequences and the H-operator will disappear from the scene. Then references to the basic free ultrafilter U will no longer be required (except when it is necessary to go back to basic principles).

Theorem 2.10.2 $\varepsilon \sim 0$ *if and only if* $1/\varepsilon \sim +\infty$ *or* $1/\varepsilon \sim -\infty$, *hence s is appreciable if and only if* $1/s$ *is appreciable.*

Let $\varepsilon \sim 0$, $\varepsilon' \sim 0$, s and s' be appreciable, and $\omega \sim \infty$, $\omega' \sim \infty$. Then,

$\varepsilon + \varepsilon' \simeq 0$, $\varepsilon - \varepsilon' \simeq 0$, $\varepsilon \cdot \varepsilon' \simeq 0$,

$\varepsilon + s$ and $\varepsilon - s$ are appreciable, and $\varepsilon \cdot s \sim 0$,

$\varepsilon + \omega \sim \infty$, $\varepsilon - \omega \sim -\infty$, and $\varepsilon \cdot \omega \sim +\infty$ or $-\infty$ or 0, or $\varepsilon \cdot \omega$ is appreciable,

$s + s'$ and $s - s'$ are appreciable or $\simeq 0$, and $s \cdot s'$ is appreciable,

$s + \omega \sim +\infty$, $s - \omega \sim -\infty$, $s \cdot \omega \sim +\infty$ or $-\infty$,

$\omega + \omega' \sim +\infty$, $\omega - \omega' \sim +\infty$ or $-\infty$ or 0, or $\omega - \omega'$ is appreciable, and

$\omega \cdot \omega' \sim +\infty$.

Exercise: Show this, and provide examples for all possibilities in case there are more than one. □

The results of this section so far show that at long last Leibniz' theory of hypersmall and hyperlarge numbers can be given a sound mathematical basis. It was Robinson who in 1961 for the first time formulated a complete theory of nonstandard analysis. See Sections 1.8 and 1.9 for more details.

In Section 2.6 it was shown that $\mathbb{N}$ is an external subset of $^*\mathbb{N}$. An alternative proof can be given by showing that $\mathbb{N}$ has a property it would not have were it internal. This happens to be a proof technique that can be applied to many external notions. Some even define external sets as sets that have such a property, but they fail to tell what that property is given that set, so that this definition would not seem to be very practical. In the case of $\mathbb{N}$ the property is that a bounded internal subset S of $^*\mathbb{N}$ has a maximum. The proof of this statement is not difficult: let $S = H(S_i)$ be bounded by $b = H(b_i)$, then S_i is bounded by

b_i, at least for all $i \in \{i : S_i$ is bounded by $b_i\}$ which is an element of the free ultrafilter U, but as usual we may assume that this is true for all i (why?). Hence each S_i has a maximum m_i and $H(m_i)$ is the maximum of S. Now $\mathbb{N}$ is bounded in $^*\mathbb{N}$ by any hyperlarge natural number, and if it were internal it would have a maximum, which it has not. Therefore, $\mathbb{N}$ is external. $\qquad\square$

In a similar way it can be shown that the set of all infinitesimals is external. Since this set is bounded in $^*\mathbb{R}$, say by 1, it must have a least upper bound β if it were internal. But this would imply that β itself would be an infinitesimal, so that 2β would be an infinitesimal as well, but $2\beta > \beta$.

Exercise: Show that β would indeed be an infinitesimal itself.

Theorem 2.10.3 *In $^*\mathbb{R}$ the set of all infinitesimals is external.* $\qquad\square$

Another variation of the theme is the next result.

Theorem 2.10.4 *In $^*\mathbb{R}$ the set of all positive hyperlarge numbers is external.*

Proof: Left as an exercise. Hint: use lower bounds. $\qquad\square$

2.11 Overflow and underflow

This section is concerned with the existence in certain internal sets of an element that depending on the internal set given either is infinitely large, or is limited, or is an infinitesimal, or is not an infinitesimal.

Theorem 2.11.1 *(Overflow or overspill.)*
Let S be an internal subset of $^*T$, where T is either $\mathbb{N}$ or $\mathbb{Z}$, or $\mathbb{Q}$, or $\mathbb{R}$, such that $\forall m \in \mathbb{N} : \exists s(m) \in S : s(m) \geq m$, i.e. such that from a classical point of view S contains arbitrarily large elements, then $\exists s \in S : s \sim \infty$, i.e. S contains some infinitely large element.

Proof: If $\forall b \in {}^*T : \exists s(b) \in S : s(b) > b$, then take $b \sim \infty$, which implies that $s(b) \sim \infty$. If this is not true, then $\exists b \in {}^*T : \forall s \in S : s \leq b$, so that by Łoš' theorem, $H(\exists b_i \in T : \forall s_i \in S_i : s_i \leq b_i)$, where $S = H(S_i)$ and $b = H(b_i)$, which by the classical least upper bound theorem implies that,

$$H(\exists \beta_i \in T : [\forall s_i \in S_i : s_i \leq \beta_i] \wedge [\exists s_i' \in S_i : s_i' > \beta_i - 1]),$$

hence, again by Łoš' theorem,

$$\exists \beta \in {}^*T : [\forall s \in S : s \leq \beta] \wedge [\exists s' \in S : s' > \beta - 1],$$

and it follows that $\forall m \in \mathbb{N} : \beta \geq s(m) \geq m$, hence that $\beta \sim \infty$, so that $s' \sim \infty$. $\square$

Theorem 2.11.2 *(Underflow or underspill.)*
Let S be an internal subset of ${}^*T$, with T as before, such that $\forall \omega \in {}^*\mathbb{N}, \omega \sim \infty : \exists s(\omega) \in S : s(\omega) < \omega \wedge s(\omega) \sim \infty$, i.e. such that S contains infinitely large elements that are arbitrarily small, then $\exists s \in S : s$ is limited.

Proof: Let $S_1 = \{s \in S : s \geq 1\}$. Clearly, S_1 is not empty, so that by the classical greatest lower bound theorem,

$$\exists \beta \in {}^*\mathbb{N} : [\forall s \in S_1 : s \geq \beta] \wedge [\exists s' \in S_1 : s' < \beta + 1],$$

as can be shown by an argument similar to that used in the preceding proof. It follows that β is limited, as otherwise $\beta \leq s(\beta) < \beta$, so that s' is limited as well. $\square$

Since in ${}^*\mathbb{Q}$ and ${}^*\mathbb{R}$, x is infinitely large if and only if $1/x$ is a nonzero infinitesimal, these two theorems have the following counterparts.

Theorem 2.11.3 *('Inverse' overflow.)*
Let S be an internal subset of ${}^*\mathbb{Q}$ or ${}^*\mathbb{R}$, such that $\forall m \in \mathbb{N} : \exists s(m) \in S : |s(m)| \leq 1/m$, i.e. such that from a classical point of view S contains arbitrarily small elements, then $\exists s \in S : s \simeq 0$.

Proof: It is no restriction to assume that $0 \notin S$. Apply overflow to $S' = \{t : 1/t \in S\}$ and use the fact that S' is internal if (and only if) S is internal. $\square$

Theorem 2.11.4 *('Inverse' underflow.)*
Let S be an internal subset of ${}^*\mathbb{Q}$ or ${}^*\mathbb{R}$, such that $\forall \varepsilon, \varepsilon \sim 0, \varepsilon > 0 : \exists s \in S : s \geq \varepsilon$, then $\exists s \in S : s$ is not an infinitesimal.

Proof: Similar to the preceding proof. $\square$

The overflow theorem immediately implies that S is an external subset of ${}^*S$ in case S is equal to $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ or $\mathbb{R}$, a fact we knew already. The underflow theorem immediately implies that ${}^*\mathbb{N} \backslash \mathbb{N}$ too is an external subset of ${}^*\mathbb{N}$, and similarly

for the other sets of numbers. This also follows from the fact that a subset of a standard set $^*S$ is internal if and only if its complement with respect to $^*S$ is internal. The 'inverse' underflow theorem implies that the subset of $^*\mathbb{R}$ of all hypersmall elements of $^*\mathbb{R}$ is external, and similarly for $^*\mathbb{Q}$.

2.12 $^*\mathbb{N}$ and $^*\mathbb{Z}$: more properties

Theorem 2.12.1 *If* $n \in {}^*\mathbb{Q}$, *then either* $n \in \mathbb{N}$ *or* n *is hyperlarge.*

Proof: If $n = H(n_i)$ is not hyperlarge, then $n \leq m$ for some $m \in \mathbb{N}$, so that, $\{i : 0 < n_i \leq m\} \in U$, but then $\{i : n_i = m'\} \in U$ for precisely one $m' \in \mathbb{N}$, $m' \leq m$, as follows from the properties of U, so that $n = m'$. $\square$

Theorem 2.12.2 *Given any* $\omega \in {}^*\mathbb{N}$, $\omega \sim \infty$, *then* $S = [1, \omega]$ *is uncountable.*

Proof: The proof is given by constructing a bijection between S and the set of all infinite sequences of 0's and 1's, which is known to be uncountable.

Given $\omega_1 = \omega$, the interval S is split into intervals $S_0 = (0, \omega_0]$ and $S_1 = (\omega_0, \omega_1]$ of approximately equal length (see the details below), S_0 is in a similar way split into $S_{00} = (0, \omega_{00}]$ and $S_{01} = (\omega_{00}, \omega_{01}]$, and S_1 into $S_{10} = (\omega_{01}, \omega_{10}]$ and $S_{11} = (\omega_{10}, \omega_{11}]$, etc., where at each split the limits of all subintervals involved are given by a column of:

$$
\begin{aligned}
\omega_1 = \omega_{11} = \omega_{111} &= \ldots \\
\omega_{110} &= \ldots \\
\omega_{10} = \omega_{101} &= \ldots \\
\omega_{100} &= \ldots \\
\omega_0 = \omega_{01} = \omega_{011} &= \ldots \\
\omega_{010} &= \ldots \\
\omega_{00} = \omega_{001} &= \ldots \\
\omega_{000} &= \ldots \\
0 = 0 = 0 &= \ldots ,
\end{aligned}
$$

where $\omega_{0\ldots00} = \lfloor (\omega_{0\ldots01} + 0)/2 \rfloor$, and, given that b is any string of 0's and 1's with an even positive binary value and that c is any such string with an odd binary

value, $\omega_b = \lfloor(\omega_{b-1} + \omega_{b+1})/2\rfloor$, $\omega_c = \omega_{(c-1)/2}$, so that, for example,

$$\omega_{0111} = \omega_{011} = \omega_{01} = \omega_0 = \lfloor(\omega_1 + 0)/2\rfloor, \qquad \omega_{0011} = \omega_{001} = \omega_{00} = \lfloor(\omega_{01} + 0)/2\rfloor,$$
$$\omega_{1011} = \omega_{101} = \omega_{10} = \lfloor(\omega_{01} + \omega_{11})/2\rfloor, \qquad \omega_{1111} = \omega_{111} = \omega_{11} = \omega_1 = \omega,$$
$$\omega_{0001} = \omega_{000} = \lfloor(\omega_{001} + 0)/2\rfloor, \qquad \omega_{0101} = \omega_{010} = \lfloor(\omega_{001} + \omega_{011})/2\rfloor,$$
$$\omega_{1001} = \omega_{100} = \lfloor(\omega_{011} + \omega_{101})/2\rfloor, \qquad \omega_{1101} = \omega_{110} = \lfloor(\omega_{101} + \omega_{111})/2\rfloor.$$

Each of these intervals contains a hyperinfinite number of elements. For example, if $\omega = H(1, 2, 3, 4, 5, 6, 7, 8, \ldots)$, then

$$
\begin{aligned}
S &= H((0,1], (0,2], (0,3], (0,4], (0,5], (0,6], (0,7], (0,8], \ldots), \\
S_0 &= H(\emptyset, (0,1], (0,1], (0,2], (0,2], (0,3], (0,3], (0,4], \ldots), \\
S_1 &= H((0,1], (1,2], (1,3], (2,4], (2,5], (3,6], (3,7], (4,8], \ldots), \\
S_{00} &= H(\emptyset, \emptyset, \emptyset, (0,1], (0,1], (0,1], (0,1], (0,2], \ldots), \\
S_{01} &= H(\emptyset, (0,1], (0,1], (1,2], (1,2], (1,3], (1,3], (2,4], \ldots), \\
S_{10} &= H(\emptyset, \emptyset, (1,2], (2,3], (2,3], (3,4], (3,5], (4,6], \ldots), \\
S_{11} &= H((0,1], (1,2], (2,3], (3,4], (3,5], (4,6], (5,7], (6,8], \ldots), \text{ etc.}
\end{aligned}
$$

Let $\omega = H(n_i)$. Given any $x = H(x_i) \in S$, it may be assumed that $0 < x_i \le n_i$ for all i. Either $x \in S_0$ or $x \in S_1$, if $x \in S_0$, then either $x \in S_{00}$ or $x \in S_{01}$, and if $x \in S_1$, then either $x \in S_{10}$ or $x \in S_{11}$, etc., which defines a unique infinite sequence of 0's and 1's. If, for example, $x \in S_0$, $x \in S_{01}$, $x \in S_{011}, \ldots$, then the sequence is $(0, 1, 1, \ldots)$.

Conversely, any infinite sequence of 0's and 1's defines a sequence of intervals for some $x = H(x_i)$, inducing for each i a sequence of intervals for x_i. Continuing the example, let $i = 7$, then, if the infinite sequence of 0's and 1's is $(0, 1, 1, \ldots)$, the sequence of intervals for x_7 is,

$$(0, 3], (1, 3], (2, 3], \ldots,$$

because $(0,3]$ is the 7th term of the sequence generating S_0, $(1,3]$ that of the sequence generating S_{01}, $(2,3]$ that of the sequence generating S_{011}, etc., and subsequent terms are either $(2,3]$ or $\emptyset$, and an $\emptyset$ is followed by $\emptyset$'s only. Now no matter how large is n_i, the interval sequence for x_i will certainly contain a term of length 1, as, the length of $(a, b]$ being $b - a$, at each split an interval is split into subintervals of proximately equal length. In fact it will take no more than $\lceil^2\log n_i\rceil$ splits to find an interval of length 1. For each i with $n_i \ge 2$ let x_i be the upper limit of any term of the interval sequence for x_i whose length is 1 (all

those terms have the same upper limit), and for each i with $n_i = 1$ let $x_i = 1$. Then $x \in S$. In the example, $x_7 = 3$. $\qquad\square$

According to the construction in this proof if as before $\omega = H(i)$, then,

$$\begin{aligned}
(1,1,1,\ldots) &\quad \text{leads to } x = \omega, \\
(0,0,0,\ldots) &\quad \text{to } x = 1, \\
(0,1,1,\ldots) &\quad \text{to } x = H(1,1,1,2,2,3,3,\ldots), \\
(1,0,0,\ldots) &\quad \text{to } x = H(1,2,2,3,3,4,4,\ldots), \text{ etc.}
\end{aligned}$$

Note that the last two x's differ by $1 = H(0,1,1,\ldots)$, and that the last sequence of 0's and 1's can be found from the one but last by 'adding 1 at infinity'.

From the construction it also follows that if the generating sequence (n_i) for ω is nondecreasing (as it is in the example given), then the generating sequence (x_i) for x too is nondecreasing. Now let x be any hyperlarge element of $^*\mathbb{N}$, then there exists an $\omega = H(n_i) \geq x$ such that (n_i) is nondecreasing, for simply let $n_i = \max\{x_j : j \leq i\}$. Starting from this ω and computing the x_i from the infinite sequence of 0's and 1's that corresponds to x, it follows that (x_i) is nondecreasing as well. This shows the following corollary, that probably does not have much practical value.

Corollary 2.12.1 *If $x \in {}^*\mathbb{N}$ and $x \sim \infty$, then there exists a nondecreasing infinite sequence (x_i) tending to infinity such that $x = H(x_i)$.* $\qquad\square$

Example: Let $x = H(y_i)$, where $y_i = 1$ if $i = 2j + 1$ for some j, $y_i = 2$ if $i = 4j + 2$ for some j, $y_i = 3$ for $i = 8j + 4$ for some j, etc. Hence for each n the number of $y_i = n$ is infinitely large, and (y_i) is certainly not nondecreasing, and this sequence has no limit. Assume that $x \sim \infty$, which is possible, as e.g. $\{1, 2, 4, 8, 16, \ldots\}$ could be an element of the filter U. Note that $y_1 = 1$, $y_2 = 2$, $y_4 = 4$, etc. Then,

$$\omega = H(1,2,2,3,3,3,3,4,4,4,4,4,4,4,4,8,\ldots).$$

To x there corresponds an infinite sequence of 0'1 and 1's. Which one is difficult to tell, because this entirely depends on the underlying filter U. To this sequence there corresponds the desired nondecreasing sequence (x_i), and again it is difficult to tell which one. Even though the desired (x_i) cannot be determined constructively, the following result has been shown.

Corollary 2.12.2 *$^*\mathbb{N}$ is already generated by all nondecreasing infinite sequences that tend to infinity.* $\qquad\square$

Subtracting well chosen $H(x_i)$ and $H(y_i)$ from each other will destroy this nice property, however, see Section 1.10.

Another corollary of the theorem is, that since $\mathbb{N}$ is countable, it follows that if $\omega \sim \infty$, then $[1, \omega] - \mathbb{N}$ is uncountable, and that $*\mathbb{N}$ and $*\mathbb{N} - \mathbb{N}$ are uncountable as well. A direct proof of this is as follows.

