ANALYSIS
using
ultrasmall numbers

Teacher's Manual

Contents

Part I
Why?

In the writing of a limit

$$\lim_{x \to a} f(x) = L$$

we will say "$f(x)$ tends to L when x tends to a" but one of the difficulties is that each part of the sentence is meaningless when considered alone; x cannot tend to a by itself. Similarly with "$f(x)$ tends to L". The concept must be grasped in its entirety or not understood at all.

For pre-university and introductory level, showing the full ε-δ definition certainly does not make things easier with δ depending on ε.

Defining the derivative as the slope of the tangent leaves us with the difficult task of defining the tangent before the derivative.

For the integral, a sum of thin slices approximates the area but the limit when slices tend to a thickness of zero seems to imply that the area is a sum of zeroes.

These difficulties are well known to analysis teachers. Many avoid them by resorting to "hand waving" definitions and proofs where informally "approaching" is used, or "arbitrarily close". Mathematical rigour is then lost.

Another way to circumvent the difficulty is to change the approach and use a version of nonstandard analysis – this is the choice we have made.

Many difficulties disappear or become more palatable when using the concept of ultrasmall numbers. In particular, the limit is defined in such a way that each part of the sentence has a meaning on its own – hence didactic and intellectual steps are smaller.

Most mathematicians have an intuitive idea of infinitesimals. These mental representations are often used to explain the fundamental concepts before rigorous formalisations are given. Recent work by Karel Hrbacek, based on Yves Péraire's research offers a new formalisation of these ideas, mathematically rigorous yet still reasonably close to intuitive ideas and with a lower level of technical complexity. We have adapted this work to Geneva high school level.

There have been several attempts to use nonstandard analysis for teaching, by Keisler [5], Stroyan [7] or Robert [6] for instance. These previous theories used different approaches but one limitation was common to all: if h is infinitely small (in a way clearly defined in each theory) the derivative of $f : x \mapsto x^2$ was easy at $x = 2$ but difficult at $x = 2 + h$. The approach used here does not have this drawback.

This paper is a companion to the book "Analysis with ultrasmall numbers"

[1]. The purpose, here, is to offer the possibility of a quick start if one wants to use it as a teaching approach. The book offers deeper insights, far more rigour and more formal justifications of the theory. In particular, the discussion about the fact that theorems which are true in classical mathematics (without the concept of ultrasmall number) remain true in this approach is found in the book.

This approach has been used for several years in several high schools in Geneva (Switzerland).

[1]Analysis with ultrasmall numbers, Karel Hrbacek, Olivier Lessmann, Richard O'Donovan, CRC Press, 2015, ISBN: 9781498702652

Part II

How to read a proof

The goal of this part is to help instructors or examiners understand a proof which uses ultrasmall numbers. This can happen when, in our school, final exams have an external juror in addition to the teacher. It can also be used to see that for the trained mathematician, the translation back to "classical" methods is quite straightforward.
This part does not explain how to write the proofs, only how to read them.

As in any book about analysis, we do not give the axioms of set theory, but state instead, properties of real numbers.

Observability

Extra Axioms – called also Principles – are used which allow to make an extra distinction within the real numbers: observability.

The intuition is that “ordinary numbers” are observable but that there are extremely small numbers (ultrasmall) which are so tiny that they are not observable. But if one zooms in to observe these tiny numbers, one can still see the previously observable numbers. And if such a tiny number h is added to 2, then $2+h$ is less observable than 2, which remains observable when $2+h$ is observable.

Given x and y, then x is as observable as y, or y is as observable as x. They may have the same observability. Observability is transitive.

One can consider the metaphor of scales of observation.

Numbers defined without the concept of observability are observable relative to any real number: they are always observable – or standard.

Ultrasmall

Relative to any real number, there exist ultrasmall numbers: numbers which are less, in absolute value, than any strictly positive observable number, yet not zero.

Relative to a, if h is ultrasmall, then $\frac{1}{h}$ is ultralarge. But then we also have that a and $a+h$ are extremely close, written $a \simeq a+h$, where the only new symbol is "$\simeq$" which reads "ultraclose": a difference which is ultrasmall or zero. Note that $a+h$ is not as observable as a.

An ultrasmall number has the "flavour" of an infinitesimal.

Context

Given a formula which has some parameters, the concept of ultrasmall, ultralarge or ultraclose always refer to the whole list of parameters. If h is ultrasmall, it must be ultrasmall relative to each parameter. Thus "$\simeq$" by its definition refers to all of the parameters.

The context of a statement is the list of parameters used in that statement. Some instructors use the word "observability" instead of context. This is a pedagogical choice which does not change the theory.

Closure

Non observable numbers do not show up as results of operations if they are not introduced explicitly. Given f and a, then $f(a)$ is observable (the context being a and the parameters of f).

Observable neighbour

For any number x which is not ultralarge, there is an observable number a such that $x \simeq a$.

Or:

Any number x which is not ultralarge can has be written in the form $x = a+h$ where a is observable and h is ultraclose to zero. Then a is the observable neighbour of x.

The existence of the observable neighbour (or observable part) is equivalent to the completeness of $\mathbb{R}$.

Limit

$\lim_{x \to a} f(x) = L$ is, here, defined by, L is observable and

$$x \simeq a \ (x \neq a) \ \Rightarrow f(x) \simeq L$$

and does not depend on the choice of x.

Since the sum of two ultrasmall numbers is ultrasmall or zero and similar algebraic properties, it is easy to prove that the sum, resp. product, quotient of limits is the limit of the sum, resp. product, quotient.

Translation

When reading dx or $h \simeq 0$, consider that h is a real number with the idea of it being arbitrarily small in absolute value. When we keep only the observable part of the result (if it exists), this translates to "limit when h tends to 0".

Examples

Continuity

f is continuous at a if $f(a+dx) \simeq f(a)$ whenever $dx \simeq 0$. This translates to $\lim_{h \to 0} f(a+h) = f(a)$ as usual. It is also written in a simple form: $x \simeq a \Rightarrow f(x) \simeq f(a)$

Derivative of f at a

If there is an observable value D such that for any dx we have $\dfrac{f(a+dx)-f(a)}{dx} \simeq D$ then $f'(a) = D$. This translates to $\lim_{h \to 0} \dfrac{f(a+h)-f(a)}{h} = D$.

