Article Submitted to Journal of Symbolic Computation

The meaning of infinity in calculus and computer algebra systems

Michael Beeson[1] and Freek Wiedijk[2]

[1] *San José State University*
[2] *University of Nijmegen*

Abstract

We use filters of open sets to provide a semantics justifying the use of infinity in informal limit calculations in calculus, and in the same kind of calculations in computer algebra. We compare the behavior of these filters to the way Mathematica behaves when calculating with infinity.

We stress the need to have a proper semantics for computer algebra expressions, especially if one wants to use results and methods from computer algebra in theorem provers. The computer algebra method under discussion in this paper is the use of rewrite rules to evaluate limits involving infinity.

1. Introduction

1.1. Problem

In calculus, when calculating limits, one often first uses the heuristic of 'calculating with infinity' before trying to evaluate the limit in a more formal way. For instance one 'calculates':

$$\lim_{x\to\infty} \frac{1}{x+1} = \frac{1}{\infty+1} = \frac{1}{\infty} = 0$$

which indeed gives the correct answer. However, it is not clear what the *meaning* of this use of the symbol '∞' is, and why this method works. This problem arises in calculus textbooks, which usually avoid examples of such calculations for fear of 'lack of rigor', although students are taught these methods at the blackboard. It arose in the design of the first author's software, MathXpert (1; 2; 3). This software, which is designed to assist a student in producing step-by-step solutions to calculus problems, had to be able to produce 'ideal' step-by-step solutions of limit problems. Are such 'ideal solutions' allowed to use calculations involving infinity? Or are those calculations just private preliminary considerations intended to guide a rigorous proof? MathXpert does allow calculations involving infinity, but not the full system justified in this paper, since that goes beyond what one finds in calculus textbooks.

In the Mathematica system (9) the approach of calculating with infinity is used. Since Mathematica gives answers, rather than step-by-step solutions, one will not notice the calculations with infinity, in cases where the limit turns out to exist (and be a finite number). But in fact, in Mathematica there is a complete 'calculus of infinity' (and some related symbols):

```
In[1]:= 1/(Infinity + 1)
Out[1]= 0
In[2]:= Sqrt[Infinity]
Out[2]= Infinity
In[3]:= Infinity - Infinity
Infinity::indet:
   Indeterminate expression (-Infinity) + (Infinity) encountered.
Out[3]= Indeterminate
In[4]:= Indeterminate + Infinity
Out[4]= Indeterminate
In[5]:= Sin[Infinity]
Out[5]= Interval[{-1, 1}]
In[6]:= 1/Interval[{-1, 1}]
Out[6]= Interval[{-Infinity, -1}, {1, Infinity}]
In[7]:= Interval[{-1, 1}]*Interval[{-1, 1}]
Out[7]= Interval[{-1, 1}]
In[8]:= Interval[{-1, 1}]^2
Out[8]= Interval[{0, 1}]
In[9]:= 0*Sin[Infinity]
Out[9]= Interval[{0, 0}]
In[10]:= Infinity/Sin[Infinity]
Out[10]= Interval[{-Infinity, -Infinity}, {Infinity, Infinity}]
In[11]:= Infinity/Sin[Infinity]^2
Out[11]= Interval[{Infinity, Infinity}]
```

Other computer algebra systems implement similar calculi. For instance, the Maple system (6) uses the symbols `infinity` and `undefined` in answers to limit problems.*

*There is also some notion of interval in Maple, written as `1 .. 2`, but our attempts to calculate with these terms led only to error messages. These terms seem primarily to be used for generating integer sequences, although the answer to $\lim_{x\to\infty} \sin x$ comes out as `-1 .. 1`.

It is well known that many computer algebra packages make errors. One of the reasons for that is that they fail to check the pre-conditions or 'side conditions' that must be satisfied for a simplification rule to be applicable. For example, before applying $\sqrt{x^2} = x$ we need to check that $x \geq 0$. Systematically keeping track of such assumptions is difficult. The errors in computer algebra systems sometimes give the impression that those systems place a higher priority on performing as many simplifications as possible than on ensuring that only correct computations are performed. Generally, 'evaluation errors' which users complain about are taken care of on an ad hoc basis only, to get rid of the most embarrassing ones.