Proof: The proof uses a variation of Cantor's diagonal method. If $*\mathbb{N}$ were countable, let $s(n) = H(s_i(n)) = H(s_1(n), s_2(n), \ldots)$ be its n-th element, $n \in \mathbb{N}$. Let,

$$t_1 = 1 + s_1(1), \ t_2 = 1 + \max(s_2(1), s_2(2)), \ t_3 = 1 + \max(s_3(1), s_3(2)_3, s_3(3)),$$

etc. Then $t = H(t_1, t_2, t_3, \ldots)$ is larger than each $s(n)$, including itself, as $t = s(j)$ for some j, contradiction. $\square$

As regards externality, recall from Section 2.6 that $\mathbb{N}$ is an external set, and so is $*\mathbb{N} - \mathbb{N}$, as these two sets are each others complement with respect to $*\mathbb{N}$, see Section 2.5. Seemingly friendly functions from $*\mathbb{N}$ to $*\mathbb{N}$ turn out to be external as well. For example, let f be a function from $*\mathbb{N}$ to $*\mathbb{N}$ such that $f(n) = 1$ if $n \in \mathbb{N}$ and $f(n) \neq 1$ if $n \notin \mathbb{N}$. Then f is external. For if f were internal, then by the internal definition principle the set $T = \{x \in *\mathbb{N} : f(x) = 1\}$ would be internal, but $T = \mathbb{N}$. After all, f is not so friendly, as the external $\mathbb{N}$ is involved in its definition.

Exercise: Show that f is external if $f(n) = 1$ if $n \sim \infty$, and $f(n) \neq 1$ if not.

Exercise: Let f be a function from $*\mathbb{N}$ to $\{1, 2\}$ such that $f(n) = 1$ if and only if $n = 1$, 2, or 3. Show that f is a standard function.

Exercise: Let f be an internal function from $*\mathbb{N}$ to itself, such that $f(n) = 1$ if $n \in \mathbb{N}$. Show that $f(n) = 1$ for all $n \in *\mathbb{N}$ and that f is standard.

Theorem 2.12.3 *Let f be an internal function from $*\mathbb{N}$ to $\{1, 2\}$, such that both 1 and 2 are assumed somewhere, i.e. such that f is onto. Moreover, let $f(n) = 2$ for all $n \geq b$ for some $b \in *\mathbb{N}$. Then there is a $\beta < b$, $\beta \in *\mathbb{N}$ such that $f(\beta) = 1$ and $f(\beta) = 2$ if $n > \beta$. Hence there is a last such that $f(\beta) = 1$, even though $b \sim \infty$ is allowed, and $\{n : n \leq b\}$ is uncountable in that case.*

Proof: The set $\{x \in *\mathbb{N} : f(x) = 1\}$ is internal and is bounded above by b, hence by the least upper bound theorem in its internal form it has a least upper bound β, which must be a maximum. $\square$

The results of this section have obvious counterparts for $*\mathbb{Z}$.

2.13 *$\mathbb{Q}$ and *$\mathbb{R}$: more properties; standard part

Theorem 2.13.1 *(Standard part theorem.)*
*If $x \in$ *$\mathbb{R}$ then either $\mid x \mid \sim \infty$, or $x = r + \varepsilon$, $r \in \mathbb{R}$, $\varepsilon \simeq 0$ for unique r and ε.*

Proof: Let x not be hyperlarge. Then the uniqueness of r and ε is easily shown, for let $r + \varepsilon = r' + \varepsilon'$, $r, r' \in \mathbb{R}$, $\varepsilon, \varepsilon' \simeq 0$, then $r - r' \simeq 0$, but $r - r' \in \mathbb{R}$, hence $r - r' = 0$, so that $r = r'$ and $\varepsilon = \varepsilon'$. Since x is limited, it follows that for some $b \in \mathbb{N}$, $\mid x \mid \leq b$. Hence $S = \{s : s \in \mathbb{R}, s \leq x\}$ is a nonempty subset of $\mathbb{R}$ that is bounded above by b. Indeed, S is nonempty because $-b \in S$. By the least upper bound theorem S has a least upper bound $\beta \in \mathbb{R}$. If $\beta < x - 1/m$ for some $m \in \mathbb{N}$, then β would not be an upper bound of S. If $\beta > x + 1/m$ for some $m \in \mathbb{N}$, then would not be the *least* upper bound of S. Therefore, for all $m \in \mathbb{N}$, $\mid \beta - x \mid \leq 1/m$, i.e. $\beta - x \simeq 0$, so that $r = \beta$ is the desired real. $\qquad\square$

Definition: If x is a limited hyperreal, then the (unique) standard hyperreal r that is infinitely close to x is called the *standard part* of x, which is denoted by $\mathrm{st}(x)$.

The theorem is false if *$\mathbb{R}$ is replaced by *$\mathbb{Q}$.

Counterexample: Let (x_i) be a Cauchy sequence of rationals, such that $R(x_i) = \sqrt{2}$. Then $x = H(x_i) \in$ *$\mathbb{Q}$ and x is limited. Suppose $x = r + \varepsilon$, $r \in \mathbb{Q}$, $\varepsilon \simeq 0$. Since also $x \in$ *$\mathbb{R}$, it follows from the theorem that $x = \sqrt{2}$ and $r = \sqrt{2}$ would be the only possibility, a contradiction. $\qquad\square$

It follows that not every limited hyperrational has a standard part in $\mathbb{Q}$. Yet limited hyperrationals do have a standard part, in $\mathbb{R}$ and in fact the entire $\mathbb{R}$ can thus be obtained.

Theorem 2.13.2 *Given any $r \in \mathbb{R}$, there exists an $x \in$ *$\mathbb{Q}$ such that $\mathrm{st}(x) = r$.*

Proof: If $r \in \mathbb{R}$, then there exists a sequence (r_n), $r_n \in \mathbb{Q}$, such that r_n tends to r if n tends to infinity, hence for all $m \in \mathbb{N}$, $\mid r_n - r \mid < 1/m$ if n is large enough. This implies that,

$$\forall m \in \mathbb{N} : \{i :\mid r_i - r \mid 1/m\} \in U,$$

hence that $x = H(r_i) \simeq r$. $\qquad\square$

Hence in some very special way *$\mathbb{Q}$ completely defines $\mathbb{R}$.

Theorem 2.13.3 *Let $a, b \in \mathbb{R}$, $a < b$. If $x \in {}^*[a, b]$, then $\mathrm{st}(x) \in [a, b]$, but if the interval in R is not closed, this is sometimes not true.*

Proof: Since x is limited, $\mathrm{st}(x)$ is well defined. If $\mathrm{st}(x) \notin [a, b]$, then either $\mathrm{st}(x) = a - \delta$ or $\mathrm{st}(x) = b + \delta$ for some $\delta \in \mathbb{R}$, $\delta > 0$. But $x - \mathrm{st}(x) \simeq 0$, hence either $x < a - \delta/2$ or $x > b + \delta/2$, but then x is not in ${}^*[a, b]$.

If, for example, $x \in {}^* (a, b]$, let a_i in $\mathbb{R}$ converge to a such that $a < a_i \leq b$, then $x = H(a_i) \in {}^* (a, b]$ but $\mathrm{st}(x) = a \notin (a, b]$. $\square$

2.14 An alternative to introducing $^*\mathbb{Z}$, $^*\mathbb{Q}$ and $^*\mathbb{R}$

In Sections 2.1 and 2.10 the following scheme for the introduction of $^*\mathbb{N}$, $^*\mathbb{Z}$, $^*\mathbb{Q}$ and $^*\mathbb{R}$ was used,

$$
\begin{array}{ccccccc}
\mathbb{N} & \rightarrow & \mathbb{Z} & \rightarrow & \mathbb{Q} & \rightarrow & \mathbb{R} \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
{}^*\mathbb{N} & & {}^*\mathbb{Z} & & {}^*\mathbb{Q} & & {}^*\mathbb{R}
\end{array}
$$

but there is an alternative, namely,

$$
\begin{array}{ccccccc}
\mathbb{N} & & \mathbb{Z} & & \mathbb{Q} & & \mathbb{R} \\
\downarrow & & \uparrow & & \uparrow & & \uparrow \\
{}^*\mathbb{N} & \rightarrow & {}^*\mathbb{Z} & \rightarrow & {}^*\mathbb{Q} & \rightarrow & {}^*\mathbb{R}
\end{array}
$$

Extending $^*\mathbb{N}$ to $^*\mathbb{Z}$ directly. Consider pairs $\langle m, n \rangle$ of elements m and n of $^*\mathbb{N}$, and let these pairs generate constants $Z'(m, n)$, subject to exactly the same identification and equality rules as were given in Section 2.1 for $Z(m, n)$, $m, n \in \mathbb{N}$. Let $\mathbb{Z}' = \{Z'(m, n) : m, n \in {}^*\mathbb{N}\}$. In $\mathbb{Z}'$ (non)negativity, (non)positivity, absolute value, addition, subtraction, multiplication and the inequalities are defined in exactly the same way as they were defined for $\mathbb{Z}$.

A 'natural' bijection between $\mathbb{Z}'$ and $^*\mathbb{Z}$ can now be established as follows. Given any $z' \in \mathbb{Z}'$, there are infinite sequences (m_i), and (n_i), $m_i, n_i \in \mathbb{N}$, such that

$$z' = Z'(H(m_i), H(n_i)).$$

Each pair $\langle m_i, n_i \rangle$ defines the element $Z(m_i, n_i) \in \mathbb{Z}$, hence the two sequences define the element $z = H(Z(m_i, n_i)) \in {}^*\mathbb{Z}$. So each $z' \in \mathbb{Z}'$ defines a $z \in {}^*\mathbb{Z}$.

Conversely, each $z \in {}^*\mathbb{Z}$ defines a $z' \in \mathbb{Z}'$. It is not difficult to see that the bijection that is implied by this preserves (non)negativity, (non)positivity, absolute value, addition, subtraction, multiplication and the inequalities, so that $\mathbb{Z}'$ may be identified with ${}^*\mathbb{Z}$ by identifying corresponding z' and z.

Extending ${}^*\mathbb{Z}$ to ${}^*\mathbb{Q}$ directly. This time consider pairs $\langle m, n \rangle$ of elements m and n of ${}^*\mathbb{Z}$, and let these pairs generate constants $Q'(m, n)$, subject to exactly the same identification and equality rules as were given in Section 2.1 for $Q(m, n)$, $m, n \in \mathbb{Z}$. Let $\mathbb{Q}' = \{Q'(m, n) : m, n \in {}^*\mathbb{Z}\}$. For the elements of $\mathbb{Q}'$ everything is defined in exactly the same way as it was done for the elements of $\mathbb{Q}$. Now a bijection between $\mathbb{Q}'$ and ${}^*\mathbb{Q}$ can be established that preserves everything, so that $\mathbb{Q}'$ may be identified with ${}^*\mathbb{Q}$.

In the preceding section we have shown that given any $r \in \mathbb{R}$ there exists an $x \in {}^*\mathbb{Q}$ such that $\mathrm{st}(x) = r$. An alternative proof is now as follows. Given any $r \in \mathbb{R}$, $\forall n \in \mathbb{Z} : \exists m \in \mathbb{Z} : m \leq nr < m + 1$, hence, by transfer,

$$\forall n \in {}^*\mathbb{Z} : \exists m \in {}^*\mathbb{Z} : m \leq n \cdot {}^*r < m + 1,$$

but ${}^*r = r$, hence $m/n \leq r < m/n + 1/n$. Take $n \sim \infty$, then, as $1/n \sim 0$, $\mathrm{st}(m/n) = r$, where $x = m/n \in \mathbb{Q}'$, as $m, n \in {}^*\mathbb{Z}$.

Extending ${}^*\mathbb{Q}$ to ${}^*\mathbb{R}$ directly. This direct extension is more involved than the preceding two because in order to generate $\mathbb{R}$ from $\mathbb{Q}$ infinite sequences (i.e. Cauchy sequences) of rationals are required, whereas for the preceding two extensions only pairs of natural numbers or integers were required, for recall that the internal version of a pair is still a pair, but that the internal version of an infinite sequence is not an infinite sequence. What is needed are internal Cauchy sequences with terms in ${}^*\mathbb{Q}$. Now what is an internal Cauchy sequence in the first place? A classical sequence $(r(n))$ of rationals $r(n)$ is a Cauchy sequence if,

$$\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n, p \in \mathbb{N}, n, p > k :\mid r(n) - r(p) \mid < 1/m.$$

Consequently, an internal Cauchy sequence $(r(n))$, $n \in {}^*\mathbb{N}$, $r(n) \in {}^*\mathbb{Q}$, is characterized by classical Cauchy sequences $(r_i(n))$, $n \in \mathbb{N}$, $r_i(n) \in \mathbb{Q}$, $i = 1, 2, 3, \ldots$, such that $r(H(n_i)) = H(r_i(n_i))$, hence by the hyperstatement,

$$H[\forall m_i \in \mathbb{N} : \exists k_i \in \mathbb{N} : \forall n_i, p_i \in \mathbb{N}, n_i, p_i > k_i :\mid r_i(n_i) - r_i(p_i) \mid < 1/m_i],$$

which can be simplified to,

$$H[\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n, p \in \mathbb{N}, n, p > k :\mid r_i(n) - r_i(p) \mid < 1/m].$$

In an analogous way, internal concurrency between the internal Cauchy sequences $r(n)$ and $s(n)$ is characterized by the hyperstatement,

$$H[\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n \in \mathbb{N}, n > k :\mid r_i(n) - s_i(n) \mid < 1/m].$$

Now let each internal Cauchy sequence $r(n)$ of hyperrationals generate a constant $R'(r(n))$ and let,

$$\mathbb{R}' = \{R'(r(n)) : n \in {}^*\mathbb{N}, r(n) \in {}^*\mathbb{Q}, \text{ and } (r(n)) \text{ is an internal Cauchy sequence.}\}$$

Equality. $R'(r(n)) = R'(s(n))$ if and only if $r(n)$ and $s(n)$ are internally concurrent.

Identification. $R'(r(n)) = r$ if for all $n \in \mathbb{N}$ (hence for all $n \in {}^*\mathbb{N}$), $r(n) = r$ for some $r \in {}^*\mathbb{Q}$.

The definitions of absolute value, addition, subtraction, multiplication, division and the inequalities are similar to those given before in Section 2.1 and all this is preserved by the bijection between $\mathbb{R}'$ and ${}^*\mathbb{R}$ that is defined as follows. Let $x' \in \mathbb{R}'$, then $x' = R'(r(n))$ is generated by the internal Cauchy sequence $r(H(n_i)) = H(r_i(n_i))$, which for each i defines the classical Cauchy sequence $(r_i(n))$, $n \in {}^*\mathbb{N}$, which in turn defines $x_i = R(r_i(n)) \in \mathbb{R}$, and hence $x = H(x_i) \in {}^*\mathbb{R}$. Conversely, each $x \in {}^*\mathbb{R}$ defines an $x' \in \mathbb{R}'$, for if $r_i(n)$ is given for all $n \in \mathbb{N}$, then this defines $r_i(n)$ for all $n \in {}^*\mathbb{N}$. Therefore, $\mathbb{R}'$ can be identified with ${}^*\mathbb{R}$.

2.15 Getting away with generating sequences and $H(s_i)$; summary

Recall from classical analysis that the real numbers were introduced by means of Cauchy sequences of rational numbers. In more formalistic mathematics a real number 'is' the set of all Cauchy sequences concurrent with a given one, but in practice one is seldom working with these sequences. In a similar way we have been introducing the hypernumbers, say the hyperreals, by means of infinite sequences (x_i) of reals x_i. Rather than letting a hyperreal be some set, it was preferred to let it be something new, i.e. $H(x_i)$, generated by (x_i), but a more formalistic procedure could have been followed just as well. Anyway, so many facts regarding nonstandard mathematics have become known in the preceding sections that, just as in classical real number analysis, we can in most cases do without generating sequences. This section serves to summarize these facts.

Just as in classical mathematics in nonstandard mathematics there are the well-known notions of number, set, n-tuple, standard, internal and external notions. Any standard notion is internal, but no notion can be internal and external at the same time (see the figure in Section 1.6). There exists a bijection between the collection of all classical notions and the collection of all standard notions. If s is

a classical notion, then $^*s$, called its $*$-transform, is the corresponding standard one. Although the notation $H(s_i)$ will largely disappear, the asterisks will not. Elements of internal sets are internal, but subsets of internal sets may or may not be internal. For example, $\{1, 2, 3\}$ is an internal, but $\mathbb{N}$ is an external subset of $^*\mathbb{N}$. When external sets directly or indirectly enter the definition of, say, some function, then the latter may turn out to be external as well.

A fairly complete list of all results found sofar that are free from the H-operator is given below.

1.   A constant is internal if and only if it is an element of some standard set.

2.   Let S be a classical set such that $^*s = s$ for each $s \in S$. Then $^*S = S$ if and only if S is finite. Otherwise $S \subseteq {}^*S$.

3.   Let S be a classical set such that each of its elements is a classical set of which each element is equal to its $*$-transform. Then $^*S = S$ if and only if S is finite and all its elements are finite. A similar results holds if S is of any level.

4.   Let f be a function from X onto Y, and assume that X is finite, that $^*x = x$ for all $x \in X$ and that $^*y = y$ for all $y \in Y$. Then $^*X = X$, $^*Y = Y$ and $^*f = f$.

5.   Let $g : W \to X$ and $f : X \to Y$, both g and f be surjective, and assume that X is finite, that $^*x = x$ for all $x \in X$ and that $^*y = y$ for all $y \in Y$. Then,

$$
\begin{aligned}
{}^*(f {\circ} g)(H(w_i)) &= (f {\circ} {}^*g)(H(w_i)) \text{ for all } H(w_i) \in {}^*W, \text{ and} \\
{}^*(f {\circ} g)({}^*w) &= (f {\circ} g)(w) \text{ for all } w \in W,
\end{aligned}
$$

even if W is not finite.