(The context is given by the parameters of f and a)

Example proof: Continuity of $f \circ g$

Assume that g is continuous at a and f is continuous at $g(a)$. Then if $x \simeq a$ we have $g(x) \simeq g(a)$ by continuity of g and $f(g(x)) \simeq f(g(a))$ by continuity of f.

We now proceed to show how to write such proofs by considering some specific examples and discuss some pedagogical issues.

Part III
Basic Principles

The following principles are consequences of axioms added to the classical axioms of set theory. When teaching analysis one usually does not study axioms but their consequences on real numbers. Theses will thus be considered axiomatically together with other properties of real numbers..

Definition 1
*The **context** (or **observability**) of a property, function or set is the list of parameters used in its definition.*

Observability Principle

- A number is observable relative to a context if it is observable relative to at least one parameter of the context.
- Every number is observable relative to some context.
- Two numbers a and b will always have a common context. If a is not observable relative to b, then b will be observable relative to a.

The word "observable" , by convention, refers to a context. The context is the parameters, sets and functions the statement is about. Therefore to determine the context of a statement, one must be able to describe about what it says something.

It is not a bad exercise – pedagogically – to ask students what a theorem, definition, or statement is about. The derivative of f at a for instance, is only about f and a, not about the dx or the ε and δ which are all dummy variables.

If a number is observable whenever any other number is observable, we say that it is *always* observable.

Closure Principle
Numbers defined without reference to observability are always observable. If a number satisfies a given property, then there is an observable number satisfying that property

(Observability here being given by the property).

The closure principle tells us that all "familiar" numbers such as $1;\ 3;\ 10^{10};\ \sqrt{2}$ or π are always observable

But also that if a number is calculated using some parameters, the resulting number will be observable. Non observable results do not show up unless explicitly summoned.

Example

Let $f : x \mapsto x^2 + 3$ The constants of f are 2 and 3 which are always observable. There is no parameter. The number $f(4)$ is thus also always observable.

Let $g : x \mapsto ax^2 + b$. The parameters are a and b. Thus $f(4)$ is as observable as a and b.

In general $f(x)$ is as observable as x.

Definition 2

*A real number is **ultrasmall** if it is non zero and smaller in absolute value than any strictly positive observable number*

This definition makes an implicit reference to a context which therefore must be determined before an ultrasmall number is referred to.

Principle of ultrasmallness

Whatever the number x, there exist ultrasmall real numbers relative to x.

Note in particular that if ε is ultrasmall relative to some context containing a then ε is not observable — neither is $a + \varepsilon$.

Definition 3

*A real number is **ultralarge** if it is larger in absolute value than any strictly positive observable number*

⚠ Note that 0 is not ultrasmall. This can be justified by observing that the reciprocal of an ultrasmall is ultralarge and 0 has no reciprocal.

⚠ Note the asymmetry: if h is ultrasmall relative to x, then it is not observable. But then x is observable relative to h, hence x **is not** ultralarge relative to h.

Definition 4

*Let a, b be real numbers. We say that a is **ultraclose** to b (relative to some context), written*

$$a \simeq b,$$

if $b - a$ is ultrasmall or if $a = b$.

In particular, $x \simeq 0$ if x is ultrasmall or zero.

Definition Principle
The only acceptable properties are those that do not refer to observability ("classical" definitions) or those that use the symbol "$\simeq$".

As one would expect (see below for a proof) the reciprocal of an ultralarge is an ultrasmall, thus n is ultralarge may be characterised by stating that $1/n \simeq 0$.

In class, it is often possible to replace the second part of the closure principle by one of its consequences, using contextual notation:

$$\boxed{f(a) \quad \text{is observable}}$$

This refers to the context, by the word "observable". The only parameters of this property are f and a.

A context is *extended* if parameters are added to the list.

Stability Principle
A property is true if and only if it is true when its context is replaced by an extended context.

This principle ensure, for example, that if f and g are functions, then the derivatives $f'(a)$ and $g'(a)$ remain the same even if $f'(a)$ is calculated using the additional parameters of g. For all functions given by explicit rules it is seen by inspection. Formally, it is a consequence of stability.

"$\simeq$" is the only new symbol introduced.

Why ultralarge rather than infinity?

The classical definition that integers are finite cardinalities remains true hence ultralarge integers cannot be infinite. They are huge, very huge, but not infinite.
Consequently, their reciprocals will be ultrasmall. These numbers are real numbers.
This approach contradicts no classical mathematical statement but by using an extra distinction, it can make statements that cannot be expressed without the concept of observability.

If $a \simeq b$ then a and b are said to be neighbours. If a is a neighbour of b and is observable (relative to some context) then a is the observable neighbour of b.

Principle of the observable neighbour
(We also say "observable part".
Relative to a context, for any real number x which is not ultralarge, there is an observable number a such that $x \simeq a$.

Such an x can be written in the form $a + h$ – where a is observable and $h \simeq 0$.

a is the **observable part** of x.

This principle is similar to the completeness of the reals.

Exercise 1 (answer page 30)
Using the principles and definitions, show that (relative to a given context)

(1) If ε is ultrasmall, then $\frac{1}{\varepsilon}$ is ultralarge.

(2) If M is ultralarge then $\frac{1}{M}$ is ultrasmall.

Rule 1
Given a context. Let a be observable and non zero and $h \simeq 0$ (non zero) and $\varepsilon \simeq 0$. Then

(1) $a \cdot h \simeq 0$

(2) $\dfrac{a}{h}$ *is ultralarge.*

(3) $\varepsilon \cdot h \simeq 0$

(4) $\varepsilon + h \simeq 0$

Exercise 2 (answer page 30) Prove rule 1.

Rule 2
Given a context. Let a and b be observable and x and y be such that $a \simeq x$ and $b \simeq y$. Then

(1) $a \pm b \simeq x \pm y$.

(2) $a \cdot b \simeq x \cdot y$.