Related to these errors is the fact that these systems have no unified semantics for their expression language. In this paper we focus on the apparatus for limits and offer a solution: a semantics explaining and supporting the use of infinities in limit calculations. We will present a formal semantics of limits, which not only explains the calculations usually performed with infinities, but offers some extensions by introducing some other symbols for common ways in which a function can fail to have a limit. Thus, we will be able to get an answer by calculation for such a limit as $\lim_{x\to\infty} 1/(2 + \sin x)$ which will be 'oscillations through the interval $[\frac{1}{3}, 1]$'. We then compare the resulting semantics to the behavior of Mathematica as illustrated above. There is a rough general correspondence, and our semantics agrees with some of the examples above, but in some instances Mathematica does give incorrect answers, and in some cases we are able to distinguish between identical Mathematica expressions which are different in our semantics.

1.2. Approach

We will represent ∞ and its cousins indeterminate and interval by *filters* over some underlying topological space (which in calculus textbooks and Mathematica is the space of real numbers, but could also be the complex numbers or more general spaces). For each point of the space there will be a filter associated with it, which is called the principal filter of the point. For each function on the space there will be a *lifted* version that works on the filters instead of on the points.

Furthermore we will define classes of filters called the *interval* filters and the *connected* filters. It will turn out that those two classes coincide and that connectedness of filters is preserved under continuous mappings. Also we will define the *join* and the *meet* of two filters.

It turns out that the calculus used in Mathematica corresponds directly to the set of finite joins of interval filters.

1.3. Related work

First, in topology, the two standard approaches for defining limits in topological spaces make use of *nets* or *filters*. There is therefore nothing original in the use of filters to analyze the notion of limits. However, our focus to use them in an applied setting, and identify specific filters associated with 'extra-mathematical' symbols such as ∞, seems to be new.

Second, the interval filters are directly related to the active field of *interval arithmetic*. We throw a new light on the calculations with intervals by looking at them as filters.

Third, justifying 'calculations with infinite objects' rigorously is close to doing the same with 'calculations with infinitesimal objects', which is the domain of *nonstandard analysis*. In nonstandard analysis one also has infinity as a first class citizen, but in nonstandard analysis there is not *one*, designated, infinity; instead there are many infinite nonstandard numbers, without a 'canonical' one. In our case there *is* a canonical infinity. To illustrate this difference concretely, let ω be the infinity of nonstandard analysis and let ∞ be the infinity of our filter calculus. Then we have $\omega + 1 \neq \omega$, but $\infty + 1 = \infty$.

The simplest way to get non-standard objects also employs a filter, but this has to be a special kind of filter called an *ultrafilter*. Also this filter contains arbitrary sets instead of only open sets (as is the case for the filters that are studied in this paper). Another important difference is that, in contrast to our approach, in this construction there is only *one* filter involved. The non-standard objects in this construction are not filters themselves.

Nonstandard analysis has been used in (4) to help in the computation of limits in a computer algebra system.

2. Filters, lifting, refinement and limits

Definition: Let X be a topological space. Denote the open sets of X by $\mathcal{O}(X)$. A *filter on* X is a set $A \subseteq \mathcal{O}(X)$ that satisfies:

$$\forall U \in A.\, \forall V \in \mathcal{O}(X).\, U \subseteq V \Rightarrow V \in A$$
$$\forall U \in A.\, \forall V \in A.\, U \cap V \in A$$

In words: a filter is a set of open sets that is closed under supersets and finite intersections. The collection of filters on X is written $\bar{X}$.

A filter that does not contain the empty set is called *proper*. A filter that does not contain any set at all is called *empty*.

Often the property of being proper is made part of the definition of a filter. We did not do this, because then we would be unable to define the notion of *meet* on page 11 below. Sometimes the property of being non-empty is made part of the definition of a filter too. However the empty filter, which is called domain-error below, is essential to our application. We found variants of the definition of filter in the literature, both allowing for improper (5) and for empty (7) filters. Therefore we feel free to define the concept of filter to suit our purposes.