6.   Summary of useful results.
         a) $^*\emptyset = \emptyset$.
         b) $^*S = {}^*T$ if and only if $S = T$,
              $^*S \neq {}^*T$ if and only if $S \neq T$,
              $^*s \in {}^*S$ if and only if $s \in S$,
              $^*S \subseteq {}^*T$ if and only if $S \subseteq T$,
              and similarly for $\subset$, $\supseteq$ and $\supset$.
         c) $^*(S \cup T) = {}^*S \cup {}^*T$,
              $^*(S \cap T) = {}^*S \cap {}^*T$,
              $^*(S - T) = {}^*S - {}^*T$,

if $S \subseteq T$ then $*(S^c) = (*S)^c$, where taking the complement is with respect to T and $*T$, respectively.

d) $*\langle s, t \rangle = \langle *s, *t \rangle$,

and similar equalities hold for n-tuples, $n = 3, 4, \ldots$.

e) $*(f(x)) = *f(*x)$.

f) $*[(f \circ g)(w)] = (*f \circ *g)(*w)$.

7.   If S is a classical infinite set of numbers, then S is external.

8.   If A is a set of numbers, the inclusion $*(\mathcal{P}(A)) \subseteq \mathcal{P}(*A)$ always holds and is strict if and only if A is an infinite set, and if so, $\mathcal{P}(*A)$ contains an element that is an external subset of A.

9.   (The internal definition principle.) Let any statement be given, say $P(x, X, s)$, where X is some set and $x \in X$ makes sense. Let X and s be internal, then so is the set,

$$\{x \in X : P(x, X, s)\}.$$

10.   (Transfer, first formulation.) Let $R(X, X', X'', \ldots; s, s', s'', \ldots)$ be a given statement with constants or free variables $X, X', X'', \ldots$ and $s, s', s'', \ldots$, and bound variables $x, x', x'', \ldots$, where $X, X', X'', \ldots$ are sets, and where x occurs in either $\exists x \in X$ or $\forall x \in X$, and similarly for $x', x'', \ldots$. Then,

$$R(X, X', X'', \ldots; \quad s, s', s'', \ldots) \equiv$$
$$R(*X, *X', *X'', \ldots; \quad *s, *s', *s'', \ldots).$$

11.   (Transfer, second formulation.) Given any internal statement, replacing everything, including every bound variable, by its standard version is equivalent to replacing everything except every bound variable by its standard version.

12.   (The standard definition principle.) Let $*X$ be a standard set, $x \in *X$ make sense, $*s$ be standard, and $P(x, *X, *s)$ be any statement. Then also,

$$\{x \in *X : P(x, *X, *s)\}$$

is standard.

13.   Infinitely large numbers and nonzero infinitesimals exist.

14. $\varepsilon \sim 0$ if and only if $1/\varepsilon \sim +\infty$ or $1/\varepsilon \sim -\infty$, hence s is appreciable if and only if $1/s$ is appreciable.

Let $\varepsilon \sim 0$, $\varepsilon' \sim 0$, s and s' be appreciable, and $\omega \sim \infty$, $\omega' \sim \infty$. Then,

$\varepsilon + \varepsilon' \sim 0$, $\varepsilon - \varepsilon' \sim 0$, $\varepsilon \cdot \varepsilon' \sim 0$,

$\varepsilon + s$ and $\varepsilon - s$ are appreciable, and $\varepsilon \cdot s \sim 0$,

$\varepsilon + \omega \sim \infty$, $\varepsilon - \omega \sim -\infty$, and $\varepsilon \cdot \omega \sim +\infty$ or $-\infty$ or 0, or

$\varepsilon \cdot \omega$ is appreciable,

$s + s'$ and $s - s'$ are appreciable or ~ 0, and $s \cdot s'$ is appreciable,

$s + \omega \sim +\infty$, $s - \omega \sim -\infty$, $s \cdot \omega \sim +\infty$ or $-\infty$,

$\omega + \omega' \sim +\infty$, $\omega - \omega' \sim +\infty$ or $-\infty$ or $\simeq 0$, or $\omega - \omega'$ is appreciable,

and $\omega \cdot \omega' \simeq +\infty$.

15. In $^*\mathbb{R}$ the set of all infinitesimals is external.

16. In $^*\mathbb{R}$ the set of all positive hyperlarge numbers is external.

17. (Overflow or overspill.) Let S be an internal subset of $^*T$, where T is either $\mathbb{N}$ or $\mathbb{Z}$, or $\mathbb{Q}$, or $\mathbb{R}$, such that $\forall m \in \mathbb{N} : \exists s(m) \in S : s(m) \geq m$, i.e. such that from a classical point of view S contains arbitrarily large elements, then $\exists s \in S : s \sim \infty$, i.e. S contains some infinitely large element.

18. (Underflow or underspill.) Let S be an internal subset of $^*T$, with T as before, such that $\forall \omega \in {^*\mathbb{N}}, \omega \sim \infty : \exists s(\omega) \in S : s(\omega) < \omega \wedge s(\omega) \sim \infty$, i.e. such that S contains infinitely large elements that are arbitrarily small, then $\exists s \in S : s$ is limited.

19. ('Inverse' overflow.) Let S be an internal subset of $^*\mathbb{Q}$ or $^*\mathbb{R}$, such that $\forall m \in \mathbb{N} : \exists s(m) \in S :\mid s(m) \mid \leq 1/m$, i.e. such that from a classical point of view S contains arbitrarily small elements, then $\exists s \in S : s \simeq 0$.

20. ('Inverse' underflow.) Let S be an internal subset of $^*\mathbb{Q}$ or $^*\mathbb{R}$, such that $\forall \varepsilon, \varepsilon \sim 0 : \exists s \in S : s \geq \varepsilon$, then $\exists s \in S : s$ is not an infinitesimal.

21. If $n \in {^*\mathbb{N}}$, then either $n \in \mathbb{N}$ or n is hyperlarge.

22. Given any $\omega \in {^*\mathbb{N}}$, $\omega \sim \infty$, then $S = [1, \omega]$ is uncountable.

23. Let f be an internal function from $^*\mathbb{N}$ to $\{1, 2\}$, such that both 1 and 2 are assumed somewhere, i.e. such that f is onto. Moreover, let $f(n) = 2$ for all $n \geq b$ for some $b \in {^*\mathbb{N}}$. Then there is a $\beta < b$, $\beta \in {^*\mathbb{N}}$ such that $f(\beta) = 1$ and $f(n) = 2$ if $n > \beta$. Hence there is a last such that $f(\beta) = 1$, even though $b \sim \infty$ is allowed, and $\{n : n \leq b\}$ is uncountable in that case.

24. (Standard part theorem.) If $x \in {}^*\mathbb{R}$ then either $\mid x \mid \sim \infty$, or $x = r + \varepsilon$, $r \in \mathbb{R}$, $\varepsilon \simeq 0$ for unique r and ε.

25. Given any $r \in \mathbb{R}$, there exists an $x \in {}^*\mathbb{Q}$ such that $\mathrm{st}(x) = r$.

26. Let $a, b \in \mathbb{R}$, $a < b$. If $x \in {}^*[a, b]$, then $\mathrm{st}(x) \in [a, b]$, but if the interval in $\mathbb{R}$ is not closed, this is sometimes not true.

Chapter 3

Some applications

3.1 Introduction and least upper bound theorem

The aim of this chapter is to show how many definitions and proofs of elementary calculus can be simplified by means of nonstandard analysis. Only a number of important examples will be considered. A much more complete treatment is Keisler [26], where the existence of nonstandard numbers is taken for granted, however, and a simplified form of transfer is introduced in an axiomatic kind of way.

Theorem 3.1.1 *(The least upper bound theorem.)*
Let S be a nonempty subset of $\mathbb{R}$ that is bounded above by some (classical) real number. Then S has a least upper bound in $\mathbb{R}$.

Proof: Taking any $c \in S$, instead of S we may consider $\{s : s \in S, s \geq c\}$, that is to say we may assume that $s \geq c$ for all $s \in S$. Then $c, b \in \mathbb{R}$, $c < b$, exist such that $\forall s \in S : c \leq s \leq b$, so that, by transfer, $\forall s \in^* S : c \leq s \leq b$. Let $\omega \in^* \mathbb{N}$, $\omega \sim \infty$ be arbitrary and divide $^*[c,b]$ in ω equal subintervals of length $\delta = (b-c)/\omega$, so that $\delta \sim 0$, and consider the points a, $a + \delta$, $a + 2\delta$, ..., $a + \omega\delta = b$. Then,

$$\exists j \in {}^*\mathbb{N} : [\forall s \in {}^*S : s \leq a + j\delta] \quad \wedge \quad [\exists s' \in {}^*S : s' > a + j\delta - \delta].$$

Let $\beta = \mathrm{st}(a + j\delta)$, which is well defined as $a + j\delta$ is limited. Then β is a (hence the) least upper bound of S. For first of all if $s \in S$ then $s \in {}^*S$, hence $s \leq a + j\delta = \beta + \varepsilon$ for some $\varepsilon \simeq 0$, but since $s, \beta \in \mathbb{R}$ this means that $s \leq \beta$. And secondly, if β' were a smaller upper bound of S, then $\beta > \beta' + 1/m$ for some $m \in \mathbb{N}$, hence

$$\beta' \geq s' > a+j\delta-\delta = \beta+\varepsilon-\delta > \beta'+1/m+\varepsilon-\delta, \text{ or } \delta-\varepsilon > 1/m, \text{ a contradiction.}$$
□

Note that this proof is not much shorter than its classical counterpart, that essentially runs as follows. Let $a_1 = a$, $b_1 = b$ and $n = 1$.

($*$) If $S \subseteq [a_n, (a_n + b_n)/2]$, then let $a_{n+1} = a_n$, $b_{n+1} = (a_n + b_n)/2$, otherwise let $a_{n+1} = (a_n + b_n)/2$, $b_n + 1 = b_n$. In either case replace n by $n + 1$ and start again from ($*$).

This procedure defines two concurrent Cauchy sequences (a_n) and (b_n), both converging to the least upper bound of S (the reader may work out the details).

Since the theorem is wrong if $\mathbb{R}$ is replaced by $\mathbb{Q}$, both proofs must use something that is typical for $\mathbb{R}$. Indeed each limited element of $^*\mathbb{R}$ (not $^*\mathbb{Q}$) has a standard part, and any Cauchy sequence converges to some element of $\mathbb{R}$ (not $\mathbb{Q}$). This illustrates the obvious fact that a nonstandard proof must contain all essential steps – perhaps in disguise – of the corresponding classical proof.

3.2   Simplifying definitions and proofs of elementary calculus

First of all recall from Sections 1.4 and 1.5 that a function f from $\mathbb{R}$ to $\mathbb{R}$ is continuous at $c \in \mathbb{R}$ if,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists \delta \in \mathbb{R}, \delta > 0 : \forall x \in \mathbb{R}, \mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon$$

or, equivalently, if,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta : \mid {}^*f(x) - {}^*f(c) \mid < \varepsilon$$

or, equivalently, if,

$$\forall \delta \in {}^*\mathbb{R}, \delta \simeq 0 : {}^*f(c + \delta) - {}^*f(c) \simeq 0.$$

The first simplification, therefore, reads as follows.

Theorem 3.2.1 *(Simplified definition of the continuity of real-valued real functions.)*
$f : \mathbb{R} \to \mathbb{R}$ *is continuous at* $c \in \mathbb{R}$ *if and only if,*

$$\forall \delta \in {}^*\mathbb{R}, \delta \simeq 0 : {}^*f(c + \delta) - {}^*f(c) \simeq 0.$$

□

Theorem 3.2.2 *(Simplified definition of uniform continuity.)*
$f : S \to \mathbb{R}$, $S \subseteq \mathbb{R}$ *is uniformly continuous in* S *if and only if,*

$$\forall x, y \in {}^*S, x - y \simeq 0 : {}^*f(x) - {}^*f(y) \simeq 0.$$

Proof: Recall that $f : S \to \mathbb{R}$, $S \subseteq \mathbb{R}$ is uniformly continuous in S if,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0, \exists \delta \in \mathbb{R}, \delta > 0 : \forall x, y \in S, \mid x - y \mid < \delta : \mid f(x) - f(y) \mid < \varepsilon.$$

By transfer, this is equivalent to,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0,$$
$$\forall x, y \in {}^*S, \mid x - y \mid < \delta : \mid {}^*f(x) - {}^*f(y) \mid < \varepsilon.$$

But this can be simplified to,

$$\forall x, y \in {}^*S, x - y \simeq 0 : {}^*f(x) - {}^*f(y) \simeq 0.$$

For let (3.1) be true and let $m \in \mathbb{N}$ be given arbitrarily. Then there exist $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$ such that $\varepsilon < 1/m$, and δ as in (3.1). Hence $\forall x, y \in S, \mid x - y \mid < \delta : \mid f(x) - f(y) \mid < 1/m$, or, by transfer,

$$\forall x, y \in {}^*S, \mid x - y \mid < \delta : \mid {}^*f(x) - {}^*f(y) \mid < 1/m,$$

so that, as m was arbitrary, ${}^*f(x) - {}^*f(y) \simeq 0$, which proves (3.1), since in absolute value any infinitesimal is smaller than δ.

Conversely, let (3.1) be true, let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$ be arbitrary and let $\delta \simeq 0$, $\delta > 0$. Hence if $x, y \in {}^*S$, $\mid x - y \mid < \delta$ then ${}^*f(x) - {}^*f(y) = \varepsilon'$ for some $\varepsilon' \simeq 0$. In other words, since $\mid \varepsilon' \mid < \varepsilon$,

$$\exists \delta' \in {}^*\mathbb{R}, \delta' > 0 : \forall x, y \in {}^*S, \mid x - y \mid < \delta' : \mid {}^*f(x) - {}^*f(y) \mid < \varepsilon,$$

(take, for example $\delta' = \delta$) or, by transfer (in the opposite direction),

$$\exists \delta' \in \mathbb{R}, \delta' > 0 : \forall x, y \in S, \mid x - y \mid < \delta' : \mid {}^*f(x) - {}^*f(y) \mid < \varepsilon,$$

which proves (3.1) since ε was arbitrary. $\qquad\square$

Theorem 3.2.3 *If f is continuous at each $x \in [a, b]$, a, $b \in \mathbb{R}$, $a < b$, then f is uniformly continuous in $[a, b]$.*

Simplified proof: Let x, y be in $^*[a, b]$, then by Theorem 2.13.3, $\text{st}(x) \in [a, b]$ and $\text{st}(y) \in [a, b]$. Let $x - y \simeq 0$. Since $y - \text{st}(y) \simeq 0$ it follows that $x - \text{st}(y) \simeq 0$. By continuity, $^*f(y) - {}^*f(\text{st}(y)) \simeq 0$ and $^*f(x) - {}^*f(st(y)) \simeq 0$, so that $^*f(x) - {}^*f(y) \simeq 0$. $\qquad \square$

Theorem 3.2.4 *(Simplified limit definition.)*
Let $f : \mathbb{R} \to \mathbb{R}$, then $\lim\limits_{x \to c} f(x) = k$, $c, k \in \mathbb{R}$ if and only if,

$$\forall \delta \in {}^*\mathbb{R}, \delta \sim 0 : {}^*f(c + \delta) - k \simeq 0,$$

so that $k = \text{st}[^*f(c + \delta)]$.

Proof: By definition, the limit exists if and only if,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists \delta \in \mathbb{R}, \delta > 0 : \forall x \in \mathbb{R}, 0 <| x - c |< \delta :| f(x) - k |< \varepsilon,$$

or, by transfer,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, 0 <| x - c |< \delta :| {}^*f(x) - k |< \varepsilon,$$

which can be simplified to,

$$\forall \delta \in {}^*\mathbb{R}, \delta \sim 0 : {}^*f(c + \delta) - k \simeq 0.$$

$\qquad \square$

Exercise: Complete this proof.

Now let $f : \mathbb{N} \to \mathbb{R}$, so that f is an infinite sequence, and let c be replaced by ∞.

Theorem 3.2.5 *(Another simplified limit definition.)*
Let $f : \mathbb{N} \to \mathbb{R}$ and $k \in \mathbb{R}$, then $\lim\limits_{x \to \infty} f(n) = k$, $k \in \mathbb{R}$, if and only if,

$$\forall n \in {}^*\mathbb{N}, n \sim \infty : {}^*f(n) - k \simeq 0,$$

so that $k = \text{st}[^*f(n)]$.

Proof: By definition, the limit exists if and only if,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists n' \in \mathbb{N} : \forall n \in \mathbb{N}, n > n' :| f(n) - k |< \varepsilon,$$

or, by transfer,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists n' \in {}^* N : \forall n \in {}^*\mathbb{N}, n > n' :| {}^*f(n) - k |< \varepsilon,$$

which can be simplified to,

$$\forall n \in {}^*\mathbb{N}, n \sim \infty : {}^*f(n) - k \simeq 0.$$

$\square$

Exercise: Again complete the proof.

Exercise: Treat the cases where the limit itself is infinite.

Theorem 3.2.6 *If f is a nondecreasing infinite sequence, that is bounded above, then $f(n)$ has a finite limit for n tending to ∞.*

Classical proof: The set $\{f(n) : n \in \mathbb{N}\}$ is bounded above, hence in $\mathbb{R}$ has a least upper bound β, so that $f(n) \leq \beta$ for all $n \in \mathbb{N}$, and for each $m \in \mathbb{N}$, $f(n') > \beta - 1/m$ for some $n' \in \mathbb{N}$, and since f is nondecreasing this implies that $\lim_{n \to \infty} f(n) = \beta$. $\square$

Exercise: Give a direct nonstandard proof similar to that of Theorem 3.1.1 (the least upper bound theorem), not using Theorem 3.1.1.

Theorem 3.2.7 *(The intermediate value theorem.)*
If $a, b \in \mathbb{R}$, $a < b$, and $f(a) < 0$, $f(b) > 0$, then $f(c) = 0$ for some c, $a < c < b$.

Simplified proof: See Section 1.4. $\square$

Theorem 3.2.8 *(The extreme value theorem.)*
Let $f : [a,b] \to \mathbb{R}$, $a, b \in \mathbb{R}$, $a < b$, and let f be continuous at each point of $[a,b]$. Then $f(x) \leq f(c)$ for some $c \in [a,b]$ and all $x \in [a,b]$, i.e. f has a maximum somewhere in the closed interval between a and b. And similarly for minimum.