(3) If $b \neq 0$ *then* $\dfrac{a}{b} \simeq \dfrac{x}{y}$.

(4) $a \simeq b \Rightarrow a = b$.

Exercise 3 (answer page 30) Prove rule 2.

We refer to these rules as "ultracalculus". The proofs are algebraic and a good training for students to learn how to work with definitions.

A consequence of (4) is that the observable part is unique. If $a \simeq x \simeq b$ with a and b observable, then $a = b$. This is equivalent to the uniqueness of the limit!

The existence of the observable neighbour is not guaranteed in $\mathbb{Q}$. Let x be a rational ultraclose to $\sqrt{2}$ (for ultralarge whole number N, take the first N digits of $\sqrt{2}$). $\sqrt{2}$ is standard, by closure: it is the unique positive solution of $x^2 = 2$ which has no parameters. Since $\sqrt{2}$ is always observable and not rational, the observable neighbour of x is not in $\mathbb{Q}$.

Part IV
The derivative

We will not show the proofs of all theorems about derivatives but we hope to show enough that the reader can perform the remaining proofs as exercises.

The definition of the derivative at a point requires that the function be defined at least on an open interval $]b, c[$ containing a. Since the domain is determined by f (closure), it is observable and we can always suppose that b, c are observable.

Since x is the independent variable, its increment can always be chosen to be non zero. It can be positive or negative. We write dx for this ultrasmall (non zero) increment.

Definition 5
Let f by a real function defined on an open interval containing a.
We say that f is differentiable at a if there is an observable number D such that for any $dx \simeq 0$ we have

$$\frac{f(a+dx)-f(a)}{dx} \simeq D$$

We write $D = f'(a)$, the derivative of f at a.

(The context is given by f and a)

- The result must not depend on dx
- When it exists, the derivative is the observable neighbour of $(f(a+dx)-f(a))/dx$.

Example
Let

$$f : x \mapsto x^2 + 3x$$

For the derivative at $x = 5$. The parameters are 2,3 and 5 (the context). Let dx be ultrasmall. Then

$$\frac{f(5+dx)-f(5)}{dx} = \frac{((5+dx)^2 + 3(5+dx)) - (25+15)}{dx} = \frac{10dx + 3dx + dx^2}{dx} = 10+3+dx.$$

Then $13 + dx \simeq 13$ which is observable and does not depend on dx, hence it is the derivative.

⚠ The same proof could be done with $x = a$ in general. Then a is part of the context. For those familiar with other forms of nonstandard calculus: note that here we could have directly shown that $f'(a) = 2a + 3$ for all a whether always observable or less observable. This is one of the main features of this approach.

Note that it is possible to be reasonably "careless" about the context. When the formula is expanded, dx must be ultrasmall relative all other terms in the expansion.

Example

Let

$$g : x \mapsto |x|$$

at 0. *Let* dx *be ultrasmall. If* $dx > 0$, *then*

$$\frac{g(0+dx) - g(0)}{dx} = \frac{g(dx) - g(0)}{dx} = \frac{dx - 0}{dx} = \frac{dx}{dx} = 1.$$

But if $dx < 0$, *then*

$$\frac{g(0+dx) - g(0)}{dx} = \frac{g(dx) - g(0)}{dx} = \frac{-dx - 0}{dx} = \frac{-dx}{dx} = -1.$$

There is thus no unique real number satisfying the condition independently of dx. *The conclusion is that the derivative of* g *does not exist for* $x = 0$.

The modulus function is defined with no reference to observability, it is thus a function which is always observable. But in fact, one can ignore this and simply use that dx is ultrasmall relative to the function and 0.

Differentiation rules

Let dx be ultrasmall relative to a and f. We write

$$\Delta f(a) = f(a + dx) - f(a).$$

Then

$$\frac{\Delta f(x)}{dx} \simeq f'(x).$$

or:

$$\frac{\Delta f(x)}{dx} = f'(x) + \varepsilon \qquad \text{with } \varepsilon \simeq 0.$$

And also $f(x + dx) = f(x) + \Delta f(x)$.

We now show without further comment the proofs of some the usual rules of differentiation and leave the others as exercises.

We do not specify the context explicitly since it should by clear now that it is the list of parameters. (dx is not a parameter but a dummy variable, as are the ε and δ of classical proofs: the context is "what we are talking about".) dx is always chosen ultrasmall not zero, hence we do not specify it every time.

Theorem 1
If $f'(a)$ exists, then $\Delta f(a) \simeq 0$.

Proof: $\Delta f(a) = \frac{\Delta f(a)}{dx} \cdot dx \simeq f'(a) \cdot 0 = 0$

□

Alternative proof: $\Delta f(a) = f(a+dx) - f(a) = \underbrace{f'(a) \cdot dx}_{\simeq 0} + \underbrace{\varepsilon \cdot dx}_{\simeq 0}$ for $\varepsilon \simeq 0$. It is then immediate that $\Delta f(a) \simeq 0$.

This shows that differentiability implies continuity, but since here, we choose to study derivatives before continuity, the concept is not mentioned.

Theorem 2
Let f and g be functions differentiable at a. Then the function $f \cdot g$ is differentiable at a and

$$(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a).$$

Proof:

$$\begin{aligned}
\frac{\Delta(f \cdot g)(a)}{dx} &= \frac{f(a+dx) \cdot g(a+dx) - f(a) \cdot g(a)}{dx} \\
&= \frac{\Big(f(a) + \Delta f(a)\Big) \cdot \Big(g(a) + \Delta g(a)\Big) - f(a) \cdot g(a)}{dx} \\
&= \frac{f(a) \cdot \Delta g(a) + \Delta f(a) \cdot g(a) + \Delta f(a) \cdot \Delta g(a)}{dx} \\
&= f(a) \cdot \frac{\Delta g(a)}{dx} + \frac{\Delta f(a)}{dx} \cdot g(a) + \frac{\Delta f(a) \cdot \Delta g(a)}{dx} \\
&\simeq f'(a) \cdot g(a) + f(a) \cdot g'(a),
\end{aligned}$$

since, in particular, $\dfrac{\Delta f(a) \cdot \Delta g(a)}{dx} = \dfrac{\Delta f(a)}{dx} \cdot \Delta g(a) \simeq f'(a) \cdot 0 = 0$

($\Delta g(a) \simeq 0$ by theorem 1 and $f'(a)$ is observable by its definition, and observable$\times$ ultrasmall is ultrasmall, by rule 1).