In the topological literature a filter is generally not defined on a topological space but on an arbitrary set. In that case the restriction to open sets is not present. However, for our application it is more natural to restrict ourselves to filters of open sets.

Definition: Here are some common filters on the real numbers, where $a \in \mathbb{R}$ is an arbitrary real number:

$$\begin{aligned}
\mathsf{improper} \equiv \dagger &\equiv \mathcal{O}(\mathbb{R}) \\
\mathsf{domain\text{-}error} \equiv \bot &\equiv \emptyset \\
\mathsf{indeterminate} \equiv \leftrightarrow &\equiv \{\mathbb{R}\} \\
\mathsf{principal}(a) \equiv \bar{a} &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid a \in U\} \\
&= \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (a-\varepsilon, a+\varepsilon) \subseteq U\} \\
\mathsf{left}(a) \equiv a^{-} &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (a-\varepsilon, a) \subseteq U\} \\
\mathsf{right}(a) \equiv a^{+} &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (a, a+\varepsilon) \subseteq U\} \\
\mathsf{punctured}(a) \equiv a^{\pm} &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (a-\varepsilon, a) \cup (a, a+\varepsilon) \subseteq U\} \\
\mathsf{infinity} \equiv \infty &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (1/\varepsilon, \infty) \subseteq U\} \\
\mathsf{minus\text{-}infinity} \equiv -\infty &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (-\infty, -1/\varepsilon) \subseteq U\} \\
\mathsf{bi\text{-}infinity} \equiv \pm\infty &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists \varepsilon \in \mathbb{R}_{>0}.\, (-\infty, -1/\varepsilon) \cup (1/\varepsilon, \infty) \subseteq U\} \\
\mathsf{positive} \equiv \rightarrow &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid (0, \infty) \subseteq U\} \\
\mathsf{negative} \equiv \leftarrow &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid (-\infty, 0) \subseteq U\} \\
\mathsf{non\text{-}zero} \equiv \pm\!\!\rightarrow &\equiv \{U \in \mathcal{O}(\mathbb{R}) \mid (-\infty, 0) \cup (0, \infty) \subseteq U\}
\end{aligned}$$

For each of these filters we have a 'long' and a 'short' notation. The first four filters can be defined for any topological space. The other filters have analogues in any order topology.

Definition: Let again X be a topological space. Let A be a collection of subsets of X (not necessarily open) that satisfies:

$$\forall U \in A.\, \forall V \in A.\, \exists W \in A.\, W \subseteq U \cap V \qquad (*)$$

The *filter generated by* A is defined to be:

$$\mathsf{generated\text{-}by}(A) \equiv \{U \in \mathcal{O}(X) \mid \exists V \in A.\, V \subseteq U\}$$

The collection of sets A is called the *basis* of the filter $\mathsf{generated\text{-}by}(A)$.

Being closed under finite intersections implies $(*)$. If all elements of A are open sets, the filter generated by A is the intersection of all filters that contain A as a subset.

The filters defined on page 5 can be defined more naturally using the notion of a generated filter. For instance, we have:

$$\begin{aligned}
\mathsf{improper} &= \mathsf{generated\text{-}by}(\{\emptyset\}) \\
\mathsf{principal}(a) &= \mathsf{generated\text{-}by}(\{\{a\}\}) \\
\mathsf{right}(a) &= \mathsf{generated\text{-}by}(\{(a, a+\varepsilon) \mid \varepsilon \in \mathbb{R}_{>0}\} \\
\mathsf{infinity} &= \mathsf{generated\text{-}by}(\{(1/\varepsilon, \infty) \mid \varepsilon \in \mathbb{R}_{>0}\}
\end{aligned}$$

All other filters that are introduced on page 5 can be defined in a similar way.