Simplified proof: Let $\omega \in {}^*\mathbb{N}$, $\omega \sim \infty$, be arbitrary, and divide ${}^*[a,b]$ in ω equal subintervals of length $\delta = (b-a)/\omega$. Let $n \in {}^*\mathbb{N}$ be such that ${}^*f(a + n\delta) \geq {}^*f(a + i\delta)$ for all $i = 0, 1, \ldots, \omega$. The existence of n follows by transfer, since any finite set has a maximum, hence so has any hyperfinite set. Obviously, $a + n\delta$ is limited, hence $c = \mathrm{st}(a + n\delta)$ is well defined and by continuity, ${}^*f(a + n\delta) - f(c) = \varepsilon$ for some $\varepsilon \simeq 0$. Each $x \in [a,b]$ is within the distance δ of some $a + i\delta$ and $\delta \sim 0$, hence, again by continuity, $f(x) = {}^*f(a + i\delta) + \varepsilon'$ for some $\varepsilon' \simeq 0$, hence,

$$f(x) \leq {}^*f(a + n\delta) + \varepsilon' = f(c) + \varepsilon + \varepsilon', \text{ i.e. } f(x) \leq f(c).$$

$\square$

Theorem 3.2.9 *(The composite function theorem.)*
Let $g(w)$ be defined for w in a neighborhood of $c \in \mathbb{R}$, and let $f(x)$ be defined for x in a neighborhood of $g(c)$. Then $f \circ g$ is continuous at c if g is continuous at c and f is continuous at $g(c)$.

Simplified proof: Let $\delta \simeq 0$, then ${}^*g(c + \delta) - g(c) \simeq 0$, hence it follows that ${}^*f({}^*g(c + \delta)) - f(g(c)) \simeq 0$. $\qquad\square$

3.3 Continuity and limits for internal functions

So far nonstandard characterizations were given for continuity and limits of classical functions. How about arbitrary internal functions? Let f be some internal function from ${}^*\mathbb{R}$ to ${}^*\mathbb{R}$, and let $c \in {}^*\mathbb{R}$, so that c may be hyperlarge or 'almost standard' i.e. be the sum of a real number and a nonzero infinitesimal. Or rather, let F be the set of all classical functions from $\mathbb{R}$ to $\mathbb{R}$ and let $f \in {}^*F$. Then, by definition D) of Section 2.9, for suitable f_i and c_i, $f = H(f_i)$ is *continuous at $c = H(c_i)$ if for all $i \in \mathbb{N}$, f_i is continuous at c_i. Here $f_i : \mathbb{R} \to \mathbb{R}$ and $c_i \in \mathbb{R}$.

Theorem 3.3.1 *(Continuity of internal functions.)*
The internal $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ is *continuous at $c \in {}^*\mathbb{R}$ if and only if in the classical definition $\mathbb{R}$ is replaced by ${}^*\mathbb{R}$, i.e. if,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta :\mid f(x) - f(c) \mid < \varepsilon.$$

Proof: Letting $f = H(f_i)$ and $c = H(c_i)$, f is *continuous at c if and only if,

$$H[\forall \varepsilon_i \in \mathbb{R}, \varepsilon_i > 0 \quad : \quad \exists \delta_i \in \mathbb{R}, \delta_i > 0 :$$
$$\forall x_i \in \mathbb{R}, \mid x_i - c_i \mid < \delta_i \quad : \quad \mid f_i(x_i) - f_i(c_i) \mid < \varepsilon_i].$$

By Łoš' theorem (Theorem 2.7.1) this is equivalent to,

$$\forall H(\varepsilon_i) \in {}^*\mathbb{R}, H(\varepsilon_i) > 0 \quad : \quad \exists H(\delta_i) \in {}^*\mathbb{R}, H(\delta_i) > 0 :$$
$$\forall H(x_i) \in {}^*\mathbb{R}, \mid H(x_i) - H(c_i) \mid < H(\delta_i) \quad : \quad \mid H(f_i)(H(x_i)) - H(f_i)(H(c_i)) \mid < H(\varepsilon_i)$$

and hence to what has to be proved. $\qquad\square$

Warning: If f or c is nonstandard, *continuity is not always equivalent to,

$$\forall \delta \in {}^*\mathbb{R}, \delta \simeq 0 : f(c + \delta) - f(c) \simeq 0.$$

Counterexamples:

a) c standard, but f nonstandard; let $\omega \sim \infty$, $f(x) = \omega x$, $c = 1$, $\delta = 1/\omega^{1/2}$; then $f(c + \delta) - f(c) = \omega\delta \sim \infty$.

b) f standard, but c nonstandard; let $f(x) = x^2$, $\delta \sim 0$, $c = 1/\delta$; then $f(c + \delta) - f(c) = 2 + \delta^2 \sim 2$.

Yet, $\forall \delta \in {}^*\mathbb{R}$, $\delta \simeq 0 : f(c + \delta) - f(c) \simeq 0$ makes sense for arbitrary internal f and c. If this is true, then f is called *S-continuous* at c.

Examples:

a) Let α be a positive infinitesimal, and $f(x) = \alpha x$ if $x \geq 0$, $f(x) = 0$ if $x < 0$. Then f is S-continuous everywhere in ${}^*\mathbb{R}$. It is *continuous at $c \in {}^*\mathbb{R}$ if $c \neq 0$, but not at $c = 0$.

b) Let $\omega \sim \infty$ and $f(x) = \omega x$. Then f is nowhere S-continuous, since $f(x) - f(c) = \omega(x - c) = \omega^{1/2}$ if $x - c = \omega^{-1/2} \sim 0$. It is continuous everywhere in ${}^*\mathbb{R}$.

c) $f(x) = x^2$. Then f is not S-continuous if $c \sim \infty$. It is *continuous everywhere in ${}^*\mathbb{R}$.

Theorem 3.3.2 *The internal function $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ is S-continuous at $c \in {}^*\mathbb{R}$ if and only if,*

$$\forall \varepsilon \in \mathbb{R}, \varepsilon < 0 : \exists \delta \in \mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon.$$

(Note that both ε and δ are standard, but that x is internal.)

Proof: The if part. Let $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$ and $\delta \in {}^*\mathbb{R}$, $\delta \simeq 0$ be given arbitrarily. Then there is a $\delta' \in \mathbb{R}$, $\delta' > 0$ such that,

$$\forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta' : \mid f(x) - f(c) \mid < \varepsilon.$$

As $\mid \delta \mid < \delta'$, so that $\mid x - c \mid < \delta'$ if $x = c + \delta$, it follows that $\mid f(c + \delta) - f(c) \mid < \varepsilon$, and since ε is arbitrary that $f(c + \delta) - f(c) \simeq 0$.

The only-if part. Conversely, let ε and δ be as before but such that $\delta > 0$. Then $\forall x \in {}^*\mathbb{R}$, $\mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon$. Now let the set S be defined by,

$$S = \{\delta \in {}^*\mathbb{R} : \delta > 0 \text{ and } \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon\}.$$

Then if $\delta \in S$ every number between 0 and δ is in S. By the internal definition principle (Corollary 2.7.1) S is internal, but it contains as a subset the set of all positive infinitesimals and the latter is external (see Theorems 2.10.3 and 2.11.4),

so that S must contain some $\delta > 0$ that is not an infinitesimal, and it follows that S must contain some $\delta > 0$, $\delta \in \mathbb{R}$. $\qquad\qquad\square$

Remark: In this proof the fact that an external set is not internal has been used. This fact is called *Cauchy's principle*. It is an example of a *principle of permanence*. In general this is a statement that if some set S contains some subset T, the latter is strictly contained in S, because S and T happen to be different *kinds* of set. In classical mathematics such principles do not seem to play a real part, but in nonstandard mathematics there are several of them, although there are only a few primary forms, or perhaps only one, i.e. Cauchy's principle, which obviously really is a matter of definition. See Section 4.1 for more details.

In a similar way the $*$-transform of uniform continuity can be introduced: simply copy the classical definition and replace $\mathbb{R}$ by $^*\mathbb{R}$. And the simplified form of uniform continuity leads to S-uniform continuity. Hence the internal $f : {^*}\mathbb{R} \to {^*}\mathbb{R}$ is *S-uniformly continuous* in $^*\mathbb{R}$ if,

$$\forall x, y \in {^*}\mathbb{R}, x - y \simeq 0 : f(x) - f(y) \simeq 0.$$

Exercise: Formulate and show a theorem similar to Theorem 3.3.2, but for uniform continuity.

It is generally agreed to drop the asterisks in both $*$continuity and $*$uniform continuity, as well as in similar indications. But keep in mind that S-continuity is then not a special form of continuity, and similarly for other attributes.

Turning to $*$limits, the definitions similar to those in Section 3.2 are, dropping the asterisks: the internal $f : {^*}\mathbb{R} \to {^*}\mathbb{R}$ tends to the limit $k \in {^*}\mathbb{R}$ for $x \in {^*}\mathbb{R}$ tending to $c \in {^*}\mathbb{R}$, if,

$$\forall \varepsilon \in {^*}\mathbb{R}, \varepsilon > 0 : \exists \delta \in {^*}\mathbb{R}, \delta > 0 : \forall x \in {^*}\mathbb{R}, 0 <\mid x - c \mid< \delta : \mid f(x) - k \mid< \varepsilon.$$

And the internal $f : {^*}\mathbb{N} \to {^*}\mathbb{R}$ tends to the limit $k \in {^*}\mathbb{R}$ for $n \in {^*}\mathbb{N}$ tending to infinity, if,

$$\forall \varepsilon \in {^*}\mathbb{R}, \varepsilon > 0 : \exists n' \in {^*}\mathbb{N} : \forall n \in {^*}\mathbb{N}, n > n' : \mid f(n) - k \mid< \varepsilon.$$

Similar to S-continuity the definitions of S-limit are as follows. The internal $f : {^*}\mathbb{R} \to {^*}\mathbb{R}$ tends to the *S-limit* $k \in {^*}\mathbb{R}$ for $x \in {^*}\mathbb{R}$ tending to $c \in {^*}\mathbb{R}$, if,

$$\forall \delta \sim 0 : f(c + \delta) - k \simeq 0.$$

And the internal $f : {^*}\mathbb{N} \to {^*}\mathbb{R}$ tends to the *S-limit* $k \in {^*}\mathbb{R}$ for $n \in {^*}\mathbb{N}$ tending to infinity, if,

$$\forall n' \in {^*}\mathbb{N}, n' \sim \infty : f(n') - k \simeq 0.$$

Theorem 3.3.3 *The internal* $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$ *tends to the S-limit* $k \in {}^*\mathbb{R}$ *for* $x \in {}^*\mathbb{R}$ *tending to* $c \in {}^*\mathbb{R}$, *if and only if,*

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists \delta \in \mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, 0 <\mid x - c \mid< \delta : \mid f(x) - k \mid< \varepsilon.$$

Proof: Left as an exercise. □

Theorem 3.3.4 *The internal* $f : {}^*\mathbb{N} \to {}^*\mathbb{R}$ *tends to the S-limit* $k \in {}^*\mathbb{R}$ *for* $n \in {}^*\mathbb{N}$ *tending to infinity, if and only if,*

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists n' \in \mathbb{N} : \forall n \in {}^*\mathbb{N}, n > n' : \mid f(n) - k \mid< \varepsilon.$$

Proof: Left as an exercise. □

A special case arises when k is finite, because then $\mathrm{st}(k)$ is well defined. Then also $f(x)$ or $f(n)$ are finite for x close enough to c or n large enough. Assuming in the first case that c is finite as well, it follows that the classical limits for x tending to $\mathrm{st}(c)$ or n tending to ∞ are equal to $\mathrm{st}(k)$.

Theorem 3.3.5 *Let* f, c *and* k *be as before, and let* k *and* c *be finite. If* $f(x)$ *tends to* k *for* $x \in {}^*\mathbb{R}$ *tending to* c, *then,*

$$\lim_{x \to st(c)} st(f(x)) = st(k), \ \text{where } x \in \mathbb{R},$$

and if $f(n)$ *tends to* k *for* $n \in {}^*\mathbb{N}$ *tending to* ∞, *then,*

$$\lim_{n \to \infty} st(f(n)) = st(k), \ \text{where } n \in \mathbb{N}.$$

Proof: In the first case

$$\exists \delta_1 \in \mathbb{R}, \delta_1 > 0 : \forall x \in {}^*\mathbb{R}, 0 <\mid x - c \mid< \delta_1 :\mid f(x) - k \mid< 1,$$

so that $\forall x \in \mathbb{R}, 0 <\mid x - \mathrm{st}(c) \mid< \delta_1/2 :\mid f(x) \mid<\mid \mathrm{st}(k) \mid +2$, which means that $f(x)$ is finite for these x's, from which the first claim follows. The second claim is shown in a similar way. □

Exercise: Treat the cases where the limit itself is infinite. And define the corresponding S-limits.

3.4 More nonstandard characterizations of classical notions

Theorem 3.4.1 *(Nonstandard characterization of Cauchy sequence.)*
$s(n)$ is a classical Cauchy sequence if and only if,

$$\forall n, p \in {}^*\mathbb{N}, n, p \sim \infty : {}^*s(n) - {}^*s(p) \simeq 0.$$

Proof: By definition,

$$\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : \forall n, p \in \mathbb{N}, n, p > k :\mid s(n) - s(p) \mid < 1/m,$$

and by transfer, fixing $m \in \mathbb{N}$ and $k \in \mathbb{N}$,

$$\forall n, p \in \mathbb{N}, n, p > k :\mid s(n) - s(p) \mid < 1/m,$$

is equivalent to,

$$\forall n, p \in {}^*\mathbb{N}, n, p > k :\mid {}^*s(n) - {}^*s(p) \mid < 1/m.$$

Now let $n, p \sim \infty$, so that automatically $n, p > k$, no matter the value of $k \in {}^*\mathbb{N}$, then (3.1) implies that,

$$\forall m \in \mathbb{N} : \exists k \in \mathbb{N} : n, p \in {}^*\mathbb{N}, n, p \sim \infty :\mid {}^*s(n) - {}^*s(p) \mid < 1/m.$$

But since k plays no part any more this can be simplified to,

$$\forall m \in \mathbb{N} : \forall n, p \in {}^*\mathbb{N}, n, p \sim \infty :\mid {}^*s(n) - {}^*s(p) \mid < 1/m,$$

hence to,

$$\forall n, p \in {}^*\mathbb{N}, n, p \sim \infty : \forall m \in \mathbb{N} :\mid {}^*s(n) - {}^*s(p) \mid < 1/m,$$

hence to,

$$\forall n, p \in {}^*\mathbb{N}, n, p \sim \infty : {}^*s(n) - {}^*s(p) \simeq 0,$$

which is (3.1).

Conversely, consider the negation of (3.5), that is,

$$\exists m \in \mathbb{N} : \forall k \in \mathbb{N} : \exists n, p \in \mathbb{N}, n, p > k :\mid s(n) - s(p) \mid \geq 1/m,$$

fix $m \in \mathbb{N}$ and apply transfer to,

$$\forall k \in \mathbb{N} : \exists n, p \in \mathbb{N}, n, p > k :\mid s(n) - s(p) \mid \geq 1/m,$$

giving,

$$\forall k \in {}^{*}\mathbb{N} : \exists n, p \in {}^{*}\mathbb{N}, n, p > k :\mid {}^{*}s(n) - {}^{*}s(p) \mid \geq 1/m,$$

which implies, fixing $k \sim \infty$ arbitrarily, that $n, p \sim \infty$, hence,

$$\exists n, p \in {}^{*}\mathbb{N}, n, p \sim \infty :\mid {}^{*}s(n) - {}^{*}s(p) \mid \geq 1/m,$$

so that (3.1) implies that,

$$\exists m \in \mathbb{N} : \exists n, p \in {}^{*}\mathbb{N}, n, p \sim \infty :\mid {}^{*}s(n) - {}^{*}s(p) \mid \geq 1/m,$$

or,

$$\exists n, p \in {}^{*}\mathbb{N}, n, p \sim \infty : \exists m \in \mathbb{N} :\mid {}^{*}s(n) - {}^{*}s(p) \mid \geq 1/m,$$

or,

$$\exists n, p \in {}^{*}\mathbb{N}, n, p \sim \infty : \neg[{}^{*}s(n) - {}^{*}s(p) \simeq 0],$$

which is the negation of (3.4). $\qquad\square$

Theorem 3.4.2 *(Nonstandard characterization of bounded set.)*
Let L be the set of all limited elements of ${}^{*}\mathbb{R}$. Then $S \subseteq \mathbb{R}$ is bounded if and only if ${}^{*}S \subseteq L$.

Proof: S is bounded if,

$$\exists m \in \mathbb{N} : \forall s \in S :\mid s \mid \leq m,$$

hence, by transfer, if,

$$\exists m \in \mathbb{N} : \forall s \in {}^{*}S :\mid s \mid \leq m,$$

so that ${}^{*}S \subseteq L$.