Since $f'(a) \cdot g(a) + f(a) \cdot g'(a)$ is observable, it is the derivative.

□

⚠ Stability is in fact hidden in the proof. If $g(a)$ is not as observable as $f(a)$ then the context for f' is extended to contain also $g(a)$ and this becomes the general context. In class, almost all functions are always observable so this subtlety is not really an issue, and in the case of explicitly given functions, it is true by inspection.

Theorem 3 (Chain Rule)
Let g be a function differentiable at a and f a function differentiable at $g(a)$. Then the function $f \circ g$ is differentiable a and

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a).$$

Proof: We consider two cases (1) $\Delta g(x) = 0$ and (2) $\Delta g(x) \neq 0$.

(1) If $\Delta g(x) \neq 0$, then

$$\begin{aligned}
\frac{f(g(x+dx)) - f(g(x))}{dx} &= \frac{f(g(x) + \Delta g(x)) - f(g(x))}{dx} \\
&= \frac{f(g(x) + \Delta g(x)) - f(g(x))}{\Delta g(x)} \cdot \frac{\Delta g(x)}{dx} \\
&\simeq f'(g(x)) \cdot g'(x)
\end{aligned}$$

since f is differentiable at $g(x)$ and since g is differentiable x we have $\Delta g(x) \simeq 0$.

The proof may be easier to read if we use Leibniz' notation, replacing $g(x)$ by y and $\Delta g(x)$ by Δy, then

$$\frac{\Delta f(y)}{dx} = \frac{\Delta f(y)}{\Delta y} \cdot \frac{\Delta y}{dx} \simeq f'(y) \cdot y'$$

(2) If $\Delta g(x) = 0$ then $g(x+dx) = g(x)$ and $g'(x) = 0$, therefore $\dfrac{f(g(x+dx)) - f(g(x))}{dx} = 0$. And $f(g(x))' = f'(g(x)) \cdot g'(x)$.

□

Definition 6
If f is differentiable at a, the quantity $f'(a) \cdot dx$ is noted $df(a)$: it is the ***differential*** *of f at a.*

Then we have

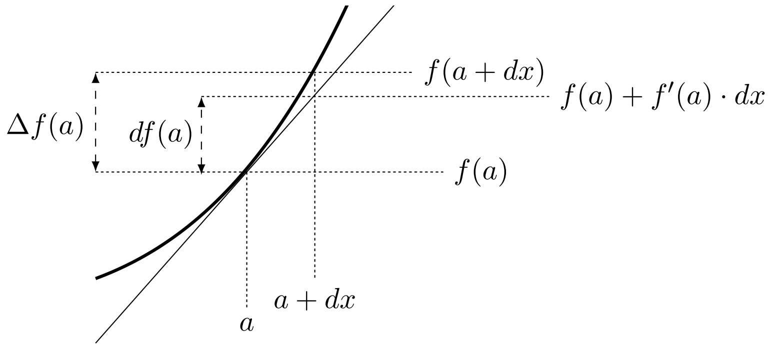
$$\frac{df(a)}{dx} = f'(a)$$

which is an equality. The expression $df(a)/dx$ here, really is a quotient.

⚠ Note the difference between Δy and dy i.e., between the variation and the differential. We have $\dfrac{\Delta y}{dx} \simeq y'$ hence $\dfrac{\Delta y}{dx} = y' + \varepsilon$ (for $\varepsilon \simeq 0$). Thus

$$\begin{aligned} \Delta y &= y' \cdot dx + \varepsilon \cdot dx \\ &= dy + \varepsilon \cdot dx \end{aligned}$$

The differential is the variation along the tangent line.



Theorem 4 (de l'Hospital's Rule for $0/0$ – simple form)
Let f and g be functions differentiable at a. Suppose that $f(a) = g(a) = 0$ but that $g'(a) \neq 0$. Then

$$\frac{f(a+dx)}{g(a+dx)} \simeq \frac{f'(a)}{g'(a)}$$

Exercise 4 (answer page 31) Prove theorem 4.

Part V
Continuity

Definition 7
*Let f be a function defined on an open interval containing a. We say that f is **continuous at** a if*

$$f(x) \simeq f(a) \quad \textit{whenever } x \simeq a.$$

This is a property of f and a hence this determines the context.
Alternatively: f is continuous at a if

$$f(a + dx) \simeq f(a)$$

or still

$$\Delta f(a) \simeq 0$$

If f is differentiable at a then f is continuous at a. This is a restatement of theorem 1.

Theorem 5
Let f and g be two functions continuous at a. Then

(1) $f \pm g$ is continuous at a.

(2) $f \cdot g$ is continuous at a.

(3) $\frac{f}{g}$ is continuous at a if $g(a) \neq 0$.

Proof: This theorem is a direct application of the rules that $x \simeq a$ and $y \simeq b$ imply $x \cdot y \simeq a \cdot b$ et $x + y \simeq a + b$ and $x/y \simeq a/b$ (rule 2).

□

Theorem 6
Suppose that g is continuous at a and that f is continuous at $g(a)$. Then $f \circ g$ is continuous at a.

Proof: Let $x \simeq a$. Then $g(x) \simeq g(a)$ by continuity of g at a so $f(g(x)) \simeq f(g(a))$ by continuity of f at $g(a)$.

□

Example

By rule 2 $x \mapsto x^2$ is continuous- We use this to show that $x \mapsto \sqrt{x}$ is also continuous.

Proof:

Let $x \simeq a$.

$\sqrt{x}$ is either less than x if $x > 1$ or less than 1, hence it is not ultralarge.

Consider its observable neighbour $b \simeq \sqrt{x}$.

By continuity of $x \mapsto x^2$ we have $b^2 \simeq x$.