Definition: Let $f : X \to X$ be some (possibly partial) function with domain $\mathrm{dom}(f)$. The *lift of* f is a function $\bar{f} : \bar{X} \to \bar{X}$, defined by:

$$\bar{f}(A) \equiv \textsf{generated-by}(\{f[U] \mid U \subseteq \mathrm{dom}(f) \wedge U \in A\})$$

This definition can be generalized to arbitrary arities. The function $\bar{f} : \bar{X} \times \bar{X} \times \ldots \times \bar{X} \to \bar{X}$ is defined by:

$$\begin{aligned}&\bar{f}(A_1, A_2, \ldots, A_n) \equiv \\ &\quad \textsf{generated-by}(\{f[U] \mid U \subseteq \mathrm{dom}(f) \wedge U = U_1 \times U_2 \times \ldots \times U_n \wedge \\ &\qquad\qquad\qquad\qquad\qquad\qquad U_1 \in A_1 \wedge U_2 \in A_2 \wedge \ldots \wedge U_n \in A_n\})\end{aligned}$$

Although f can be a partial function, the lift of f is always total. One can get rid of the problems of partial functions in calculus by lifting the whole theory to filters. In some sense by going to filters we are adding a 'bottom element' $\perp$ to the values of the theory. Looked at in this way, we have a *strict* partial logic, because a function applied to $\perp$ will always give $\perp$ again.

Note also that the definitions of $\bar{a}$ as a principal filter and as lift of a 0-ary constant function coincide. This justifies using one notation for both.

From now on we will often write f instead of $\bar{f}$ when one or more of the arguments of f are filters. So we will write $\sin(A)$ instead of $\overline{\sin}(A)$. This will allow us to write things like $\sqrt{A}$, and mean the lift of the square root function.

To state this convention more precisely: if $t[x_1, \ldots, x_n]$ is a term that does not involve filters (so $x_1, \ldots, x_n$ are variables ranging over the ordinary reals) then $t[A_1, \ldots, A_n]$ will mean the lift of the function $\lambda x_1 \cdots x_n.\, t[x_1, \ldots, x_n]$ applied to the filters $A_1, \ldots, A_n$. Note that with this convention $1/A$ means something different from $\bar{1}/A$. The first is the lift of the unary function $\lambda x.\, 1/x$ applied to A. The second is the lift of the binary function $\lambda x\, y.\, x/y$ applied to $\bar{1}$ and A. Those are not necessarily equal: $1/1^+ = 1^-$ but $\bar{1}/1^+ = \bar{1}$.

Definition: The *filter limit* of the function $f : X \to X$ when taking the limit to the filter A is defined to be:

$$\operatorname*{Lim}_{x \to A} f(x) \equiv \bar{f}(A)$$

We distinguish a filter limit from an ordinary limit by writing 'Lim' with a capital letter L. Note that the filter limit is always defined, even for non-continuous f. It might seem silly to introduce a new notation for this when we already have defined lifting, as it is the same operation. However, now we can write:

$$\operatorname*{Lim}_{x \to 0^+} x/x$$

which is something different from

$$0^+ / \, 0^+$$

The first is the lift of the unary function $\lambda x.\, x/x$ applied to 0^+ and has as value $\bar{1}$. The second is the lift of the binary function $\lambda x\, y.\, x/y$ applied to the pair $(0^+, 0^+)$ and has as value $\to$.

Definition: A filter A *refines* a filter B, notation $A \sqsubseteq B$ when $A \supseteq B$ as collections of open sets. When the two filters A and B differ we write $A \sqsubset B$.

Here are some refinement relations between the filters defined on page 5. For any proper and non-empty filter A we have:

$$\dagger \sqsubset A \sqsubseteq \leftrightarrow \sqsubset \bot$$

At any real number $a \in \mathbb{R}$ we have:

$$a^-, a^+ \sqsubset a^\pm \sqsubset \bar{a}$$

and the 'infinite' filters are related by:

$$-\infty, \infty \sqsubset \pm\infty \sqsubset \leftrightarrow, \quad -\infty \sqsubset \leftarrow \sqsubset \pm\!\rightarrow \sqsubset \leftrightarrow, \quad \infty \sqsubset \rightarrow \sqsubset \pm\!\rightarrow \sqsubset \leftrightarrow$$

Note that the filters defined on page 5 are not the only ones. There are many 'wild' filters refining $\bar{a}$ and ∞. For instance the filter generated by the sets $\{2\pi n \mid n > N\}$ is a filter which refines ∞. It has the property that the filter limit of sin to this filter is $\bar{0}$.