Conversely, if S is not bounded then,

$$\forall m \in \mathbb{N} : \exists s \in S :\mid s \mid > m,$$

hence, by transfer,

$$\forall m \in {}^{*}\mathbb{N} : \exists s \in {}^{*}S :\mid s \mid > m,$$

and taking m hyperlarge it follows that $\mid s \mid$ is hyperlarge for some $s \in {}^{*}S$, so that $s \notin L$. $\qquad\square$

Theorem 3.4.3 *(Nonstandard characterization of open set.)*
Let $S \subseteq \mathbb{R}$ and let $h(S) = \{t \in {}^*\mathbb{R} : t \simeq s$ for some $s \in S\}$. Then S is open if and only if,

$$h(S) \subseteq {}^*S.$$

Proof: S is open if,

$$\forall s \in S : \exists m \in \mathbb{N} : \forall t \in \mathbb{R}, \mid t - s \mid < 1/m : t \in S,$$

hence, by transfer,

$$\forall s \in S : \exists m \in \mathbb{N} : \forall t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \in {}^*S,$$

so that, restricting t such that $t \simeq s$,

$$\forall s \in S : \forall t \in {}^*\mathbb{R}, t \simeq s : t \in {}^*S, \text{ i.e. } h(S) \subseteq {}^*S.$$

Conversely, if S is not open, then,

$$\exists s \in S : \forall m \in \mathbb{N} : \exists t \in \mathbb{R}, \mid t - s \mid < 1/m : t \notin S,$$

hence, by transfer,

$$\exists s \in S : \forall m \in {}^*\mathbb{N} : \exists t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \notin {}^*S,$$

and taking m hyperlarge it follows that for some $s \in S$ and some $t \in {}^*\mathbb{R}$ we have that $t \simeq s$, but $t \notin {}^*S$, hence that $h(S)$ is not a subset of ${}^*S$. $\qquad\square$

Remark: $h(S)$ is called the *halo* (or the *monad*) of S.

Theorem 3.4.4 *(Nonstandard characterization of closed set.)*
$S \subseteq \mathbb{R}$ is closed if and only if,

$$h(S^c) \subseteq {}^*(S^c) = ({}^*S)^c.$$

Proof: Follows directly from the previous theorem. $\qquad\square$

Exercise: Show the last theorem independently of the previous theorem.

Theorem 3.4.5 *(Nonstandard characterization of interior point.)*
Let $s \in \mathbb{R}$ and let $h(s) = \{t \in \mathbb{R} : t \simeq s\}$. Then s is an interior point of $S \subseteq \mathbb{R}$ if and only if $h(s) \subseteq {}^*S$.

Proof: Since s is an interior point of S if,

$$\exists m \in \mathbb{N} : \forall t \in \mathbb{R}, \mid t - s \mid < 1/m : t \in S,$$

the proof is a simplified version of that of Theorem 3.4.3. The details are left as an exercise. □

Remark: In view of the previous remark, $h(s)$ is of course called the *halo* of the point s. Note that $h(0)$ is the set of all infinitesimals.

Theorem 3.4.6 *(Nonstandard characterization of boundary point.)*
$s \in \mathbb{R}$ *is a boundary point of* $S \subseteq \mathbb{R}$ *if and only if both* $h(s) \cap {}^*S$ *and* $h(s) \cap ({}^*S)^c$ *are nonempty.*

Proof: If s is a boundary point of S then,

$$\forall m \in \mathbb{N} : [\exists t \in \mathbb{R}, \mid t - s \mid < 1/m : t \in S] \wedge [\exists t \in \mathbb{R}, \mid t - s \mid < 1/m : t \notin S]$$

hence, by transfer,

$$\forall m \in {}^*\mathbb{N} : [\exists t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \in {}^*S] \wedge [\exists t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \notin {}^*S]$$

so that, taking $m \sim \infty$,

$$[\exists t : t \simeq s, t \in {}^*S] \wedge [\exists t : t \simeq s, t \notin {}^*S],$$

hence $h(s) \cap {}^*S \neq \emptyset$ and $h(s) \cap ({}^*S)^c \neq \emptyset$.

Conversely, if s is not a boundary point of S, then,

$$\exists m \in \mathbb{N} : [\forall t \in \mathbb{R}, \mid t - s \mid < 1/m : t \notin S] \vee [\forall t \in \mathbb{R}, \mid t - s \mid < 1/m : t \in S],$$

hence, by transfer,

$$\exists m \in \mathbb{N} : [\forall t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \notin {}^*S] \vee [\forall t \in {}^*\mathbb{R}, \mid t - s \mid < 1/m : t \in {}^*S].$$

The first substatement between square brackets implies that if $t \simeq s$ then $t \notin {}^*S$, so that $h(s) \subseteq ({}^*S)^c$, and the second one similarly that $h(s) \subseteq {}^*S$, so that either $h(s) \cap {}^*S = \emptyset$ or $h(s) \cap ({}^*S)^c = \emptyset$. □

Theorem 3.4.7 *(Nonstandard characterizations of accumulation point and closure.)*
$s \in \mathbb{R}$ *is an accumulation point (or limit point) of* $S \subseteq \mathbb{R}$ *if and only if,*

$$\exists t \in {}^*S, t \neq s : t \sim s.$$

Let cl S be the closure of S. Then $s \in cl\ S$ if and only if,

$$\exists t \in {}^*S : t \simeq s.$$

Proof: If s is an accumulation point of S then,

$$\forall m \in \mathbb{N} : \exists t \in S, t \neq s :\mid t - s \mid < 1/m,$$

hence, by transfer,

$$\forall m \in {}^*\mathbb{N} : \exists t \in {}^*S, t \neq s :\mid t - s \mid < 1/m,$$

so that, taking m hyperlarge, $\exists t \in {}^*S$, $t \neq s$: $t \sim s$.

Conversely, if s is not an accumulation point of S, then,

$$\exists m \in \mathbb{N} : \forall t \in S, t \neq s :\mid t - s \mid \geq 1/m,$$

hence, by transfer,

$$\exists m \in \mathbb{N} : \forall t \in {}^*S, t \neq s :\mid t - s \mid \geq 1/m,$$

so that,

$$\forall t \in {}^*S, t \neq s : \neg[t \simeq s].$$

The second part of the theorem follows by observing that $s \in \mathrm{cl}\, S$ if and only if $s \in S$ or else if s is an accumulation point of S. $\qquad \square$

Exercise: Give an alternative proof of Theorem 3.4.4, using the fact that S is closed if and only if $S = \mathrm{cl}\, S$.

3.5 Inverse functions; b^c

Recall that a function $f : S \to T$ has an *inverse* f^{-1} if and only if f is bijective, and then $f^{-1}(t) = s$ if $f(s) = t$.

Theorem 3.5.1 *Let a function f be monotonically increasing (or decreasing) and be continuous in $[a,b]$, $a, b \in \mathbb{R}$, $a < b$. Then,*

1)   *range (f), the range of f, is a finite closed interval,*
1)   *f has an inverse,*
1)   *f^{-1} too is monotonically increasing (or decreasing), and*
1)   *f^{-1} is continuous in its domain.*

Proof: Only the case where f is increasing is considered.

1) As $a \leq x \leq b$ implies that $f(a) \leq f(x) \leq f(b)$, range $(f) \subseteq [f(a), f(b)]$. And if $f(a) \leq w \leq f(b)$ then by the intermediate value theorem (Theorem 3.2.7) there is a $c \in [a, b]$ such that $f(c) = w$ which means that $[f(a), f(b)] \subseteq$ range (f). Therefore, range $(f) = [f(a), f(b)]$.

2, 3) Clearly, f^{-1} exists, $[f(a), f(b)]$ is its domain, and it is increasing.

4) Let $w = f(c) \in [f(a), f(b)]$, and let $\varepsilon \sim 0$. If $\varepsilon > 0$ then, of course, w must be smaller than $f(b)$, and if $\varepsilon < 0$ then w must be larger than $f(a)$. Assume $\varepsilon > 0$. Note that $*(f^{-1}) = (*f)^{-1}$, so that parentheses are not required here. Let $c' = *f^{-1}(w + \varepsilon)$ so that $c' > c$. If c' were not infinitesimally close to c, then $c' > c + 1/m$ for some $m \in \mathbb{N}$. As $c, m \in \mathbb{R}$. $f(c + 1/m) > f(c) + 1/n$ for some $n \in \mathbb{N}$, so that $*f(c') > f(c + 1/m) > f(c) + 1/n = w + 1/n$, but $*f(c') = w + \varepsilon$, hence $\varepsilon > 1/n$, a contradiction. It follows that $c' \sim c$, which shows the continuity of f^{-1} at w. $\square$

The next subject of this section is the introduction of b^c, for $b, c \in \mathbb{R}$, $b > 0$. If $c \in \mathbb{Q}$ this can best be done in the classical way, using the functions x^c and b^x as well the properties of inverse functions. Hence begin with x^n, $n \in \mathbb{N}$, $x \in \mathbb{R}$, $x > 0$, which is monotonically increasing and continuous, leading to the definition of $x^{1/n}$ as its inverse, and hence to $x^{m/n}$, $m, n \in \mathbb{N}$, $x \in \mathbb{R}$, $x > 0$, either as $(x^{1/n})^m$ or as $(x^m)^{1/n}$. To see that the two are identical, note that $(y^n)^{1/n} = y$, so that, taking $y = (x^{1/n})^m$ it follows that,

$$(((x^{1/n})^m)^n)^{1/n} = (x^{1/n})^m,$$

and taking $y = x$ it follows that,

$$(((x^{1/n})^m)^n)^{1/n} = ((x)^m)^{1/n} = (x^m)^{1/n}.$$

For $c \in \mathbb{Q}$, $c > 0$, let $x^c = 1/x^{-c}$, and in view of $x^c \cdot x^d = x^{c+d}$, let $x^0 = 1$. Then x^c, $c \in \mathbb{Q}$, $x \in \mathbb{R}$, $x > 0$ is increasing if $c > 0$, decreasing if $c > 0$, and equal to 1 if $c = 0$.

Next consider b^x, $b \in \mathbb{R}$, $b > 0$, $x \in \mathbb{Q}$, which is now well defined, then b^x is increasing if $b > 1$, decreasing if $b < 1$, and equal to 1 if $b = 1$; and continuous at each $x \in \mathbb{Q}$.

For $c \in \mathbb{R}$ there is a nonstandard alternative. Given any $c \in \mathbb{R}$, by Theorem 2.13.2, $c = \text{st}(c')$ for some $c' \in {}^*\mathbb{Q}$, where c' is determined uniquely up to a hyperrational infinitesimal. Now let,

$$g(c) = \text{st}(b^{c'}),$$

then $g(c) = b^c$, where, $b^c = \lim_{x \to c} b^x$, $x \in \mathbb{Q}$.

To see this note first of all that $g(c)$ is defined uniquely, for if $\varepsilon \simeq 0$, $\varepsilon \in {}^*\mathbb{Q}$, then $\mathrm{st}(b^{c'+\varepsilon}) = \mathrm{st}(b^{c'}) \cdot \mathrm{st}(b^{\varepsilon}) = \mathrm{st}(b^{c'}) \cdot 1$.

Now let $d \in \mathbb{Q}$, such that $c = d + r$ for some $r > 0$, then $g(d) = b^d$. Since $c' = c + \varepsilon$ for some $\varepsilon \simeq 0$, $c' = d + r + \varepsilon$ and,

$$g(c) - g(d) = \mathrm{st}(b^{d+r+\varepsilon}) - b^d = b^d \cdot [\mathrm{st}(b^{r+\varepsilon}) - 1].$$

In this product the first factor is positive, and the second one is positive (or negative) if $b > 1$ (or $b < 1$), and zero if $b = 1$. A similar result is obtained if $r < 0$. Hence the function g is monotonous if not constant, so that if c is between the rationals d and e, then $g(c)$ is between b^d and b^e. This shows both the monotonicity and the continuity of $g(x)$, $x \in \mathbb{R}$, and hence that $g(c) = b^c$. $\qquad\qquad\square$

Exercise: Show that $\mathrm{st}(p \cdot q) = \mathrm{st}(p) \cdot \mathrm{st}(q)$ and that $\mathrm{st}(b^{\varepsilon}) = 1$ if $b > 0$ and $\varepsilon \simeq 0$, $\varepsilon \in {}^*\mathbb{Q}$.

3.6   Differentiation

Let $f : {}^*\mathbb{R} \to {}^*\mathbb{R}$, $c \in {}^*\mathbb{R}$, then f is said to be *differentiable at* c if for some $k \in {}^*\mathbb{R}$,

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c} \text{ exists and is equal to } k.$$

Then this limit, which obviously is a *limit, is called the *derivative* of f at c, and usually k is replaced by $f'(c)$ or by $\dfrac{df(c)}{dx}$.

In case everything is standard, this definition becomes the classical definition of differentiability, and from Section 3.3 it follows that $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $c \in \mathbb{R}$ if for some $k \in \mathbb{R}$,

$$\forall \delta \sim 0 : \frac{{}^*f(c + \delta) - f(c)}{\delta} \simeq k = f'(c),$$

so that,

$$f'(c) = \mathrm{st}\left[\frac{{}^*f(c + \delta) - f(c)}{\delta}\right].$$

This means that $f'(c)$ is infinitesimally close to a quotient, justifying to a certain extent calling $f'(c)$ a differential quotient, even though $f'(c)$ is a limit of a quotient.

As before this need not be true if f, and c are internal; then f is called *S-differentiable* at c if for some internal k,

$$\forall \delta \sim 0 : \frac{f(c + \delta) - f(c)}{\delta} \simeq k = f'(c).$$

From the definition of $f'(c)$ in the standard case it follows that,

$$\forall \delta \simeq 0 : {}^*f(c + \delta) - f(c) - f'(c) \cdot \delta = \tau \delta,$$

for some $\tau \simeq 0$. This is known as the *increment theorem*. Conversely, if, for some $k \in \mathbb{R}$ and,

$$\forall \delta \simeq 0 : {}^*f(c + \delta) - f(c) - k\delta = \tau \delta, \text{ for some } \tau \simeq 0,$$

then $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $c \in \mathbb{R}$ and $f'(c) = k$.

Theorem 3.6.1

1)   *If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $c \in \mathbb{R}$, then f is continuous at c.*
1)   *If f is differentiable at c, and $b \in \mathbb{R}$, then $f + b$ and $b \cdot f$ are differentiable at c as well, and,*

$$(f + b)'(c) = f'(c), (b \cdot f)'(c) = b \cdot f'(c).$$

1)   *If both f and g are differentiable at c, then*

$$(f + g)'(c) = f'(c) + g'(c), (f \cdot g)'(c) = f'(c) \cdot g(c) + f(c) \cdot g'(c),$$

and, if $g(c) \neq 0$, then,

$$(f/g)'(c) = (f'(c) \cdot g(c) - f(c) \cdot g'(c))/g^2(c).$$

Note that in the last case that $g(x) \neq 0$ if $x = c + \delta$, for $\delta \simeq 0$, as g is continuous at c. In particular, if $g(x) = 1/f(x)$, and $f(c) \neq 0$, then, $g'(c) = -f'(c)/f^2(c)$.

1)   *Chain rule. If g is differentiable at c, and f is differentiable at $g(c)$, then the composite function $F = f \circ g$, too is differentiable at c, and,*

$$F'(c) = (f \circ g)'(c) = f'(g(c)) \cdot g'(c).$$

Proof:

1)   Follows immediately from ${}^*f(c + \delta) - f(c) - f'(c) \cdot \delta = \tau \delta$, so that,

$${}^*f(c + \delta) - f(c) \simeq 0 \text{ if } \delta \simeq 0.$$

2) and 3) are left as exercises.

4) For any $\delta \simeq 0$, and for some $\mu \simeq 0$,

$$^*f(g(c) + \delta) - f(g(c)) - f'(g(c)) \cdot \delta = \mu\delta$$

Let, given any $\varepsilon \simeq 0$,

$$\delta = {}^*g(c + \varepsilon) - g(c) = g'(c) \cdot \varepsilon + \tau\varepsilon,$$

for some $\tau \simeq 0$. Then $\delta \simeq 0$ and,

$$^*f({}^*g(c + \varepsilon)) - f(g(c)) - f'(g(c)) \cdot (g'(c) + \tau) \cdot \varepsilon = \mu\delta,$$

hence,

$$^*f({}^*g(c + \varepsilon)) - f(g(c)) - f'(g(c)) \cdot g'(c) \cdot \varepsilon =$$
$$f'(g(c)) \cdot \tau\varepsilon + \mu(g'(c) \cdot \varepsilon + \tau\varepsilon) = \tau'\varepsilon,$$

for some $\tau' \simeq 0$, which means that $f \circ g$ is differentiable at c, and that its derivative at c is equal to $f'(g(c)) \cdot g'(c)$. $\qquad\square$

Theorem 3.6.2 *(The critical point theorem.)*
Let X be an interval of $\mathbb{R}$, $c \in X$, $f : X \to \mathbb{R}$ be continuous at each $x \in X$, and c be a maximum or a minimum of f over X. Then c is an endpoint of X, or $f'(c)$ is undefined, or $f'(c) = 0$.

Proof: Let c be a maximum of f over X. If c is not an endpoint of X and if $f'(c)$ exists, then,

$$\forall \delta \sim 0 : f'(c) = \mathrm{st}\left[\frac{{}^*f(c + \delta) - f(c)}{\delta}\right].$$

Taking first δ positive and then negative this gives that $f'(c) \leq 0$ and $f'(c) \geq 0$, hence $f'(c) = 0$. $\qquad\square$

Theorem 3.6.3 *(Rolle's theorem.)*
Let f be continuous in $[a, b]$, $a, b \in \mathbb{R}$, $a < b$, and be differentiable in (a, b). Moreover, let $f(a) = f(b) = 0$. Then $f'(c) = 0$ for some $c \in (a, b)$.

Proof: The proof is classical in nature, and relies on Theorems 3.2.8 and 3.6.2. $\qquad\square$

The proofs of the next corollaries too are classical in nature, and therefore are not given in detail.

Corollary 3.6.1 *(Mean value theorem.)*
Let f be as in the previous theorem, except $f(a) = f(b) = 0$ need not be true.
Then,

$$f'(c) = \frac{f(b) - f(a)}{b - a} \text{ for some } c \in (a, b).$$

Proof: The proof follows by applying Rolle's theorem to g, defined by,

$$g(x) = f(x) - f(a) \cdot \frac{f(b) - f(a)}{b - a} \cdot (x - a).$$

$\square$

Corollary 3.6.2 *(Generalized mean value theorem.)*
Let f and g both be continuous in $[a, b]$, and be differentiable in (a, b). In addition
let $g'(x) \neq 0$ for all $x \in (a, b)$. Then,

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)} \text{ for some } c \in (a, b).$$

Proof: The proof follows by applying Rolle's theorem to h, defined by,

$$h(x) = f(x)(g(b) - g(a)) - g(x)(f(b) - f(a)).$$

$\square$

Corollary 3.6.3 *(Taylor's theorem.)*
Let f and f' both be continuous in $[a, b]$, and let f'' exist in (a, b), then,

$$f(b) = f(a) + f'(a)(b - a)/1! + f''(c)(b - a)^2/2!, \text{ for some } c \in (a, b),$$

and similarly for higher derivatives.