Since $x \simeq a$ we have $b^2 \simeq a$ hence $b^2 = a$ So $b = \sqrt{a}$, and we conclude that $\sqrt{x} \simeq \sqrt{a}$.

□

In general:

Theorem 7 (Continuity of the inverse)
Let $f : I \to J$ (I and J closed and bounded) a continuous one-to-one correspondence. Then $f^{-1} : J \to I$ is continuous.

Proof: Let a be observable and in I.

Note $f(a) = d$ and $a = f^{-1}(y)$. Let $y \simeq f(a)$, we must show that $f^{-1}(y) \simeq a$

Since J is closed and bounded, every $y \in J$ has an observable neighbour in J.

Note c the observable neighbour of $f^{-1}(y)$. By continuity of f we have $f(c) \simeq f(f^{-1}(y)) = y \simeq d$

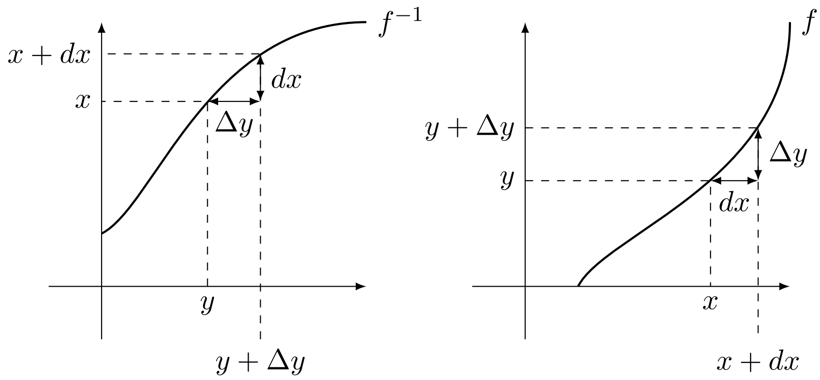
hence $f(c) = d$ and $c = f^{-1}(d) = a$ because f is one-to-one, hence $f^{-1}(y) \simeq a$.

□

Theorem 8 (Derivative of the inverse)
Let $f : (a, b) \to \mathbb{R}$ continuous and having an inverse. Let $y = f(x)$. If f is differentiable at $x \in (a, b)$ with $f'(x) \neq 0$, then f^{-1} is differentiable at y and

$$(f^{-1})'(y) = \frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(y))}.$$

Proof:



Then

$$\frac{\Delta f^{-1}(y)}{\Delta y} = \frac{dx}{\Delta y} = \frac{1}{\frac{\Delta y}{dx}} = \frac{1}{\frac{\Delta f(x)}{dx}} \simeq \frac{1}{f'(x)},$$

since $\dfrac{\Delta f(x)}{dx} \simeq f'(x) \neq 0$ by hypothesis. But $\frac{1}{f'(x)}$ is observable by closure, so $(f^{-1}(y))'$ exists and

$$(f^{-1}(y))' = \frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(y))}.$$

□

Part VI

Limit

Continuity being defined without reference to the limit, we can define the limit in terms of continuity without circularity.

The limit of f at a is the value that f *should* take to be continuous at a. Formally:

Definition 8
*Let f be function defined around a. We say that the **limit of** f **at** a **exists** if there is an observable real number L such that*

$$f(x) \simeq L \quad \textit{whenever } x \simeq a \ (x \neq a).$$

Example

Consider the function

$$f : x \mapsto \frac{2x^2 - 7x + 3}{x - 3}, \quad \textit{with Dom}(f) = \mathbb{R} \setminus \{3\},$$

and its limit at $a = 3$. The function is defined f around 3. Let $x \simeq 3$ ($x \neq 3$) . Then

$$f(x) = \frac{(x-3)(2x-1)}{x-3} = 2x - 1 \simeq 2 \cdot 3 - 1 = 5$$

Since 5 does not depend on the choice of x and is ultraclose to $f(x)$, it is the limit.

Theorem 9
If the limit of f at a exists then it is unique.

Proof: Suppose that L_1 and L_2 are observable and such that

$$f(x) \simeq L_1 \quad \text{and} \quad f(x) \simeq L_2$$

whenever $x \simeq a$ $(x \neq a)$.

But then $L_1 \simeq L_2$ and are both observable. therefore $L_1 = L_2$

□

We write

$$\lim_{x \to a} f(x) = L$$

if the limit if f at a is L.

Part VII
Intermediate Value Theorem

For local properties (derivatives, continuity) the method is to look at ulraclose neighbours to determine an approximation of the required value. The observable part being the exact value. For global properties (Intermediate Value, Extreme value, Integral) the method is to divide the interval into an ultralarge numbers of pieces, find the best approximation and use its observable part.

For $f(a) < 0 < f(b)$ we search for a c in $[a; b]$ with the such that $f(c) \simeq 0$. The context is a, b et f. We first choose an ultralarge whole number N. Then $dx = (b - a)/N$ is ultrasmall. We consider the $N + 1$ points $x_i = a + i \cdot dx$, for $i = 0, \ldots, N$. Then we find an "ultragood" approximation to the required number. The observable neighbour of this turns out to be the number having the property.

Theorem 10 (Intermediate Value)
Let f be a function continuous on $[a; b]$ such that $f(a) < 0 < f(b)$. Then there is a $c \in [a; b]$ such that $f(c) = 0$.

Proof: (The context is f, a, b and d.)