We can now state the first theorem, which lists some of the many calculation rules that one needs for arithmetic on filters:

THEOREM 2.1: *Let $a \in \mathbb{R}_{>0}$ be some positive real number. Then:*

$$\begin{array}{rclrclrclrcl}
\infty + \bar{a} & = & \infty & \bar{a}/\bar{0} & = & \bot & \bar{a}/\infty & = & 0^+ & \bar{0}\,\infty & = & \leftrightarrow \\
\infty - \bar{a} & = & \infty & \bar{a}/0^+ & = & \infty & \bar{a}/\pm\infty & = & 0^\pm & 0^+\infty & = & \rightarrow \\
\infty + \infty & = & \infty & \bar{a}/0^\pm & = & \pm\infty & \bar{a}/\rightarrow & = & \rightarrow & 0^\pm\infty & = & \pm\!\rightarrow \\
\infty - \infty & = & \leftrightarrow & 0^+/0^+ & = & \rightarrow & \bar{a}/\leftrightarrow & = & \bot & \infty\,\infty & = & \infty
\end{array}$$

Proof: We will show only the proof for the next to last calculation rule, $0^\pm\infty = \pm\!\rightarrow$. The other proofs are similar. To prove this rule, we need to show that both $0^\pm\infty \subseteq \pm\!\rightarrow$ and $0^\pm\infty \supseteq \pm\!\rightarrow$.

So suppose that we have $V \in 0^\pm\infty$. Then by the definition of lift of the multiplication function there are $U_1 \in 0^\pm$ and $U_2 \in \infty$ such that $U_1U_2 \subseteq V$. By the definitions of $0^\pm$ and ∞ we have for some positive ε_1 and ε_2 that $(-\varepsilon_1, 0) \cup (0, \varepsilon_1) \subseteq U_1$ and $(1/\varepsilon_2, \infty) \subseteq U_2$. Now to show that $V \in \pm\!\rightarrow$, we need to show that for any $x \neq 0$ we have that $x \in V$. Consider some positive ε smaller than ε_1 and ε_2/x. Then $x \in V$ follows, because we have that $\varepsilon \in U_1$, and because by $x/\varepsilon > x/(\varepsilon_2/x) = 1/\varepsilon_2$ we have that $x/\varepsilon \in U_2$.

For the other inclusion, suppose that $V \in \pm\!\rightarrow$, which by definition means that $(-\infty, 0) \cup (0, \infty) \subseteq V$. Take for U_1 and U_2 both $(-\infty, 0) \cup (0, \infty)$ too, then $U_1 \in 0^\pm$ and $U_2 \in \infty$. Clearly $U_1U_2 = (-\infty, 0) \cup (0, \infty)$ as well, so $U_1U_2 \subseteq V$, and therefore $V \in 0^\pm\infty$. $\square$

Note that, although the lift of division is a total function, 'division by zero' is still not allowed in a sense, because the result of $\bar{a}/\bar{0}$ is domain-error. This is essentially different from the way that Mathematica behaves. We will come back to this in Section 4

The next theorem tells us how to evaluate the lift of a continuous function in a point:

THEOREM 2.2: *Let f be a function that is continuous and monotonically increasing in a neighborhood of a. Then:*

$$\bar{f}(\bar{a}) = \overline{f(a)}, \quad \bar{f}(a^{\pm}) = f(a)^{\pm}, \quad \bar{f}(a^{-}) = f(a)^{-}, \quad \bar{f}(a^{+}) = f(a)^{+}$$

Proof: We will show only the proof of the last equality. The other proofs are similar. To prove that $\bar{f}(a^+) = f(a)^+$ we need to show that both $\bar{f}(a^+) \subseteq f(a)^+$ and $\bar{f}(a^+) \supseteq f(a)^+$.

