Calculus using proximities: a mathematical approach in which students can actually prove theorems

Abstract: Teaching and learning calculus are notoriously difficult and the didactic solutions may involve resorting to intuitive but vague definitions or informal gestures offered as proofs. The teaching literature is rife with examples of metaphors, adverb manipulations and descriptions of what happens “just before” the limit. It is then difficult to leave the domain of the mental image, thus losing the training in rigour. The author (with Karel Hrbacek and Olivier Lessmann) has endeavoured a radically different approach with the objective of training students to prove theorems while preserving both intuition and mathematical rigour. Hence we change the mathematical setting rather than the didactic setting. The result (which is a by-product of nonstandard analysis) has been used in several high schools in Geneva – Switzerland – for over ten years.

Keywords: Didactics, Analysis, Nonstandard analysis, Ultrasmall numbers

MSC: 03C99, 03E70, 03H05, 26A03, 26A06, 97I10, 97I40

Introduction

When instructors realise that certain mathematical subjects are difficult to teach and to learn, they usually turn to didactics to find solutions. The idea here is to change the mathematics to solve a didactic problem. This may be an unusual process but we think that the result is interesting both on the didactic and the mathematical points of view.

Of course, changing the mathematical setting will raise new questions about didactics. This also challenges some philosophical standpoints as to what mathematics is. If they are changed, are we really teaching the same subject? This question goes all the way back to the Leibniz-Newton dichotomy. We stress that this new approach does not replace but extends the classical one. Though it is necessary to acknowledge this foundational aspect, we do not wish to address these questions in full in this article. Maybe a simple example will suffice to justify the endeavour: if the interval $[0, 1]$ contains infinitely many numbers (real or rational), some of them, in some sense, must be extremely close. We may assume this as an obvious fact. The present formalisation gives a meaning to this “extreme closeness”.

Some specialists in didactics may think that the result is more mathematical than didactic. Mathematicians may say that there is too much about didactics. It is a case of mutual influence of mathematics and didactics. Please consider.
The Problem

An analysis course will deal with making the students understand the concept of limit. But how, in class, do we define a limit? What one approaches? But we know that numbers do not really move... Is it what we can be arbitrarily close to? But how close is arbitrarily close?

We have all used these metaphors of movement. But there is a circularity when we study a movement of \( y \) given by \( f \) by assuming a movement of \( x \) — and this metaphor does not resolve the fundamental difficulty that it seems that the limit may lead to a division by zero in the case of the derivative, or a sum of zeroes in the case of the integral.

Consider the concept of continuity: a topological concept of preservation of neighbourhood or proximity. The classical full definition of continuity of a function \( f \) at \( a \) is:

\[
(\forall \varepsilon > 0) \ (\exists \delta > 0) \ (\forall x) \ (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon)
\]

The concept of proximity is really well hidden in this formula! The use of several different quantifiers, their order, the fact that one chooses \( \varepsilon \) first but uses it only at the end, the fact that \( \delta \) depends on \( \varepsilon \) even though it is for any \( \varepsilon \), the fact that it is not possible to study what is on the left of the arrow independently of what is on the right: all this make this formula extremely difficult to understand when it is encountered for the first time. Hiding all technical details in \( \lim_{x \to a} f(x) = f(a) \) is similar to hiding the dust under the carpet: It looks clean but the difficulties remain. For our students (Geneva high school) the technical aspects of this definition are overwhelming and the meaning is out of reach for most.

The metaphor of “\( f(x) \) being arbitrarily close to \( f(a) \) when \( x \) is arbitrarily close to \( a \)” will not lead to divisions by zero for the derivative, but division by “arbitrarily small” values. The approach presented here may be considered as giving a more formal meaning to “arbitrarily close.”

A possible mathematical solution

The present framework was developed to be usable in class and yet remain consistent; i.e., these new concepts do not conflict with the classical approach. This was established by K. Hrbacek [1]. The idea was not to create new mathematical objects (even though it is possible) but to develop a new language for proving classical statements: the ones that appear in our analysis syllabus. This was presented in [2].

A discussion of the axioms and mathematical principles involved can be found on the website http://ultrasmall.org. There is also a teacher’s manual and student handouts. A discussion of this approach has been given in [3] and the pedagogical aspect is addressed in [4].

For this approach, a useful mental image is one of scales of observation. The usual numbers (defined without using the concept of observability) are those that can be seen without microscope or telescope. Then there are ultrasmall numbers (non zero numbers smaller in absolute value than any strictly positive observable number). A microscope is needed to see these.