Let N be a positive ultralarge integer and let $dx = (b - a)/N$. Then $dx \simeq 0$. Consider $x_i = a + i \cdot dx$, for $i = 0, \ldots, N$ (hence $x_0 = a$ and $x_N = b$). Since it is a finite collection, there is a first index j such that $f(x_{j+1}) \geq 0$. Then we have

$$f(x_j) \leq 0 \leq f(x_{j+1}).$$

Let c be the observable neighbour of x_j (it exists since x_j is bounded by observable a and b). Then $x_j \simeq c$. Furthermore, $c \in [a; b]$ et $c \simeq x_{j+1}$ because $x_j \simeq x_{j+1}$. By continuity of f at c we have

$$f(c) \simeq f(x_j) < 0 \quad \text{et} \quad f(c) \simeq f(x_{j+1}) \geq 0.$$

We deduce that

$$f(c) \simeq 0.$$

But $f(c)$ is observable by closure and also 0, so

$$f(c) = 0.$$

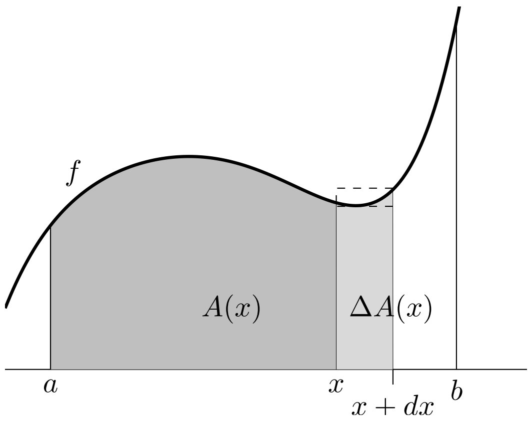
□

Part VIII
The integral

We consider two approaches to the integral. First assuming that an area function exists, then by summing an ultralarge number of ultrasmall values.

Consider a non negative function f continuous on $[a;b]$. Let $A(x)$ be the area between the function and the x-axis between a and x.

The variation of the area between x and $x+dx$ is noted $\Delta A(x)$.



Theorem 11
Let f be a non negative function continuous on $[a;b]$. Then the function

$$A : x \mapsto A(x),$$

where $A(x)$ is the area under the curve between a and x satisfies the two following properties:

(1) $A'(x) = f(x)$, for every $x \in [a;b]$.

(2) $A(a) = 0$.

Proof: (2) is obvious. We show (1).

The context is a, f and x. Let dx be ultrasmall and positive. Since f is continuous on $[x; x+dx]$ it attains its maximum and minimum on $[x; x+dx]$. Note $(x_M, f(x_M))$ for the maximum and $(x_m, f(x_m))$ for the minimum. Then

$$f(x_m) \cdot dx \leq \Delta A(x) \leq f(x_M) \cdot dx.$$

Therefore

$$f(x_m) \leq \frac{\Delta A(x)}{dx} \leq f(x_M).$$

As f is continuous at x (which is part of the context hence observable) and $x \simeq x_M$ and $x \simeq x_m$ we have $f(x) \simeq f(x_M)$ and $f(x) \simeq f(x_m)$ (hence also $f(x_M) \simeq f(x_m)$), this implies that

$$\frac{\Delta A(x)}{dx} \simeq f(x).$$

The same result follows if dx is negative and

$$A'(x) = f(x),$$

because $f(x)$ is observable.

□

Definition 9
*Let f be a function defined on $[a;b]$. We say that f **is integrable on** $[a;b]$ if there is an observable real number I such that for any ultralarge whole number n with $dx = \dfrac{b-a}{n}$ and $x_i = a + i \cdot dx$ for $i = 0, \ldots, n$, we have*

$$\left(\sum_{i=0}^{n-1} f(x_i) \cdot \Delta x\right) \simeq I.$$

*If such a number exists, we call it the '**integral of f between a and b** written*

$$\int_a^b f(x) \cdot dx.$$

Theorem 12 (Fundamental Theorem of Analysis)
Let f be a function continuous on $[a;b]$. Let F be an antiderivative of f on $[a;b]$. Then

$$\int_a^b f(x) \cdot dx = F(b) - F(a).$$

Proof:

Let N be a positive ultralarge positive integer let $dx = (b-a)/N$ and $x_i = a + i \cdot dx$ for $i = 0, \ldots, N$. We write $F(b) - F(a)$ as a telescoping sum:

$$F(b) - F(a) = \sum_{i=0}^{N-1} F(x_{i+1}) - F(x_i) = \sum_{i=0}^{N-1} F(x_i + dx) - F(x_i).$$

By the mean value theorem, there is an x in $[x_i, x_{i+1}]$ such that $F(x_{i+1}) - F(x_i) = F'(x) \cdot dx$. Since x is between a and b it is not ultralarge. Let c be the observable neighbour of x. The derivative of F is assumed to be continuous on $[a;b]$ so $F'(x) \simeq F'(c) \simeq F'(x_i)$. Hence $F'(x) = F'(x_i) + \varepsilon_i$ and

$$F(x_{i+1}) - F(x_i) = F'(x_i) \cdot dx + \varepsilon_i \cdot dx = f(x_i) \cdot dx + \varepsilon_i \cdot dx.$$

Therefore we have

$$F(b) - F(a) = \sum_{i=0}^{N-1} f(x_i) \cdot dx + \sum_{i=0}^{N-1} \varepsilon_i \cdot dx.$$

We now show that

$$\sum_{i=0}^{N-1} \varepsilon_i \cdot dx \simeq 0.$$

Take positive and observable c. Then $\dfrac{c}{b-a}$ is observable by closure and $|\varepsilon_i| < \dfrac{c}{b-a}$, for every i since $\varepsilon_i \simeq 0$. Then

$$\Big| \sum_{i=0}^{N-1} \varepsilon_i \cdot dx \Big| \leq \sum_{i=0}^{N-1} |\varepsilon_i| \cdot dx < \sum_{i=0}^{N-1} \frac{c}{b-a} \cdot \frac{b-a}{N} = c \sum_{i=0}^{N-1} \frac{1}{N} = c.$$

This means that $\left| \sum_{i=0}^{N-1} \varepsilon_i \cdot dx \right|$ is less than any observable positive number which implies $\displaystyle\sum_{i=0}^{N-1} \varepsilon_i \cdot dx \simeq 0$.

One could also take $\varepsilon = \max\{|\varepsilon_i| \mid 0 \leq i \leq N\}$*, then* $\left|\sum_{i}^{N-1} \varepsilon_i \cdot dx\right| \leq \sum_{i}^{N-1} \varepsilon \cdot dx \leq \varepsilon \cdot \sum_{i}^{N-1} dx \leq \varepsilon \cdot (b-a) \simeq 0$

Hence

$$F(b) - F(a) \simeq \sum_{i=0}^{N-1} f(x_i) \cdot dx,$$

but as $F(b) - F(a)$ is observable f is thus integrable and

$$F(b) - F(a) = \int_a^b f(x) \cdot dx.$$

□

The proofs that continuous functions are integrable and that the integral is an antiderivative is more involved.