So suppose that $V \in \bar{f}(a^+)$. Then there is a $U \in a^+$ such that $f[U] \subseteq V$, and because $U \in a^+$ there is a positive ε with $(a, a+\varepsilon) \subseteq U$. We may assume that ε is small enough that f is continuous and monotonically increasing on $[a, a+\varepsilon]$, and this implies $f[(a, a+\varepsilon)] = (f(a), f(a)+\delta)$ with $\delta = f(a+\varepsilon) - f(a)$. Therefore $(f(a), f(a)+\delta) \subseteq f[U] \subseteq V$, and hence $V \in f(a)^+$.

For the other inclusion, suppose that $V \in f(a)^+$, so for some positive δ it holds that $(f(a), f(a)+\delta) \subseteq V$. Because f is continuous in a there is a positive ε such that $f[(a-\varepsilon, a+\varepsilon)] \subseteq (f(a)-\delta, f(a)+\delta)$. Again we may assume that ε is small enough that f is monotonically increasing on $[a-\varepsilon, a+\varepsilon]$, which gives $f[(a, a+\varepsilon)] \subseteq (f(a), f(a)+\delta)$, and therefore $f[(a, a+\varepsilon)] \subseteq V$. Because $(a, a+\varepsilon) \in a^+$, this implies that $V \in \bar{f}(a^+)$. □

Similar theorems hold for decreasing functions and functions at a local maximum or minimum.

The next theorems show how to evaluate filter limits:

THEOREM 2.3: *Bringing filter limits inside expressions:*

$$\operatorname*{Lim}_{x \to A} f(g_1(x), g_2(x), \ldots, g_n(x)) \sqsubseteq \bar{f}(\operatorname*{Lim}_{x \to A} g_1(x), \operatorname*{Lim}_{x \to A} g_2(x), \ldots, \operatorname*{Lim}_{x \to A} g_n(x))$$

Proof: Define $h(x) = f(g_1(x), \ldots, g_n(x))$. To prove the statement we need to show that $\bar{h}(A) \sqsubseteq \bar{f}(\bar{g_1}(A), \ldots, \bar{g_n}(A))$, which amounts to $\bar{f}(\bar{g_1}(A), \ldots, \bar{g_n}(A)) \subseteq \bar{h}(A)$

So suppose $W \in \bar{f}(\bar{g_1}(A), \ldots, \bar{g_n}(A))$. That means that there are $V_i \in \bar{g_i}(A)$ with $V_1 \times \cdots \times V_n \subseteq \mathrm{dom}(f)$ and $f[V_1 \times \cdots \times V_n] \subseteq W$. This implies that there are $U_i \in A$ with $U_i \subseteq \mathrm{dom}(g_i)$ and $g_i[U_i] \subseteq V_i$. Now define $U = U_1 \cap \cdots \cap U_n$, then $U \in A$, $U \subseteq \mathrm{dom}(g_i)$ and $g_i[U] \subseteq V_i$, which then implies that $U \subseteq \mathrm{dom}(h)$ and $h[U] \subseteq W$. Therefore $W \in \bar{h}(A)$. □

Note that this theorem also holds for non-continuous f.

As an example of the fact that we do not always have equality here, not even when all functions are continuous, consider:

$$\begin{gathered}\mathrm{Lim}_{x\to\infty}(x-x) = \mathrm{Lim}_{x\to\infty} 0 = \bar{0} \\ (\mathrm{Lim}_{x\to\infty} x) - (\mathrm{Lim}_{x\to\infty} x) = \infty - \infty = \leftrightarrow\end{gathered}$$

This agrees with the theorem, since $\bar{0} \sqsubseteq \leftrightarrow$.

Theorem 2.4: *Monotonicity with respect to refinement:*

$$A_1 \sqsubseteq B_1, A_2 \sqsubseteq B_2, \ldots, A_n \sqsubseteq B_n \Rightarrow \bar{f}(A_1, A_2, \ldots, A_n) \sqsubseteq \bar{f}(B_1, B_2, \ldots, B_n)$$

Proof: Let V be given with $V \in \bar{f}(B_1, \ldots, B_n)$. Then there are $U_i \in B_i$ with $U_1 \times \cdots \times U_n \subseteq \mathrm{dom}(f)$ and $f[U_1 \times \cdots \times U_n] \subseteq W$. Because $A_i \sqsubseteq B_i$, also $U_i \in A_i$, which implies that also $V \in \bar{f}(A_1, \ldots, A_n)$. □