Proof: The proof is based on the previous corollary. $\square$

Theorem 3.6.4 *(L'Hopital's theorem).*
Let both f and g be differentiable, and $g'(x) \neq 0$ in a neighborhood (a, b) of

$c \in \mathbb{R}$ *with the possible exception of the point c itself. Also assume that* $\lim\limits_{x \to c} f(x) = \lim\limits_{x \to c} g(x) = 0$, *and that* $\lim\limits_{x \to c} \dfrac{f'(x)}{g'(x)}$ *exists and is equal to k. Then,*

$$\lim_{x \to c} \frac{f(x)}{g(x)} = k.$$

Similar statements hold true if k and/or c are replaced by ∞.

Proof: Only the case as stated will be proved.

Without loss of generality it may be assumed that $f(x)$ and $g(x)$ are defined at $x = c$ and that $f(x) = g(x) = 0$, so that f and g are continuous at c. It follows that the generalized mean value theorem can be applied to both $[c, c + \delta]$ and $[c - \delta, c]$, for $\delta > 0$ and close enough to zero. Taking $\delta \sim 0$ this gives that, for some c_1 and c_2, $c < c_1 < c + \delta$, $c - \delta < c_2 < c$,

$$\frac{f'(c_1)}{g'(c_1)} = \frac{f(c + \delta) - f(c)}{g(c + \delta) - g(c)} \text{ and } \frac{f'(c_2)}{g'(c_2)} = \frac{f(c) - f(c - \delta)}{g(c) - g(c - \delta)}.$$

Since both c_1 and c_2 are infinitesimally close to c this implies that these two expressions are infinitesimally close to k and the theorem follows. $\qquad\square$

3.7 Integration

Only Riemann integration of continuous functions will be considered, even though the extension to more general functions is not very difficult.

Let $a, b \in {}^*\mathbb{R}$, $a < b$ and let $f : [a, b] \to {}^*\mathbb{R}$ be a continuous function. Then, by definition, the *Riemann integral* of f over $[a, b]$ is,

$$J = \int_a^b f(x)dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(a + i(b - a)/n)(b - a)/n.$$

In case everything is standard, then, given any $\omega \in {}^*\mathbb{N}$, $\omega \sim \infty$,

$$J = \text{st}\left[\sum_{i=1}^{\omega} f(a + i(b - a)/\omega)(b - a)/\omega \right].$$

Note that in this case $dx = (b - a)/\omega \sim 0$ and $dx > 0$. And the usual formulation is,

$$J = \text{st}\left[\sum_{i=1}^{\omega} f(a + i \cdot dx)/dx \right].$$

Instead of the point $a + i \cdot dx$ any other point $x_i \in [a + (i-1)dx, a + i \cdot dx]$ may be selected, $i = 1, \ldots, \omega$ and J becomes $\mathrm{st}(S)$, where,

$$S = \sum_{i=1}^{\omega} f(x_i) dx,$$

a so-called *Riemann sum*. That S is finite follows from the extreme value theorem (Theorem 3.2.8), as this theorem implies the existence of $m, M \in \mathbb{R}$ such that,

$$\forall x \in [a, b] : m \leq f(x) \leq M,$$

so that $m(b-a) \leq S \leq M(b-a)$ and S is indeed finite.

The notation of the integral correctly suggests that J depends neither on ω, nor on how the x_i are selected. Indeed, let y_i some other element of the interval $[a + (i-1)dx, a + i \cdot dx]$, then if ω is not changed it follows from the uniform continuity of f and $dx < \delta$ for all $\delta \in \mathbb{R}$, $\delta > 0$, that,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \left| \sum_{i=1}^{\omega} f(x_i) dx - \sum_{i=1}^{\omega} f(y_i) dx \right| < \omega \cdot \varepsilon \cdot dx = \varepsilon(b-a),$$

hence that the expression to the left of the last inequality is an infinitesimal, which shows that J does not depend on the selection of the x_i. And if $\omega_1, \omega_2 \sim \infty$, and for $k = 1, 2$, $dx_k = (b-a)/\omega_k$ and,

$$S_k = \sum_{i=1}^{\omega_k} f(a + i \cdot dx_k) dx_k,$$

let $\omega = \omega_1 \omega_2$ and let dx and S be as before, then,

$$S - S_1 = \sum_{i=1}^{\omega_1} \sum_{j=1}^{\omega_2} [f(a + \{(i-1)\omega_2 + j\}dx)dx - f(a + i\omega_2 dx)dx],$$

and again this is an infinitesimal, as is $S - S_2$, and hence so is $S_1 - S_2$.

In case $a > b$, then by definition $\int_a^b f(x)dx = -\int_b^a f(x)dx$, and if $a = b$, then $\int_a^a f(x)dx = 0$.

As remarked before f need not be continuous in order for J to exist. Also in what follows derivatives need not always be continuous, but let us not go into the details as this chapter is only concerned with showing the possibilities of nonstandard analysis.

The definition of improper integrals is as in classical analysis: just consider the appropriate limits.

Exercise: Present the details of introducing improper integrals, complete with the simplified forms in case everything is standard.

In the remainder of this section it is assumed that everything is standard.

Theorem 3.7.1 *If $a < b < c$, then,*

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx.$$

Proof: Follows from the fact that the standard part of a sum is equal to the sum of the standard parts. □

Theorem 3.7.2 *F defined by $F(x) = \int_a^x f(t)dt$, $a \le x \le b$, $a < b$, is continuous. Moreover, the derivative of $F(x)$ exists and is equal to $f(x)$, $a < x < b$. Conversely, if G is such that $G'(x) = f(x)$, then,*

$$F(b) = \int_a^b f(x)dx = G(b) - G(a).$$

Proof: The continuity of F follows from the fact that, if $\delta \simeq 0$,

$$\int_a^{x+\delta} f(t)dt - \int_a^x f(t)dt = \int_x^{x+\delta} f(t)dt,$$

no matter whether δ is nonnegative or negative, in which case the right-hand side too is an infinitesimal.

The second part of the theorem follows from the fact that m and M exist as before such that,

$$m \cdot dx \le F(x + dx) - F(x) = \int_x^{x+dx} f(t)dt \le M \cdot dx.$$

If $dx \sim 0$, $dx > 0$, m and M can be taken infinitesimally small, so that for some $\delta \simeq 0$,

$$F(x + dx) - F(x) = f(x)dx + \delta \cdot dx,$$

which gives the desired result.

To show the last part of the theorem, note that for some constant c, $F(x) = G(x) + c$, $a \le x \le b$, and that $F(a) = 0$. □

The proofs of the next two theorems are not particularly of a nonstandard nature, and for that reason are kept rather short.

Theorem 3.7.3 *(Substitution rule.)*
Let $F'(x) = f(x)$ for $x \in [a, b]$, and let $x = g(w)$ for $w \in [\alpha, \beta]$ such that g maps $[\alpha, \beta]$ onto $[a, b]$, with $g(\alpha) = a$, $g(\beta) = b$. Also assume that g is continuously differentiable. Then,

$$\int_a^b f(x)dx = \int_\alpha^\beta f(g(w))g'(w)dw.$$

Proof: From Theorem 3.6.1 it follows that,

$$(F(g(w)))' = F'(g(w))g'(w) = f(g(w))g'(w),$$

hence,

$$\int_a^b f(x)dx = F(b) - F(a) = F(g(\beta)) - F(g(\alpha)) = \int_\alpha^\beta (F(g(w)))'dw =$$
$$\int_\alpha^\beta f(g(w))g'(w)dw.$$

$\square$

Theorem 3.7.4 *(Integration by parts.)*
If both f and g are continuously differentiable in $[a, b]$, then,

$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx.$$

Proof: Since $(f(x)g(x))' = f(x)g'(x) + f'(x)g(x)$ it follows that,

$$J = \int_a^b (f(x)g(x))'dx,$$

exists and is equal to the sum of the two integrals in the statement of the theorem. But $J = f(b)g(b) - f(a)g(a)$. $\square$

3.8 Pitfalls in nonstandard analysis

In this section it is shown by means of a number of examples that some care is required when applying nonstandard analysis.

A)   The existence of a positive infinitesimal ε means that,

$$\exists \varepsilon \in {}^*\mathbb{R} : \forall m \in \mathbb{N} : 0 < \varepsilon < 1/m.$$

By transfer (?) this would give,

$$?\ \exists \varepsilon \in \mathbb{R} : \forall m \in \mathbb{N} : 0 < \varepsilon < 1/m\ ?,$$

which obviously is not true. The cause of the trouble is that in the first statement the constant $\mathbb{N}$ is external, so that transfer is not allowed.

B)   Any nonempty subset X of $\mathbb{N}$ has a smallest element, that is,

$$\exists X \subseteq \mathbb{N}, X \neq \emptyset : \exists x \in X : \forall y \in X : x \leq y,$$

which by transfer (?) would lead to,

$$?\ \forall X \subseteq {}^*\mathbb{N}, X \neq \emptyset : \exists x \in X : \forall y \in X : x \leq y\ ?,$$

which is wrong as can be seen be seen by taking, for example, $X = {}^*\mathbb{N} - \mathbb{N}$ (which is external), for if x were the smallest element of X, then also $x - 1 \in X$, so $x - 1 \leq x$. The correct procedure would be:

$$\forall X \in \mathcal{P}(\mathbb{N}), X \neq \emptyset : \exists x \in X : \forall y \in X : x \leq y,$$

which by transfer gives,

$$\forall X \in {}^*(\mathcal{P}(\mathbb{N})), X \neq \emptyset : \exists x \in X : \forall y \in X : x \leq y,$$

so that X must be internal. The latter is indeed true as could be shown by returning to first principles, i.e. to write ${}^*(\mathcal{P}(\mathbb{N}))$ as $\{H(X_1, X_2, \ldots) : X_i \subseteq \mathbb{N}\}$, so that, if x_i is the smallest element of X_i, $H(x_1, x_2, \ldots)$ is the smallest element of $H(X_1, X_2, \ldots)$.

C)   As is well-known $\mathbb{R}$ has *Archimedian order*, that is, given any $x \in \mathbb{R}$ there is an $n \in \mathbb{N}$ such that $n > x$, or,

$$\forall x \in \mathbb{R} : \exists n \in \mathbb{N} : n > x.$$

But ${}^*\mathbb{R}$ has no such order,

$$?\ \forall x \in {}^*\mathbb{R} : \exists n \in \mathbb{N} : n > x\ ?,$$

for take any $x \sim \infty$. Indeed, transfer would give,

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

which is correct. One might say that ${}^*\mathbb{R}$ has *hyper-Archimedean order*.

D) In Section 1.4 it was shown by means of transfer that statements (1.2) and (1.3) are equivalent:

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta : \mid {}^*f(x) - {}^* f(c) \mid < \varepsilon$$

and,

$$\forall \delta \in {}^*\mathbb{R}, \delta \simeq 0 : {}^*f(c + \delta) - {}^*f(c) \simeq 0, \text{ where } c \in \mathbb{R} \text{ and } f : \mathbb{R} \to \mathbb{R}.$$

Note that ${}^*c = c$. Replacing c by a nonstandard constant or replacing ${}^*f$ by a nonstandard, but internal, function, this equivalence may be destroyed. Examples were already given in Section 3.3.

E) Let S be a bounded subset of $\mathbb{R}$, then S has a least upper bound in $\mathbb{R}$. Hence, by transfer (?), the set of all infinitesimals in ${}^*\mathbb{R}$ would have a least upper bound β in ${}^*\mathbb{R}$, which it has not, because β and hence 2β would have to be infinitesimals themselves, but $2\beta > \beta$. Transfer is illegal here because, by Theorem 2.10.4, the infinitesimals form an external set.

126

Chapter 4

Some special topics

4.1 Principles of permanence

In the proof of Theorem 3.3.2 the fact was used that an external set is not internal, and in the remark below that proof that this fact, which is called *Cauchy's principle*, is a *principle of permanence*. In general the latter is the statement that if a certain property P holds for all elements of a certain set A, it must hold for at least one element not in A:

$$[\forall a \in A : P(a)] \rightarrow [\exists b \notin A : P(b)].$$

The statement is based on the fact that an incompatibility between the property and the set would exist in case the property would only hold for the elements of the set. For example, let the set be $\mathbb{N}$ and let the property be being an element of some given internal subset S of $^*\mathbb{N}$:

$$[\forall a \in \mathbb{N} : a \in S] \Rightarrow [\exists b \notin \mathbb{N} : b \in S].$$

If S would only contain the elements of $\mathbb{N}$ there would be an incompatibility as S is internal and $\mathbb{N}$ is external. In other words S must contain some $\omega \sim \infty$. The aspect of permanence here lies in the fact that belonging to S is necessarily carried over from the classical natural numbers to certain hyperlarge numbers. From this point of view Cauchy's principle would not seem to be a very explicit example of a principle of permanence, and indeed some authors restrict the term 'principle of permanence' to cases where if something is true for all elements of some set, it must be true for some element outside that set. On the other hand all principles of permanence different from Cauchy's principle can be based on the latter (even Fehrele's principle to be discussed below).

It would be wrong to conclude that in the example S would contain all $\omega \sim \infty$, since taking $\omega_o \sim \infty$ arbitrarily, S defined by,

$$S = \{n : n \in {}^*\mathbb{N}, n \leq \omega_o\},$$

is internal. The latter is easily shown by returning to the basic theory, because, letting $\omega_o = H(n_{oi})$,

$$S = \{H(n_i) : n_i \leq n_{oi}\} = \{H(n_i) : n_i \in S_i\}, \text{ where } S_i = \{n : 1 \leq n \leq n_{oi}\}.$$

Note that the example is a special case of overflow (see Theorem 2.11.1), and that overflow, as well as underflow, may be seen as principles of permanence.

Theorem 4.1.1 *If $f : X \rightarrow {}^*\mathbb{R}$ is an internal function such that $\forall x \in X : f(x) \simeq 0$, then $\sup_{x \in X} | f(x) | \simeq 0$.*

Proof 1: Since $f(x)$ is bounded over X, the supremum exists. Denote it by β. Then $\forall \delta > 0 : \exists x \in X : | f(x) | \geq \beta - \delta$, that is $\beta \leq \delta + | f(x) |$. Let $\delta \sim 0$, then it follows that $\beta \simeq 0$. $\square$

Proof 2: Let $I = \{\beta \in {}^*\mathbb{R} : [\forall x \in X : | f(x) | \leq \beta]\}$. Then I is internal, but it contains the external subset $\{\beta \in {}^*\mathbb{R} : \beta > 0, \beta \text{ is not an infinitesimal}\}$, hence, by Cauchy's principle, it must contain some $\beta \sim 0$, from which the result follows. $\square$

Note that the first proof, which is classical in nature, is to be preferred. Nevertheless, the second proof is a good illustration of applying Cauchy's principle.

Corollary 4.1.1 *Let $[a, b]$ be some interval of ${}^*\mathbb{R}$, such that $b - a$ is limited, and let f and g both be Riemann integrable functions over $[a, b]$ such that $f(x) \simeq g(x)$ for all $x \in [a, b]$. Then,*

$$\int_a^b f(x)dx \simeq \int_a^b g(x)dx.$$

Proof: Let $\beta = \sup_{a \leq x \leq b} | f(x) - g(x) |$. According to the theorem $\beta \simeq 0$, hence,

$$0 \leq \left| \int_a^b f(x)dx - \int_a^b g(x)dx \right| \leq \int_a^b | f(x) - g(x) | \, dx \leq \int_a^b \beta \cdot dx = \beta(b - a),$$

which is an infinitesimal. $\square$

In Cauchy's principle an external set is 'confronted' with an internal set, but there is another principle of permanence (in the more general sense of the term) where two external sets that are of different kinds are 'confronted' with each other. A typical example of an external set of the first kind is the set of all infinitesimals in ${}^*\mathbb{R}$, and a typical example of an external set of the second kind is the set of

all limited numbers in $^*\mathbb{R}$. Obviously, one must find a general rule from which it follows that these two sets are indeed of a different kind. An external set is of the first kind if it is a *halo*, and it is of the second kind if it is a *galaxy*. The notions 'halo' and 'galaxy' are defined below, and the second principle of permanence is *Fehrele's principle*:

No halo is a galaxy, hence no galaxy is a halo.

A set H is called a halo if there exists an $\omega \in {^*\mathbb{N}}$, $\omega \sim \infty$, such that,

1) $H = \cap_{n \in \mathbb{N}} S(n)$, where $(S(n))$, $n \in [1, \omega])$ is a hyperfinite internal sequence of internal subsets $S(n)$ of some given standard set (such as $^*\mathbb{N}$ or $^*\mathbb{R}$ or something totally different), and,

2) H is external.

The second requirement is not superfluous, for take all $S(n)$ equal to some fixed internal set.

Obviously, the sequence involved may be an infinite sequence, because for any $\omega \sim \infty$ it contains a hyperfinite sequence with $[1, \omega]$ as its domain. It is also allowed to define $S(n)$ for $n \in \mathbb{N}$ only, even though the sequence would then be external. For, with the operator H as in the basic theory, let $S(n) = H(S_i(n))$, $n \in \mathbb{N}$ and define $T(n)$ for $n \in {^*\mathbb{N}}$ by, $T(H(n_i)) = H(S_i(n_i))$, so that $T = H(S_i)$ and T is internal. Moreover, $T(n) = H(S_i(n)) = S(n)$, $n \in \mathbb{N}$, hence T extends S as a function with domain $\mathbb{N}$ to a function with domain $^*\mathbb{N}$.