Integration rules

Integration rules are shown as usual by assuming the existence of antiderivatives. We do not show these here.

We do show, however integration by variable substitution by an example. It is a hidden use of the chain rule..

Example

Consider

$$\int_0^1 \sqrt{1+\sqrt{x}} \cdot dx.$$

Let $u = 1 + \sqrt{x}$*. (Therefore* $u - 1 = \sqrt{x}$*)*

$$\frac{du}{dx} = \frac{1}{2\sqrt{x}} = \frac{1}{2(u-1)}$$

hence

$$dx = 2 \cdot (u-1) \cdot du.$$

Then $\sqrt{1+\sqrt{x}} = \sqrt{u}$ *and* $\sqrt{1+\sqrt{x}} \cdot dx = \sqrt{u} \cdot 2 \cdot (u-1) \cdot du$

If $x = 0$ *then* $u = 1$ *and if* $x = 1$ *then* $u = 2$.

x	u
0	*1*
1	*2*

Replacing all terms we get

$$\int_0^1 \sqrt{1+\sqrt{x}} \cdot dx = 2\int_1^2 \sqrt{u} \cdot (u-1) \cdot du = 2\int_1^2 \left(u^{3/2} - u^{1/2}\right) \cdot du$$

so that

$$2\left(\frac{2}{5}u^{5/2} - \frac{2}{3}u^{3/2}\right)\Bigg|_1^2 = \frac{8+8\sqrt{2}}{15}.$$

Since g *is a one-to-one correspondence on* $[1;2]$ *whose inverse is differentiable* $x \mapsto 1 + \sqrt{x}$ *(except at* $x = 0$*), it is also possible to go back to the original variable* x *and find an antiderivative.*

$$\int \sqrt{1+\sqrt{x}} \cdot dx = \frac{4}{5}\left(\sqrt{1+\sqrt{x}}\right)^5 - \frac{4}{3}\left(\sqrt{1+\sqrt{x}}\right)^3 + C.$$

Part IX
Equivalence of definitions of Limits

Introduction

Extending the classical axioms with the additional axioms which enable to produce the concepts of ultrasmall numbers and related concepts has been proven to be a conservative extension.

A statement about observable numbers which can be expressed without reference to observability can be understood as a statement of classical mathematics once the additional structure is "forgotten". This is what "conservative extension" means.

There are, of course, statements with ultrasmall numbers which have no equivalent in classical mathematics, such as $x \simeq a$.

Reminder about contexts

The context of a statement is the list of parameters used within it. It does not contain variables linked by a $\forall$ or $\exists$: these are so called dummy variables. The statement is not about them: they are simply useful figments.

Definition 1

The function f has a limit at a if there is an L, such that

$$\forall \varepsilon > 0, \exists \delta > 0 \quad |x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

This statement is not about ε nor δ: It is about f, a and L.

Definition 2

The function f has a limit at a if there is an observable L such that

$$\forall x \quad x \simeq a \Rightarrow f(x) \simeq L$$

This statement is not about x: It is about f, a and L.

Theorem 13
Definition 1 $\iff$ *Definition 2*

Proof:

(1) $\Rightarrow$ (2)

Assume

$$\forall \varepsilon > 0, \exists \delta > 0 \quad |x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

The context is given by f, a et L.

Since the algorithm is to show the relation between any given ε and the corresponding δ, we fix ε and the context is extended to f, a, L and ε.

Then

$$\exists \delta > 0 \quad |x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

By closure, there is an observable δ satisfying the property.

hence, for any x, if $x \simeq a$, then $|x-a| < \delta$

Hence we have the following statement,

$$\forall x \quad x \simeq a \Rightarrow |f(x) - L| < \varepsilon$$

but since ε is in the context, this implies that $|f(x) - L| \simeq 0$ or $f(x) \simeq L$

Therefore

$$\forall x \quad x \simeq a \Rightarrow f(x) \simeq L$$

(2) $\Rightarrow$ (1)

Assume that

$$x \simeq a \Rightarrow f(x) \simeq L$$

The context is given by f, a and L.

Fix ε and extend the context to f, a, L and ε

Let $\delta \simeq 0$. If $|x-a| < \delta$ then $x \simeq a$. This implies that $f(x) \simeq L$ or $|f(x) - L| \simeq 0$.

But since ε is observable, we have $|f(x) - L| < \varepsilon$

So there is a δ such that

$$|x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

hence there is an observable such δ.

Therefore

$$\forall \varepsilon > 0, \exists \delta > 0 \quad |x-a| < \delta \Rightarrow |f(x) - L| < \varepsilon$$

$\square$

Answers to exercises

Answer to exercise 1, page 12

Fix a context.

(1) Let ε be ultrasmall (say positive). Let a be positive and observable.

Since ε is ultrasmall, we have $\varepsilon < \frac{1}{a}$ since, by closure, $\frac{1}{a}$ is observable. Therefore $\frac{1}{\varepsilon} > a$. Since a is arbitrary (yet observable) we have shown that $\frac{1}{\varepsilon}$ is ultralarge.

(2) Let M be ultralarge (say positive). Let a be positive and observable. By closure $\frac{1}{a}$ is observable, hence $M > \frac{1}{a}$, and$\frac{1}{M} < a$. Thus $\frac{1}{M}$ is ultrasmall.

Answer to exercise 2, page 12

Proof: Without loss of generality we consider all terms to be positive.

(1) By contradiction. Suppose $a \cdot h \geq b$ for some positive and observable b. Then we have $h \geq \frac{b}{a}$. By closure b/a is observable which contradicts that h is ultrasmall.

⚠ We suggest to stop a little while on this proof. It shows the power of the closure principle. This principle and the observable neighbour principle are the fundamental tools in this approach. Stability is also fundamental but is often used without noticing it.

(2) By contradiction: Suppose $a/h \leq b$ for some positive and observable b. Then we have $a \leq b \cdot h$ but we have just shown that $b \cdot h \simeq 0$. This contradicts that a is observable and non zero.