Together these two theorems allow one to evaluate a filter limit 'up to refinement' by substituting the filter inside the expression. Often this refinement does not hurt, because the right hand side will be a refinement of $\bar{a}$ or ∞ anyway, allowing us to apply the next theorem, which gives the relation between filter limits and the usual kind of limits:

Theorem 2.5: *Limit correspondence theorem:*

$$\begin{aligned}\lim_{x\to a} f(x) = b \quad&\Leftrightarrow\quad \mathop{\mathrm{Lim}}_{x\to a^{\pm}} f(x) \sqsubseteq \bar{b} \\ \lim_{x\to a^{+}} f(x) = b \quad&\Leftrightarrow\quad \mathop{\mathrm{Lim}}_{x\to a^{+}} f(x) \sqsubseteq \bar{b} \\ \lim_{x\to \infty} f(x) = b \quad&\Leftrightarrow\quad \mathop{\mathrm{Lim}}_{x\to \infty} f(x) \sqsubseteq \bar{b}\end{aligned}$$

Proof: We will show only the proof of the first equivalence. The other proofs are similar.

($\Rightarrow$) Let be given that $\lim_{x\to a} f(x) = b$: we have to show that $\bar{f}(a^{\pm}) \sqsubseteq \bar{b}$. So suppose that $V \in \bar{b}$, which means that for some $\epsilon > 0$ we have that $(b-\epsilon, b+\epsilon) \subseteq V$. Because of the limit, there is a $\delta > 0$ such that for all $x \in (a-\delta, a) \cup (a, a+\delta)$ (call this set U) we have that $f(x)$ is defined and $f(x) \in (b-\epsilon, b+\epsilon)$, or, in other words, $U \subseteq \mathrm{dom}(f)$ and $f[U] \subseteq (b-\epsilon, b+\epsilon)$. Clearly U satisfies $U \in a^{\pm}$ and $f[U] \subseteq V$, and therefore $V \in \bar{f}(a^{\pm})$.

($\Leftarrow$) Now assume that $\bar{f}(a^{\pm}) \sqsubseteq \bar{b}$. For a given $\epsilon > 0$, let $V = (b-\epsilon, b+\epsilon)$. Then $V \in \bar{b}$, so $V \in \bar{f}(a^{\pm})$, and therefore there is some $U \in a^{\pm}$ with $U \subseteq \mathrm{dom}(f)$ and $f[U] \subseteq V$. Because $U \in a^{\pm}$, for some $\delta > 0$ we have that $(a-\delta, a) \cup (a, a+\delta) \subseteq U$. Now this implies that if $x \in (a-\delta, a) \cup (a, a+\delta)$ then $f(x)$ is defined and $f(x) \in (b-\epsilon, b+\epsilon)$. Because for each $\epsilon > 0$ there is a $\delta > 0$ with this property, $\lim_{x\to a} f(x) = b$. □

Similar theorems hold at a^- and $-\infty$ and for limits to plus or minus infinity.

In Europe $\lim_{x\to a^+}$ is sometimes written as $\lim_{x\downarrow a}$. The ∞ and a^+ in the 'ordinary' limits on the left are not filters: those are just the customary notations for limits from the right and to infinity. The $a^\pm$, a^+ and ∞ on the right *are* filters.

Together these theorems now give us a method to rigorously evaluate ordinary limits using filters:

1. Replace the limit by the corresponding filter limit.
2. 'Evaluate' the filter limit using filter calculations, leading to a refinement.
3. If the right hand side of the refinement refines $\bar{a}$, $-\infty$ or ∞ then we have succeeded and can use Theorem 2.5 (or its analogue for infinite limits) to find the answer to the original question. If not, the method failed.