It is no restriction to assume that the sequence $(S(n), n \in [1, \omega])$ is nonincreasing, i.e. that $S(1) \supseteq S(2) \supseteq \ldots \supseteq S(\omega)$, for if this is not the case, let,

$$S'(n) = \cap_{k \leq n} S(k), n \in [1, \omega],$$

then,

$(S'(n), n \in [1, \omega])$ is nonincreasing, and $H = \cap_{n \in \mathbb{N}} S'(n)$.

It may even be assumed that the sequence is strictly decreasing, i.e. that $S(1) \supset S(2) \supset \ldots \supset S(\omega)$ (perhaps for some other ω). For let,

$$K = \{n \in [1, \omega] : [\exists p(n) > n : S(n) \supset S(p(n))]\},$$

then $\mathbb{N} \subseteq K$, because otherwise $\exists n' \in \mathbb{N} : \forall p > n' : S(n') = S(p)$, hence $H = S(n')$, but H is external and $S(n')$ is internal. Moreover, by the internal definition principle, K is internal, hence $\exists \omega' \sim \infty : \omega' \in K$, i.e. $[1, \omega'] \subseteq K$. Now let $m_1 = 1$, $m_2 = p(m_1), \ldots$, and $S''(n) = S(m_n)$, $n \in [1, \omega']$, then,

$(S''(n), n \in [1, \omega'])$ is strictly decreasing, and $H = \cap_{n \in \mathbb{N}} S''(n)$.

Conversely, if $H = \cap_{n \in \mathbb{N}} S(n)$ and $(S(n), n \in [1, \omega])$, where $\omega \sim \infty$, is strictly decreasing, then H is automatically external, for let $M = \{n \in {}^*\mathbb{N} : H \subset S(n)\}$. If H were internal, then, by the internal definition principle, M would be internal as well, but $M = \mathbb{N}$ and $\mathbb{N}$ is external. To see that $M = \mathbb{N}$ observe that $\mathbb{N} \subseteq M$ and suppose that $H \subset S(\omega)$ for some $\omega \sim \infty$, then $H \subset S(\omega) \subset S(n)$ for all $n \in \mathbb{N}$, so that $H \subset S(\omega) \subseteq \cap_{n \in \mathbb{N}} S(n) = H$, a contradiction. This proves the next theorem.

Theorem 4.1.2 *H is a halo if and only if $H = \cap_{n \in \mathbb{N}} S(n)$, where $(S(n), n \in [1, \omega])$ for some $\omega \sim \infty$ is a strictly decreasing internal sequence of internal sets $S(n)$.* $\qquad\square$

A set G is called a *galaxy* if there exists an $\omega \in {}^*\mathbb{N}$, $\omega \sim \infty$, such that,

1) $G = \cup_{n \in \mathbb{N}} T(n)$, where $(T(n), n \in [1, \omega])$ is a hyperfinite internal sequence of internal subsets $T(n)$ of some given standard set, and,
2) G is external.

As before the sequence may be an infinite internal sequence or be defined for $n \in \mathbb{N}$ only.

Also the sequence may assumed to be nondecreasing and even strictly increasing, as can be seen by an argument similar to the one leading to the preceding theorem.

Theorem 4.1.3 *G is a galaxy if and only if $G = \cup_{n \in \mathbb{N}} T(n)$, where $(T(n), n \in [1, \omega])$ for some $\omega \sim \infty$ is a strictly increasing internal sequence of internal sets $T(n)$.* $\qquad\square$

Remarks:

1. The given standard set is arbitrary, and hence may be some abstract set. This shows the generality of the two definitions, where numbers only play a part in the definitions of the two notions, and even this can be weakened, as can be seen from the next remark.
2. The definitions can be generalized by replacing ${}^*\mathbb{N}$ by some standard index set ${}^*A$, and letting S and T be internal functions mapping the a ${}^*A$ to internal sets $S(a)$ and $T(a)$. Then $H = \cap_{a \in A} S(a)$ and $G = \cup_{a \in A} T(a)$.

Theorem 4.1.4 *(Fehrele's principle.) No halo is a galaxy, hence no galaxy is a halo.*

Proof: (Van den Berg.) Assume to the contrary that some halo H is equal to some galaxy G. Let $H = \cap_{n \in \mathbb{N}} S(n)$ and $G = \cup_{n \in \mathbb{N}} T(n)$, with $S(n)$ and $T(n)$ internal, $(S(n))$ nonincreasing and $(T(n))$ nondecreasing. Then $T(n) \subseteq S(n)$ for all $n \in \mathbb{N}$. Let $I = \{n \in {}^*\mathbb{N} : T(n) \subseteq S(n)\}$. Then I is internal, as follows from the internal definition principle. Also $\mathbb{N} \subseteq I$, so that, since $\mathbb{N}$ is external, $\omega \in I$ for some $\omega \sim \infty$, so that $S(n) \supseteq S(\omega) \supseteq T(\omega) \supseteq T(n)$ for all $n \in \mathbb{N}$, hence $H \supseteq S(\omega) \supseteq T(\omega) \supseteq G$, or $H = S(\omega) = T(\omega) = G$, so that $H = G$ would be internal. $\qquad \square$

Exercise: Show that the subset G of an internal set S is a galaxy if and only if $S - G$ is a halo.

Corollary 4.1.2 *(Robinson's lemma.)*
Let $(s(n), n \in {}^*\mathbb{N})$, $s(n) \in {}^*\mathbb{R}$, be an internal sequence, such that $s(n) \simeq 0$ for all $n \in \mathbb{N}$. Then,

$$\exists \omega \in {}^*\mathbb{N}, \omega \sim \infty : \forall k \in {}^*\mathbb{N}, k \leq \omega : s(k) \simeq 0.$$

Proof 1: (Van den Berg.) Let $H = \{n \in {}^*\mathbb{N} : [\forall k \leq n : s(k) \simeq 0]\}$ and $G = \mathbb{N}$. Then G is a galaxy and $H \supseteq \mathbb{N} = G$. If H is external, then $G \subset H$ (by Fehrele's principle), and if H is internal then trivially $G \subset H$, hence $G \subset H$ anyway. Let $\omega \in H - G$, then $\omega \sim \infty$, and $s(k) \simeq 0$ for all $k \leq \omega$. $\qquad \square$

Proof 2: (Robinson; in time preceding the first proof, and using Cauchy's principle.) Let $S = \{n \in {}^*\mathbb{N} : [\forall k \leq n :| s(k) | \leq 1/k]\}$, then S is internal and $S \supseteq \mathbb{N}$, hence (by Cauchy's principle) $S \supset \mathbb{N}$, so that,

$$\exists \omega \in {}^*\mathbb{N}, \omega \sim \infty : \forall k \leq \omega :| s(k) | \leq 1/k.$$

Let $k \leq \omega$. If $k \sim \infty$, then $s(k) \simeq 0$, as then $1/k \simeq 0$, and if $k \in \mathbb{N}$, then by assumption $s(k) \simeq 0$. $\qquad \square$

Corollary 4.1.3 *(Dominated approximation.)*
Let f, g and h be functions from ${}^*\mathbb{R}$ to ${}^*\mathbb{R}$, Riemann integrable over $(-\infty, +\infty)$. Let f and g be internal but h be standard. Assume that $f(x) \simeq g(x)$ for all limited x, and that $| f(x) |, | g(x) | \leq h(x)$ for all $x \in {}^*\mathbb{R}$. Then,

$$\int_{-\infty}^{+\infty} f(x)dx \simeq \int_{-\infty}^{+\infty} g(x)dx.$$

Proof: Given any $n \in \mathbb{N}$, let $\beta = \sup_{|x| \leq n} | f(x) - g(x) |$, then $\beta \simeq 0$, as follows from Theorem 4.1.1, and Corollary 4.1.1 implies that,

$$\forall n \in \mathbb{N} : \int_{-n}^{+n} f(x)dx \simeq \int_{-n}^{+n} g(x)dx,$$

so that, by Robinson's lemma, there exists an $\omega \sim \infty$ such that,

$$\int_{-\omega}^{+\omega} f(x)dx \simeq \int_{-\omega}^{+\omega} g(x)dx.$$

Since, as h is standard, $\int_{|x| \geq \omega} h(x)dx \simeq 0$, it follows that,

$$\int_{|x| \geq \omega} f(x)dx \simeq 0 \text{ and } \int_{-\omega}^{+\omega} f(x)dx \simeq \int_{-\infty}^{+\infty} f(x)dx,$$

and similarly for g instead of f, from which the desired result follows. $\square$

Exercise: Show that if $(s(n), n \in {}^*\mathbb{N})$, $s(n) \in {}^*\mathbb{R}$, is an internal sequence such that $\forall n \in \mathbb{N} : s(n) |< 1/n$, then $\exists \omega \sim \infty : \forall n \in [1, \omega] : s(n) |< 1/n$.

4.2 The saturation principle

The saturation principle is concerned with infinite sequences of internal sets and does not hold in classical mathematics. An infinite sequence $(S(n), n \in \mathbb{N})$ of sets – internal or not – has the finite intersection property if $\cap_{k=1}^n S(k) \neq \emptyset$ for all $n \in \mathbb{N}$.

Theorem 4.2.1 *(Saturation.)*
Let the infinite sequence $(S(n), n \in \mathbb{N})$ of internal sets $S(n)$ have the finite intersection property, then the intersection of all of them is nonempty, i.e. $\cap_{n \in \mathbb{N}} S(n) \neq \emptyset$.

Proof 1: (Not using permanence.) Let $S(n) = H(S_i(n))$, where H is the H-operator of the basic theory. For all $n \in \mathbb{N}$, let,

$$T(n) = \cap_{k=1}^n S(k), \text{ and } T_i(n) = \cap_{k=1}^n S_i(k),$$

then for $n \geq 2$, $T(1) \supseteq T(2) \supseteq \ldots \supseteq T(n)$, hence,

$$\{i : T_i(1) \supseteq T_i(2) \supseteq \ldots \supseteq T_i(n)\} \in U,$$

where U is the basic free ultrafilter. Also $\{i : T_i(n) \neq \emptyset \in U\}$, and since $i : i \leq n\}$ is a finite set,

$$Q_n = \{i : i \geq n, T_i(1) \supseteq T_i(2) \supseteq \ldots \supseteq T_i(n), T_i(n) \neq \emptyset\} \in U.$$

Obviously, $Q_n \supseteq Q_{n+1}$. For $i \in Q_2$, $i \geq 2$, let n_i be the maximal $n \geq 2$ such that,

$$T_i(1) \supseteq T_i(2) \supseteq \ldots \supseteq T_i(n), \ T_i(n) \neq \emptyset, \ \text{and} \ n \leq i.$$

This n_i is well-defined. Then $\{n_2, n_3, \ldots\}$ is not bounded, because if it were, so that $n_i \leq m$ for some $m \in \mathbb{N}$, then $Q_{m+1} = \emptyset$, but $Q_{m+1} \in U$. Now for each $i \in Q_2$, $i \geq 2$ take $s_i \in T_i(n_i)$ and take s_i arbitrarily otherwise, then for each $n \in \mathbb{N}$, $H(s_i) \in S(n)$, because for each $n \in \mathbb{N}$ there is an $n_i \geq n$, so that as $s_i \in T_i(n_i)$, also $s_i \in T_i(n) \subseteq S_i(n)$, and $H(s_i) \in S(n)$, as $Q_2 \in U$. $\qquad\square$

Proof 2: (Using Cauchy's principle, and more elegant.) First extend the given sequence to an internal sequence as indicated below the definition of halo (this is not necessary in the first proof). Now let $Q = \{n \in {}^*\mathbb{N} : \cap_{k=1}^n S(k) \neq \emptyset\}$, then Q is internal and $Q \supseteq \mathbb{N}$, hence $Q \supset \mathbb{N}$ and,

$$\exists \omega \sim \infty : \cap_{k=1}^\omega S(k) \neq \emptyset, \ \text{so that certainly} \ \cap_{k \in \mathbb{N}} S(k) \neq \emptyset.$$

$\qquad\square$

Note that the second proof leads to a more general result, which in fact is a principle of permanence.

In classical mathematics a counterexample to the theorem is, for example, the sequence $(S(n))$ where $S(n) = \{n, n+1, \ldots\}$, $n \in \mathbb{N}$.

Corollary 4.2.1 *Let A be a given internal set, and $(S(n))$ an infinite sequence of internal subsets of A. If for all $n \in \mathbb{N}$, $\cup_{k=1}^n S(k) \neq A$, then $\cup_{k \in \mathbb{N}} S(k) \neq A$. Hence if the union of any finite number of $S(n)$ does not fill up A, then the union of all of them does not fill up A.*

Proof: The proof follows from the fact that $(\cup S(n))^c = \cap S^c(n)$, where c denotes complementation with respect to A. $\qquad\square$

Corollary 4.2.2 *Given an infinite sequence $(S(n))$ of internal sets, then $S = \cup_{k \in \mathbb{N}} S(k)$ is internal if and only if there exists an $n \in \mathbb{N}$ such that $S = \cup_{k=1}^n S(k)$.*

Proof: The if-part follows immediately. Conversely, if S is internal, then so are all $T(n) = S - S(n)$, and $\cap_{k \in \mathbb{N}} T(k) = \emptyset$, hence there must exist an $n \in \mathbb{N}$ such that $\cap_{k=1}^n T(k) = \emptyset$, which means that $S = \cup_{k=1}^n S(k)$. $\qquad\square$

Corollary 4.2.3 *Given an infinite sequence $(S(n))$ of internals sets, then $S = \cap_{k \in \mathbf{N}} S(k)$ is internal if and only if there exists an $n \in \mathbf{N}$ such that $S = \cap_{k=1}^{n} S(k)$.*

Proof: By complementation from Corollary 4.2.1. □

4.3 Stirling's formula

In order to provide still more evidence that nonstandard mathematics can be a very elegant substitute for classical mathematics in this section Stirling's formula for large factorials will be derived by nonstandard means. The argument closely follows that given in Van den Berg and Sari [27]. It takes the definition and properties of e as the base of the natural logarithm for granted, as well as those of π as the area of the unit circle, and that,

$$\int_{-\infty}^{+\infty} \exp(-x^2)dx = \sqrt{\pi}.$$

By definition,

$$\Gamma(x) = \int_{0}^{\infty} e^{-t}t^{x-1}dt, \ x \in {}^{*}\mathbf{R}, x > 0.$$

Also the existence of this integral is taken for granted. Let ω be any positive hyperlarge element of ${}^{*}\mathbf{R}$, so that,

$$\Gamma(\omega + 1) = \int_{0}^{\infty} e^{-t}t^{\omega}dt.$$

The integrand is increasing in the interval $[0, \omega]$ and decreasing in the interval $[\omega, \infty)$, for which reason the variable t is replaced by $u = (t - \omega)/\omega$, giving,

$$\Gamma(\omega + 1)\omega^{-\omega-1}e^{\omega} = \int_{-1}^{\infty} e^{-\omega u + \omega \log(1+u)}du,$$

so that the integrand now reaches its maximum at $u = 0$. It so happens that there exists a positive infinitesimal δ such that the contributions of the integrand over the intervals $[-1, -\delta]$ and $[+\delta, \infty)$ may be ignored, so that only the interval $[-\delta, +\delta]$ need be taken into account. In other words, the 'mass' of the integrand is almost entirely concentrated in a hypersmall interval around zero. Instead of the infinitesimal δ consider for the time being any $d \in {}^{*}\mathbf{R}$, $0 < d < 1$, split $[-1, \infty)$ into $[-1, -d]$, $[-d, +d]$ and $[+d, \infty)$ and indicate the integrals of $e^{-\omega u + \omega \log(1+u)}$ over these subintervals by,

$$\int_{-1}^{-d}, \int_{-d}^{+d}, \int_{+d}^{\infty}, \text{ respectively.}$$

1) $[+d, \infty)$. Since the second derivative of $u - \log(1 + u)$ is positive, it follows from Taylor's theorem that,

$$u - \log(1 + u) > d - \log(1 + d) + (u - d)d/(1 + d),$$

giving, after replacing in the integrand the left-hand side of this inequality by its right-hand side, and evaluating the resulting integral, that,

$$0 < \int_{+d}^{\infty} < e^{-\omega d + \omega \log(1+d)}(1 + d)/(\omega d),$$

if – in view of the denominator $\omega d - d$ is *not* an infinitesimal. Since $-d + \log(1 + d) < 0$ it then further follows that,

$$\forall m \in \mathbb{N} : 0 < \int_{+d}^{\infty} < \omega^{-m}.$$

Let $G = \{d \in {}^*\mathbb{R} : 0 < d < 1, d \text{ is not an infinitesimal}\}$, then since the positive infinitesimals form a halo, G is a galaxy. But,

$$H = \cap_{m \in \mathbb{N}}\{d \in {}^*\mathbb{R} : 0 < d < 1, 0 < \int_{+d}^{\infty} < \omega^{-m},$$

is a halo that clearly contains G, hence, by Fehrele's principle, H must contain a positive infinitesimal δ', such that,

$$\forall m \in \mathbb{N} : 0 < \int_{+\delta'}^{\infty} < \omega^{-m}.$$

Obviously, δ' may be replaced by a larger infinitesimal.

2) $[-1, -d]$. Now,

$$u - \log(1 + u) > -d - \log(1 - d) - (u + d)d/(1 - d),$$

and since $d + \log(1 - d) < 0$ it follows similarly that,

$$\forall m \in \mathbb{N} : 0 < \int_{-1}^{-d} < \omega^{-m},$$

if again d is *not* an infinitesimal, but nevertheless there must be a positive infinitesimal δ'', that may be replaced by a larger one, such that,

$$\forall m \in \mathbb{N} : 0 < \int_{-\delta''}^{-\delta''} < \omega^{-m}.$$

The details are left as an exercise.

Letting $\delta = \max\{\delta', \delta''\}$ it follows that,

$$\forall m \in \mathbb{N} : 0 < \int_{+\delta}^{\infty} + \int_{-1}^{-\delta} < 2\omega^{-m},$$

showing that the contributions of the two 'tails' are extremely small, which is not yet to say that they may be ignored.