(3) Direct proof: Let b be some observable positive number. By closure $\sqrt{b}$ is observable and $\varepsilon < \sqrt{b}$ and $h < \sqrt{b}$ hence $\varepsilon \cdot h < b$. Thus $\varepsilon \cdot h \simeq 0$.

By contradiction: Suppose $\varepsilon \cdot h \geq b$ for some positive and observable b. Then $\varepsilon \geq \frac{b}{h}$. But we have just shown that b/h is ultralarge. This contradicts that ε is ultrasmall.

(4) Direct proof: $\varepsilon < b/2$ and $h < b/2$ for any positive and observable b. Then $\varepsilon + h < b$. Since b is arbitrary, $\varepsilon + h \simeq 0$.

Alternatively: $0 \leq |\varepsilon + h| \leq 2\max\{|\varepsilon|, |h|\} \simeq 0$

□

Answer to exercise 3, page 13

Proof:

(1) If $a \simeq x$ and $b \simeq y$ then by the observable neighbour principle there exist ε and δ ultrasmall (non zero) such that $x = a + \varepsilon$ et $y = b + \delta$. With this notation we have

$$x \pm y = (a+\varepsilon) \pm (b+\delta) = (a \pm b) + \underbrace{(\varepsilon \pm \delta)}_{\simeq 0},$$

therefore $a \pm b \simeq x \pm y$.

(2) Similarly for the product:

$$x \cdot y = (a+\varepsilon) \cdot (b+\delta) = a \cdot b + \underbrace{a \cdot \delta}_{\simeq 0} + \underbrace{b \cdot \varepsilon}_{\simeq 0} + \underbrace{\varepsilon \cdot \delta}_{\simeq 0}$$

by rule 1 hence $a \cdot b \simeq x \cdot y$.

(3) We first show that $\frac{1}{b} \simeq \frac{1}{y}$.

Since $b \neq 0$ and $y \simeq b$, y is not ultrasmall hence $\frac{1}{y}$ is not ultralarge. Let c be the observable neighbour of $\frac{1}{y}$. Then by the previous rules $c \cdot y \simeq \frac{1}{y} \cdot y = 1$ but also $c \cdot y \simeq c \cdot b \simeq 1$. Since c, b and 1 are observable, $c = \frac{1}{b}$ and we have shown that $\frac{1}{y} \simeq \frac{1}{b}$.

The general rule is obtained by combining previous results.

$$\frac{a}{b} = a \cdot \frac{1}{b} \simeq x \cdot \frac{1}{y}.$$

(4) If $a \simeq b$ then $a - b \simeq 0$. Since a and b are observable, their difference is observable. Therefore it cannot be ultrasmall, hence it is zero.

□

Answer to exercise 4, page 18

Proof: The context is a, f et g. Let $x \simeq a$ with $x \neq a$. We write $x = a + dx$). Then

$$\frac{f(a+dx)}{g(a+dx)} = \frac{f(a) + \Delta f(a)}{g(a) + \Delta g(a)} = \frac{\Delta f(a)}{\Delta g(a)} = \frac{\dfrac{\Delta f(a)}{dx}}{\dfrac{\Delta g(a)}{dx}} \simeq \frac{f'(a)}{g'(a)}.$$

□

It is also possible to use $f(a+dx) = f(a) + f'(a) \cdot dx + \varepsilon \cdot dx$ and $g(a+dx) = g(a) + g'(a) \cdot dx + \delta \cdot dx$, hence

$$\frac{f(a+dx)}{g(a+dx)} = \frac{f'(a) \cdot dx + \varepsilon \cdot dx}{g'(a) \cdot dx + \delta \cdot dx} = \frac{f'(a) + \varepsilon}{g'(a) + \delta} \simeq \frac{f'(a)}{g'(a)}$$

Part X
Particularities

What happens if contextual notation is not used.

The "$\simeq$" symbol is defined only relative to some context. Referring in a non contextual way would require to invent a new notation. The reader may find it amusing to try and invent such a notation of their own and see "pathological" objects. (see the paper "external functions" on this site)

We will not address this issue further except to say that contextual notation guarantees that objects they define are really sets, functions or properties in the usual sense and can be used in induction.

The problem of induction

Note for teachers: it is not possible to proceed by induction to show that all numbers are observable.

The reader familiar with induction may conclude that since 1 is always observable then $2 = 1 + 1$ is observable – which is true by closure, and that if n is always observable, then $n + 1$ is also always observable – which is also true also by closure. But it would be false to conclude that this proves that all natural numbers are always observable. They are not: there *are* ultralarge integers. Induction is a property which is valid for classical statements which do not use the concept of observability. We will see that we can extend induction to "contextual" statements. The statement “is always observable” is clearly not a classical statement. It is not a contextual statement either: $n+1$ is as observable as n is a statement about n (and $n + 1$) hence it cannot be used in a definition – see page 11 (Definition principle).

This aspect of induction is probably the most troubling to newcomers with a mathematical training.

References

[1] Karel Hrbacek. Stratified analysis? In I. van den Berg V. Neves, editor, *Nonstandard Methods and Applications in Mathematics*, pages 47–63. Springer, 2007.

[2] Karel Hrbacek, Olivier Lessmann, and Richard O'Donovan. Analysis with ultrasmall numbers. *American Mathematical Monthly*, 117(9), November 2010.

[3] Karel Hrbacek, Olivier Lessmann, and Richard O'Donovan. *Analysis With Ultrasmall Numbers*. CRC Press, 2015.

[4] H. Jerome Keisler. *Foundations of Infinitesimal Calculus*. University of Wisconsin, 2007.

[5] H. Jerome Keisler. *Elementary Calculus, An Infinitesimal Approach*. University of Wisconsin, 2013.

[6] Alain Robert. *Analyse non standard*. Presses polytechniques romandes, 1985.

[7] Keith D. Stroyan. *Calculus, The Language of Change*. Academic Press, 1993.

[8] Keith D. Stroyan. *Mathematical Background: Foundations of Infinitesimal Calculus*. Academic Press, 1997.