And if one zooms on an observable number, say 1, then numbers ultraclose to 1 will appear (noted in the figure below: \( 1, 1 + \varepsilon \) etc., for \( \varepsilon > 0 \) and very very small — we will say “ultrasmall”). It should be noticed that although more numbers are now observable, 1 is still observable.

But the microscope view is not the end of the story: the same happens all over again. We can zoom in again on \( 1 + \varepsilon \) and see yet more numbers (note \( 1 + \varepsilon + h \), etc)

---

1 We will consider that there are ultrasmall real numbers. There are several philosophical interpretations and these have been discussed widely in the nonstandard literature: one view recognises that there are no new objects, that ultrasmall numbers were there all along but our syntax was unable to distinguish them. Another interpretation follows the view that new numbers are added to the field of real numbers and that the “usual” ones are the standard numbers. Both interpretations are valid and change nothing to the mathematics. The author, after having used them for several years, has come to consider the first interpretation as natural.
There is no end to how many times we can zoom in: this is the major difference with other nonstandard methods. Depending on what point we are looking at and its level of observability, there are ultrasmall numbers relative to this observability. The parameters used in a definition determine its context: observability is always to be considered in reference to the context of a given statement.

These concepts may be made quite intuitive and they are completely formalised with their properties proven in [2]. Once these concepts have been established, no more “black boxes” are needed and the whole calculus course can be done in a deductive way.

We will give some examples (with the relevant definitions) to show how it works in class. The reader must assume that all words will have been precisely defined. The point here is not to make the reader an expert of the subject but rather a witness to the situation.

**Main properties**

The figure given above characterises the way observability stratifies the set of real numbers. We notice that when we can “see” $\varepsilon$, we can still “see” 1 even though at the original scale, we could see 1 but not $\varepsilon$. We formalise this first by defining the context: it is the list of parameters used in the definition, set or property. All concepts of observability are relative to a given context.

**Observability:**

1. Given two numbers $x$ and $y$. There always is a common context: if $x$ is not observable in the context of $y$ then $y$ is observable in the context of $x$.
2. The results of operations between $x$ and $y$ are in their common context.
3. If a number satisfies a given property, then there is an observable number satisfying that property.
4. Relative to any context, there are strictly positive numbers smaller than any strictly positive observable number (they are ultrasmall).
5. If a number is not ultralarge, then it is ultraclose to an observable real number: its observable neighbour.

The reciprocal of an ultrasmall is ultralarge and if the difference between two numbers $a$ and $b$ is ultrasmall or zero, then $a$ and $b$ are ultraclose, noted $a \simeq b$.

**New rule:**

In addition to classical statements which do not refer to observability, statements may refer to observability only by the use of the $\simeq$ symbol – which depends the context.

The term ultralarge will be used rather than infinitely large because all the numbers here are real numbers and while not contradicting that all natural numbers are finite, some are ultralarge. Their reciprocals are ultrasmall. Yet they have the “flavour” of infinitely large numbers and infinitesimals.
Examples

Continuity

The formalism used here expresses that continuity preserves proximity: a topological characterisation.

A function \( f \) is continuous at \( a \) if

\[
\forall x \quad x \simeq a \Rightarrow f(x) \simeq f(a).
\]

This reads: If \( x \) is ultraclose to \( a \), then \( f(x) \) is ultraclose to \( f(a) \). The quantifier is explained as meaning that the proximity of \( f(x) \) and \( f(a) \) must not depend on the choice of \( x \), provided it is ultraclose to \( a \). The “\( \simeq \)” refers to observability which is given by the context: in this case, the parameters used in the description of \( f \) and \( a \).

We now show a proof – as done in class with the students – of one of the theorems about continuity.

**If \( g \) is continuous at \( a \) and \( f \) is continuous at \( g(a) \); then \( f \circ g \) is continuous at \( a \).**

**Proof.** Let \( x \neq a \). Then \( g(x) \simeq g(a) \) by continuity of \( g \) at \( a \) and \( f(g(x)) \simeq f(g(a)) \) by continuity of \( f \) at \( g(a) \).

This proof is complete. It should be compared with the classical statement which starts by the rather awkward

\[
\lim_{x \to a} f(g(x)) = f(\lim_{x \to a} g(x)) = f(g(a))
\]

which involves showing that the limit passes inside.

Limits

Continuity being defined here without reference to the limit, the limit of \( f \) at \( a \) can be described as the value that \( f \) should take to be continuous at \( a \). Hence \( f \) has a limit at \( a \) if there is an observable number \( L \) such that whenever \( x \simeq a \) we have \( f(x) \simeq L \).