3) $[-\delta, +\delta]$. By Taylor's theorem, $u - \log(1+u) = (u^2/2)/(1+\theta u)^2$, for some θ, $0 < \theta < 1$, so that $u - \log(1+u) = u^2(1+\varepsilon'(u))/2$ for some $\varepsilon'(u) \simeq 0$. Replacing u by $v = u\sqrt{\omega}$ then gives, for some $\varepsilon(v) \simeq 0$,

$$\sqrt{\omega} \cdot \int_{-\delta}^{+\delta} = \int_{-\delta\sqrt{\omega}}^{+\delta\sqrt{\omega}} \exp(-v^2(1+\varepsilon(v))/2)dv.$$

Here δ is fixed such that $\delta\sqrt{\omega} \sim \infty$. Now let $f(v) = \exp(-v^2(1+\varepsilon(v))/2)$, $g(v) = \exp(-v^2/2)$, and $h(v) = \exp(-v^2/4)$, then from Theorem 4.1.3 it follows that,

$$\int_{-\delta}^{+\delta} \sim \omega^{-1/2} \cdot \int_{-\infty}^{+\infty} \exp(-v^2/2)dv = \sqrt{(2\pi/\omega)}.$$

Combining everything (in 2) it is sufficient to take m=1) finally leads to,

$$\lim_{x \to \infty} \frac{\Gamma(x+1)}{x^x e^{-x}\sqrt{(2\pi x)}} = 1.$$

$\square$

For more general results the reader may consult Van den Berg [28] and Koudjeti [29].

4.4   Nonstandard mathematics without the axiom of choice?

The preceding pages should have made it clear that nonstandard mathematics can be introduced in a way that is well-known to classical mathematicians. But logicians claim that from the point of view of logic and axiomatics when relating, say, $\mathbb{R}$ to $^*\mathbb{R}$ our naive approach obscures the insight into what is really happening. They are right, but nevertheless a naive approach could well be more understandable and also more acceptable, because there is no verdict on external sets, and because the axioms can easily be grasped (not even the Zermelo-Fraenkel axioms of set theory are necessary, requiring to look at natural numbers as sets and implying the unintended fact that there must be hyperlarge natural numbers). Yet, one stumbling stone remains: the axiom of choice. Couldn't we do without? Let us try and see what happens if the same general line of thinking is followed but the underlying free ultrafilter U over $\mathbb{N}$ is replaced by the Fréchet filter F^o (see Section 1.14). This means that again infinite sequences of classical entities will generate entities that either are new or are identified with their

classical counterparts. It also means that we no longer follow ideas of Luxemburg and others, but follow instead those of Chwistek (see Section 1.9) and perhaps those of Cauchy, who does not seem to be very explicit, however, when it comes to defining infinitesimals. Anyway, it would have been difficult for Cauchy to base his informal treatment of the infinitesimals on a free ultrafilter U, because the free ultrafilter theorem (see the Appendix) was not known to him, and moreover the axiom of choice had in his time still to be 'invented'. Chwistek is much more explicit, but does not develop anything that could be appreciated as a fully fledged infinitesimal calculus.

Clearly, $Q \in F^0$ if and only if $i \in Q$ for $i \geq n$ for some $n \in \mathbb{N}$. The latter will be rephrased as 'for i large enough'.

As has been made clear in Section 1.10 in order to introduce $^*\mathbb{R}$ (now with respect to F^0) it will again be necessary to consider all infinite sequences of real numbers. Instead of,

$$H(s_i) = H(t_i) \text{ if and only if } \{i : s_i = t_i\} \in U \text{ (Section 2.2)},$$

the definition of equality will be,

$$H(s_i) = H(t_i) \text{ if and only if } i \text{ is large enough},$$

and (Section 2.3),

$$H(s_i) = s \text{ if and only if } s_i = s \text{ for } i \text{ large enough}, s_i, s \in \mathbb{R},$$

Theorem 2.3.1 remains true, although the only-if part of the proof must be modified:
If $\{i : S_i = T_i\} \notin F^o$, then either there is a subsequence $(s_{i(j)}, j \in \mathbb{N})$ such that $s_i(j) \in S_{i(j)}$, but $s_i(j) \notin T_{i(j)}$, or there is a subsequence with the roles of S and T reversed (or both). Assume the first case, and take $s_i \in S_i$ arbitrarily if i is not an index of the subsequence. Then $H(s_i) \in S$ and $H(s_i) \notin T$, i.e. $S \neq T$, and similarly for the other case. $\qquad\square$

So, again,

$$H(s_i) = \{H(s_i) : s_i \in S_i\},$$

and Theorem 2.3.2 too remains valid.

Also the introduction of internal pairs, n-tuples in general and functions does not cause difficulties.

But with Theorem 2.4.2 the problems begin. Let $S = \{0, 1\}$, then S contains $q = H(0, 1, 0, 1, \ldots)$. Even though S is finite, $^*S \neq S$, because $q \neq 0$ and $q \neq 1$.

Moreover q is not hyperlarge, so what is it as an element of $^*\mathbb{N}$? Actually $^*S$ turns out to be infinite, and the conclusion is that $^*S$ contains far to many elements, i.e. the Fréchet filter does not operate properly and allows too much to go through. But perhaps this is not really harmful. The survey at the end of Section 2.4 reveals more trouble, however, for although $^*\emptyset$, $^* =$, $^* \in$ and $^*\cup$ are equal to or equivalent to $\emptyset$, $=$, $\in$ and $\cup$, respectively, this is not true for: $\neq$, $\not\subseteq$, $\cap$, $-$ and ^{c}. In fact, each of the relevant equivalences or equalities must be replaced by an implication or an inclusion in the correct direction, as the reader can find out for her or himself. It follows that $^* \neq$, $^* \not\subseteq$, $^*\cap$, $^*-$ and ^{*c} are all new relations or operations. Consequently, it follows from Section 2.6 (ignoring the implications of Section 2.5 on externality) that Łoś' theorem (Section 2.7) is no longer true, and that the same holds for transfer (Section 2.8). As one of the many counterexamples, consider $H(S_i)^*\cup H(T_i)$ that is no longer equal to $H(S_i)\cup T(i)$.

Section 2.7 reveals even more trouble: although $^*\wedge$ is equivalent to $\wedge$; $^*\neg$, $^*\vee$, $^* \Rightarrow$ and $^* \Leftrightarrow$ are not equivalent to $\neg$, $\wedge$, $\Rightarrow$ and $\Leftrightarrow$, respectively.

Finally, let us review the quantifiers. Although, given X_i and c_i,

$$\exists H(x_i) \in H(X_i) : H(x_i) = H(c_i)$$

is equivalent to,

$$H(\exists x_i \in X_i : x_i = c_i),$$

the equivalence is in general invalid if the simple statement $x_i = c_i$ is replaced by some other statement. Similar remarks apply to $\forall$.

Also the definition of $^*R$ with R some binary relation is cumbersome, for suppose that $s_{2i}Rt_{2i}$ but $\neg(s_{2i-1}Rt_{2i-1})$ for all $i \in \mathbb{N}$, then $H(s_iRt_i)$ is neither TRUE nor FALSE. Compare this to the example before with $H(0,1,0,1,\ldots)$. So $H(s_iRt_i)$ is not an ordinary statement, but an awkward internal something.

Yet the definitions of $^* <$ and $^* >$, for example, do not cause any problems, simply because $<$ and $>$ are not yet defined for hyperreal numbers, and the following case of transfer regarding continuity is legitimate,

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 : \exists \delta \in \mathbb{R}, \delta > 0 : \forall x \in \mathbb{R}, \mid x - c \mid < \delta : \mid f(x) - f(c) \mid < \varepsilon$$

is equivalent to,

$$\forall \varepsilon \in {}^*\mathbb{R}, \varepsilon > 0 : \exists \delta \in {}^*\mathbb{R}, \delta > 0 : \forall x \in {}^*\mathbb{R}, \mid x - c \mid < \delta : \mid {}^*f(x) - f(c) \mid < \varepsilon.$$

The definitions of infinitesimal and hyperlarge number can even be given a very simple form:

$$\varepsilon \simeq 0 \text{ if and only if } \varepsilon = H(\delta_i) \text{ where } (\delta_i) \text{ converges to } 0,$$

and,

$\omega \sim \infty$ if and only if $\omega = H(\omega_i)$ where (ω_i) converges to $+\infty$ or to $-\infty$.

Remarks:

1. It may well be that Cauchy's informal treatment comes closer to this kind of transfer, than to transfer with respect to some free ultrafilter U, but we will never know for sure.
2. Compare the definition of $\omega \sim \infty$ with Corollary 2.12.1.

Even the following remains true: f is continuous at $c \in \mathbb{R}$ if and only if,

$$\forall \delta \in {}^*\mathbb{R}, \delta \simeq 0 : {}^*f(c+\delta) - f(c) \simeq 0.$$

The proof of this equivalence is simple, because the latter statement is equivalent to,

$$\forall(\delta_i, i \in \mathbb{N}), \text{ tending to } 0 : (f(c+\delta_i) - f(c), i \in \mathbb{N}) \text{ tends to } 0,$$

which is equivalent to the continuity of f at c, and we are back at the plausible reasoning of Section 1.10. Note that Cauchy applied the simplified definition also to arbitrary $c \in {}^*\mathbb{R}$, so that he must have used some sort of S-continuity (see Section 3.3). In fact the $\varepsilon - \delta$ definition was introduced later on by Weierstrass.

The conclusion must be that with F^o instead of U nonstandard mathematics becomes a very restricted theory and nothing is left of the logician's equivalence ideal. More seriously to the ordinary mathematician, nothing is left of entirely new mathematical models that can be studied on the basis of a free ultrafilter and that cannot exist in classical mathematics, but nevertheless have turned out to be of great value not only within mathematics, but also outside it (such models have not been treated in this book). On the other hand what is left in the restricted theory can be based on well-known facts (such as the equivalence of (4.1) with ordinary continuity).

140

Appendix

The proof of the theorem below is based on the axiom of choice. It is the only instance where this axiom is required for the theory of nonstandard mathematics, assuming that the corresponding classical mathematics does not require it. The axiom of choice can be shown to be equivalent to Zorn's lemma, stating that if each totally ordered subset of a partially ordered nonempty set E has an upper bound in E with respect to the implied order, then E has at least one maximal element.

E is called *partially ordered* if there exists a binary relation ρ, called *order relation* or simply *order*, for some or all pairs of elements of E such that for $a, b, c \in E$,

1)   if $a\rho b$ and $b\rho c$ then $a\rho c$,
2)   if $a\rho b$ and $b\rho a$ then $a = b$, and
3) $a\rho a$ for all $a \in E$.

A subset G of the partially ordered E with order ρ is *totally* ordered with respect to ρ if $a\rho b$ or $b\rho a$ or both for all pairs (a, b), $a, b \in E$, and m is a *maximal element* of E if $[\forall a \in E : m\rho a]$ implies that $a\rho m$ and hence that $a = m$.

The proof showing the equivalence of the axiom of choice and Zorn's lemma is by no means trivial, and can be found in several textbooks, e.g. Dunford and Schwartz [30].

Theorem   *Free ultrafilters over* $\mathbb{N}$ *exist.*

Proof: By 'filter' will be meant 'filter over $\mathbb{N}$'. Let F^o be the Fréchet filter, i.e. the set of the complements of all finite subsets of $\mathbb{N}$, so that $Q \in F^o$ if and only if $\{n, n+1, \ldots\} \subseteq Q$ for some $n \in \mathbb{N}$. Let E be the set of all filters F such that $F \supseteq F^o$. Then E is nonempty and can be partially ordered by means of the order ρ, where $a\rho b$ if and only if $a \subseteq b$, i.e. the order is set inclusion. Let G be any totally ordered subset of E. Then, $B = \cup\{F : F \in G\}$ is an element of E, and B is an upper bound of G. Obviously $B \supseteq F^o$, and that B is a filter can be seen as follows.

1) $\mathbb{N} \in B$, as $\mathbb{N} \in F^o \subseteq B$.
2) $\emptyset \notin B$, as B is the union of filters.
3) If $Q \in B$ and $\mathbb{N} \supseteq R \supseteq Q$, then $Q \in F$ for some $F \in B$, hence $R \in F$, so $R \in B$.
4) If $Q, R \in B$, then $Q \in F$ and $R \in F'$ for certain $F, F' \in G$. As G is totally ordered $F \subseteq F'$ or $F' \subseteq F$ (or both). Assume $F \subseteq F'$, then $Q, R \in F'$, so $Q \cap R \in F' \subseteq B$.

That B is an upper bound for G is easily seen.

According to Zorn's lemma E must contain a maximal element U. Since $U \supseteq F^o$, U is a free filter. U is also an ultrafilter. For let $Q \in U$ be arbitrary. In order to show that either $Q \in U$ or $Q^c \in U$ consider the following two cases.

Case 1: Suppose $\forall Q' \in U : Q \cap Q'$ is infinite. Let,

$$V = \{T : T \subseteq \mathbb{N}, T \supseteq Q \cap Q' \text{ for some } Q' \in U\}.$$

Then V is a filter. The verification of this statement is left as an exercise. Also $V \in E$, for let $Q' = \{n, n+1, \ldots\}$, and $U \subseteq V$. By maximality $V \subseteq U$. But $Q \in V$, hence $Q \in U$.

Case 2: Suppose $\exists Q'' \in U : Q \cap Q''$ is finite. Then $\forall Q' \in U : Q^c \cap Q'$ is infinite. To see this let $Q' \in U$ be arbitrary. Since both Q' and Q'' belong to U, also $Q' \cap Q'' \in U$ and is infinite. Since $Q \cap Q' \cap Q''$ is finite $Q' \cap Q'' - Q \cap Q' \cap Q''$ is infinite, i.e. $Q^c \cap Q' \cap Q''$ is infinite and so is $Q^c \cap Q'$. Now apply Case 1 with Q^c instead of Q. $\qquad\square$

References

[1] Heath, T.L., *The Works of Archimedes, with the Method of Archimedes*, Dover Publications, 1912, chapter VII.

[2] Euler, L., *Introduction ad Analysin Infinitorum*, 1748.

[3] Luxemburg, W.A.J., *What is Nonstandard Analysis?*, American Mathematical Monthly, 80, 1973, 38–67.

[4] Cauchy, A.L., *Course d'analyse de l'école royale polytechnique*, 1821.

[5] Lakatos, Imre, *Cauchy and the Continuum: The Significance of Non-standard Analysis for the History and Philosophy of Mathematics*, The Mathematical Intelligencer, 1978, 151–161.

[6] Robinson, A., *Non-standard analysis*, Proceedings Royal Academy, Amsterdam, Series A, 64, 1961, 432–440.

[7] Robinson, A., *Nonstandard Analysis*, North-Holland, 1966 (2nd revised edition in 1974, 3rd edition in 1996, Princeton University Press).

[8] Hahn, H., *Über die nichtarchimedische Groszensysteme*, S.-B. Wiener Akademie, Math.-Natur. Kl. 116, Abt. IIa, 1907, 601–655.

[9] Skolem, T., *Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzahlbare unendlich vieler Aussagen mit ausschliesslich Zahlenvariabelen*, Fund. Math. 23, 1933, 150–161,

[10] Hewitt, E. *Rings of real-valued continuous functions I*, Trans. Amer. Math. Soc. 64, 1948, 45–99.

[11] Łoš, J., *Quelques remarques, theorèmes, et problèmes sur les classes definissables d'algebras*, in: Skolem et al., eds., Mathematical interpretations of formal systems, North-Holland, 1955, 98–113.

[12] Laugwitz, D., and C. Schmieden, *Eine Erweiterung der Infinitesimalrechnung*, Mathematische Zeitschrift 69, 1958, 01–39.

[13] Luxemburg, W.A.J., *Nonstandard Analysis, Lectures on A. Robinson's theory of infinitesimal and infinitely large numbers*, Caltech Bookstore, 1962.

[14] Nelson, E., *Internal set theory*, Bull. Amer. Math. Soc. 83, 1977, 1165–1193.

[15] Robert, A., *Nonstandard Analysis*, Wiley, 1988.

[16] Diener, F., et G. Reeb, *Analyse Non Standard*, Hermann, 1989.

144

[17] Chwistek, L., *Über die Hypothesen der Mengenlehre*, Mathematische Zeitschrift 25, 1926, 439–473.

[18] Chwistek, L., *The Limits of Science*, translated from the Polish, Routledge Kegan Paul, 1948.

[19] Chang, C.C., and H.J. Keisler, *Model Theory*, North-Holland,

[20] Beeson, M.J., *Foundations of Constructive Mathematics*, Springer, 1985.

[21] Bishop, E., and D. Bridges, *Constructive Analysis*, Springer, 1985.

[22] Beth, E.W., *The Foundation of Mathematics*, North-Holland, 1965.

[23] Beth, E.W., *Mathematical Thought*, Reidel, 1965, chapter V.

[24] Heyting, A., *Intuitionism, An Introduction*, North-Holland, 1966.

[25] Potter, M.D., *Sets: An Introduction*, Clarendon Press, 1990.

[26] Keisler, H.J., *Elementary Calculus, An Infinitesimal Approach*, Prindle, Weber Schmidt, 1986.

[27] Berg, I.P. van den, en T. Sari, *Inleiding tot de infinitesimaalrekening*, (Introduction to the infinitesimal calculus; Lecture Notes, University of Groningen, 1988; private communication; in Dutch).

[28] Berg, I.P. van den, *Nonstandard Asymptotic Analysis*, Lecture Notes in Mathematics, nr. 1249, Springer, 1987.

[29] Koudjeti, F., *Elements of External Calculus with an Application to Mathematical Finance*, thesis, University of Groningen, 1995.

[30] Dunford, N., J.T. Schwartz, *Linear Operators, part I*, Interscience, 1957.

Index