Derivatives

A function \( f \) is differentiable at \( a \) if there is an observable number \( d \) such that for any ultrasmall \( \Delta x \), we have

\[
\frac{f(a + \Delta x) - f(a)}{\Delta x} \simeq d.
\]

We write \( f'(a) = d \).

We continue with the derivative of the composition. The theorem is:

**If \( g \) is differentiable at \( a \) and \( f \) is differentiable at \( g(a) \); then \( f \circ g \) is differentiable at \( a \) and \( (f \circ g)'(a) = f'(g(a)) \cdot g'(a) \).**

**Proof.** Let \( \Delta g(a) = g(a + \Delta x) - g(a) \), then \( \frac{\Delta g(a)}{\Delta x} \simeq g'(a) \) since \( g \) is differentiable at \( a \). We have that \( \Delta g(a) \simeq 0 \) since \( \Delta g(a) \simeq g'(a) \cdot \Delta x \simeq 0 \) (in particular, this shows the continuity of a differentiable function).

Also \( g(a + \Delta x) = g(a) + \Delta g(a) \) by definition of \( \Delta g(a) \).

If \( \Delta g(a) \neq 0 \) then \( \Delta g(a) \) is ultrasmall and

\[
\frac{f(g(a + \Delta x)) - f(g(a))}{\Delta x} = \frac{f(g(a + \Delta g(a)) - f(g(a))}{\Delta x} \cdot \frac{\Delta g(a)}{\Delta x} \simeq f'(g(a)) \cdot g'(a)
\]

If \( \Delta g(a) = 0 \), then it is immediate that \( g'(a) = 0 \) and \( \frac{f(g(a + \Delta x)) - f(g(a))}{\Delta x} = 0 = f'(g(a)) \cdot g'(a) \). The formula is true in any case. \( \square \)
This proof would not be the first one in the presentation of the derivative, but a good proportion of the students can work it out alone given the following hint:

Write \( g(a) = y \) and \( \Delta g(a) = \Delta y \) which leads to the “Leibnizian” notation

\[
\frac{f(y + \Delta y) - f(y)}{\Delta y} \cdot \frac{\Delta y}{\Delta x}
\]

Graphical Representation of the Derivative

The following sketch shows the fundamental relations involved in calculating a derivative. It would be obtained by a zoom of factor \( 1/\Delta x \). The quantities shown here actually exist. If we define \( df(a) = f'(a) \cdot \Delta x \) (the first order approximation) then when writing \( f'(a) = \frac{df(a)}{\Delta x} \), we do really write a quotient, not the limit of a quotient which is not a quotient any more.

The differential

Using the notation convention that for ultrasmall \( \Delta x \) we write \( dx \), then \( y' = \frac{dy}{dx} \) which is a classical writing. And since \( dx \) and \( dy \) are specific values, separating the variables in a differential equation amounts to straightforward algebraic manipulations.

The Integral

Let \( f \) be a positive function, continuous on \([a, b]\). Consider the area of the surface bounded above by \( f \), below by the \( x \)-axis and left and right by the vertical lines at \( a \) and at \( x \). Assume there exists a function \( A(x) \) which measures this area (\(^1\)). Assume that we have already shown that a continuous function on a closed interval reaches its maximum and its minimum. So on \([x, x + \Delta x]\) the area of the rectangle is \( \Delta A(x) \) and we have \( f(x_{\text{min}}) \cdot \Delta x \leq \Delta A(x) \leq f(x_{\text{Max}}) \cdot \Delta x \). Dividing leads to \( f(x_{\text{min}}) \leq \frac{A(x)}{\Delta x} \leq f(x_{\text{Max}}) \). We have \( x_{\text{min}} \simeq x \simeq x_{\text{Max}} \) hence \( \frac{A(x)}{\Delta x} \simeq f(x) \). The same holds for negative \( \Delta x \), hence \( A'(x) = f(x) \) (\(^3\)).

---

\(^1\) At this stage, it is common to assume the existence of such a function and postpone the discussion of what this area really is for later – if and when students specialise in mathematics and do some measure theory.

\(^3\) Of course, the drawing is not correct: if \( a \) and \( b \) are observable and distinct, then \( x \) and \( x + \Delta x \) cannot be distinguished. Drawings are here – as frequent – to help the mind.
For more advanced students, a more rigorous definition will be needed:

Let $f$ be defined between $a$ and $b$, let $N$ be an ultralarge positive natural number. Then let $\Delta x = \frac{b-a}{N}$ (hence $\Delta x$ is ultrasmall). Define $x_k = a + k \cdot \Delta x$. Then if there is an observable number $I$ such that (for any ultralarge $N$)

$$
\sum_{k=0}^{N-1} f(x_k) \cdot \Delta x \simeq I
$$

then $I = \int_a^b f(x)dx$.

The integral is the observable neighbour of an ultralarge sum of ultrathin slices, not the limit when there are no slices any more.($^4$)

### Comparison with other nonstandard approaches

#### Acceptable statements

Adding the extra predicate of observability can, in principle, allow for statements which contradict classical theorems. In nonstandard analysis, these are called external statements and are to be avoided at introductory level since they introduce what would be considered pathological objects. Internal statements are the ones that we should refer to. In most nonstandard approaches, determining whether a statement is internal is a crucial and sometimes complicated issue. Here, thanks to multiplicity of levels of observability and the concept of context, even statements that do refer to observability are internal, as long as they take it relative to their context. So while in most nonstandard approaches the usual definitions of continuity, derivative, etc., are external, in our approach they are internal.

The “$\simeq$” symbol is the only new symbol introduced. By defining “ultraclose” to refer to the context, the notation is “rigged” in such a way that it is almost impossible to write an external statement without inventing an extra notation.

In short: an internal statement is either a classical statement of mathematics or one which can be shown to be equiconsistent with these classical statements. A statement which uses only the usual symbols or the additional symbol “$\simeq$” is internal.

We give here an example of an external statement. This question would not be addressed in class unless a student asked deeper questions – and this has not happened yet.

In nonstandard approaches with two levels (standard and nonstandard) (see [5–8]) there is a $\text{st}(\cdot)$ predicate for the standard part. It is an external predicate but is used in the definition of the derivative thus making the derivative an external statement. Consider the rule $x \mapsto \text{st}(x)$. If this defined a function, then zooming on the graph we would see a horizontal line on any ultrasmall neighbourhood (all points on an ultrasmall interval have the same observable neighbour.) There is no value where we could point to a discontinuity yet this everywhere horizontal “continuous” graph (if it exists) is increasing!

These nonstandard approaches must therefore deal with the difficult problem of justifying when external concepts can be used (in defining the derivative) and when they cannot.

In this approach, where observability is relative to the context, the derivative for instance, is given by the observable neighbour of the quotient $(f(a + \Delta x) - f(a))/\Delta x$, where the observability depends on $f$ and on $a$. The ultrasmallness of $\Delta x$ also refers to that context. (Because $\Delta x$ is bounded by a quantifier, it is not part of the context: this is not about $\Delta x$: as in any classical explanation, it is a dummy variable.) The derivative is now an internal object. The “new rule” does not allow the construction of a function such as the one above where $\text{st}$ would be “observable relative to 1”.

---

$^4$ If $f$ is continuous on $[a,b]$, then this integral is equivalent to the Riemann integral. A proof can be found in [2].
Pedagogical comments

From the teacher’s point of view

In most cases, introductory calculus is presented without the full formalism which is considered so difficult that too many students would be left behind; the real meaning being obscured by the technicalities. The resulting informal presentations contain some hand waving and many metaphors, such as $x$’s moving towards... For the mathematically trained teacher, this is frustrating because we are not showing the core of our passion and subject which is the argumentation and rigour of proof. Furthermore, when proofs are involved, they rely on a mixture of non-rigorous definitions, some informal explanations and some rigorous steps. Deciding which parts should be formal rests with the teacher and may amount to proof by intimidation.

In the approach we developed, the teacher must first learn new mathematical rules. But understanding a new “dialect” to describe already known notions sheds new light and the general understanding is usually improved. Colleagues who have chosen to teach calculus with ultrasmall numbers find it much more satisfactory as a teaching method. Physicists find it appealing that $\Delta x$ “is a tiny vector”.

Some of the advantages are linked to visualisation: the derivative is the observable neighbour of the slope of a segment between two ultraclose points on a curve, continuity tells us that we are ultraclose to where we expect to be and the integral is the observable neighbour of a sum of ultrathin slices.

From the student’s point of view

In class, the presentation is completely deductive and most proofs can be discovered by students after preliminary exercises.

Students are prepared for further studies in that we do give a definition of the limit after the main concepts have been understood. Thus all definitions written with the “$\varepsilon$” can be rewritten with limits. If necessary, with more advanced students, it is even possible to show the equivalence with the classical $\varepsilon$-$\delta$ definition of the limit.

We have met our students after they went to university and they claim that it is not difficult for them to adapt to another rigorous definition. For other students it was hard work to remedy the lack of formal definitions.

References