As an example, we use this method to evaluate $\lim_{x\to\infty} 1/(x+1)$:

$$\operatorname*{Lim}_{x\to\infty} \frac{1}{x+1} \sqsubseteq \frac{\bar{1}}{\operatorname{Lim}_{x\to\infty}(x+1)} \sqsubseteq \frac{\bar{1}}{\infty+\bar{1}} = \frac{\bar{1}}{\infty} = 0^+$$

(The refinements here are really equalities, but the theorems that we have do not give that, and in fact we do not need it.) Now $0^+ \sqsubseteq \bar{0}$ and so from Theorem 2.5 we find that:

$$\lim_{x\to\infty} \frac{1}{x+1} = 0$$

Here is another example of a limit evaluated using this method:

$$\operatorname*{Lim}_{x\to 0^\pm} x \sin\frac{1}{x} \sqsubseteq 0^\pm \sin\frac{1}{0^\pm} = 0^\pm \sin(\pm\infty) = 0^\pm[-1,1] = \bar{0}$$

(in this calculation '$[-1,1]$' is the filter generated by the closed interval $[-1,1]$, cf. the notation introduced in Section 3 below), which implies that:

$$\lim_{x\to 0} x \sin\frac{1}{x} = 0$$

Note that we rigorously established this limit *by calculation*, and did not need the 'squeeze theorem' which is usually used to evaluate this limit.

3. Interval filters and connected filters

Definition: We will define a class of filters on $\mathbb{R}$ called the *interval filters*. Consider the set:

$$\mathcal{R} = \{-\infty^+\} \cup \{a^- \mid a \in \mathbb{R}\} \cup \{a^+ \mid a \in \mathbb{R}\} \cup \{\infty^-\}$$

For each pair of elements α and β from $\mathcal{R}$ for which $\alpha \le \beta$ in the natural order on $\mathcal{R}$, we will define a filter $\mathsf{interval}(\alpha, \beta)$. We map the elements of $\mathcal{R}$ to formulas as:

$$\begin{array}{ccc}
\alpha & \phi_l(x,\alpha,\varepsilon) & \phi_r(x,\alpha,\varepsilon) \\
-\infty^+ & \top & x < -1/\varepsilon \\
a^- & a-\varepsilon < x & x < a \\
a^+ & a < x & x < a+\varepsilon \\
\infty^- & 1/\varepsilon < x & \top
\end{array}$$

and then we define:

$$\mathsf{interval}(\alpha,\beta) \equiv \{U \in \mathcal{O}(\mathbb{R}) \mid \exists\varepsilon \in \mathbb{R}_{>0}.\, \forall x \in \mathbb{R}.\, \phi_l(x,\alpha,\varepsilon) \wedge \phi_r(x,\beta,\varepsilon) \Rightarrow x \in U\}$$

We will write interval filters using interval notation:

$$\begin{array}{rclcrcl}
[a,b) & \equiv & \mathsf{interval}(a^-,b^-) & \quad & (a,b) & \equiv & \mathsf{interval}(a^+,b^-) \\
[a,b] & \equiv & \mathsf{interval}(a^-,b^+) & & (a,b] & \equiv & \mathsf{interval}(a^+,b^+)
\end{array}$$

We suppose that it will be clear from the context when we mean an interval as a set of real numbers and when we mean an interval as an interval filter. Generally, for finite a and b they are related like:

$$(a,b] = \mathsf{generated\text{-}by}(\{(a,b]\})$$

but not always. If $a = b$, then the left hand side is a^+ but the right hand side is $\mathsf{improper}$ because the set $(a,a]$ is empty.

When we analyze a two-sided limit into two one-sided limits, and then want to put the results back together, we need the concept of the 'join' of two filters, which we write $A \vee B$. For example, $0^- \vee 0^+ = 0^\pm$. This concept is defined as follows:

Definition: The operations *join* and *meet* on filters are defined by:

$$\begin{array}{rcl}
A \vee B & = & A \cap B \\
A \wedge B & = & \{U \cap V \mid U \in A \wedge V \in B\}
\end{array}$$

We can now write the filters defined on page 5 as interval filters or as joins of interval filters:

$$\begin{array}{rclcrcl}
\bar{a} & = & [a,a] & \quad & \leftrightarrow & = & (-\infty,\infty) \\
a^- & = & [a,a) & & \rightarrow & = & (0,\infty) \\
a^+ & = & (a,a] & & \leftarrow & = & (-\infty,0) \\
a^\pm & = & [a,a) \vee (a,a] & & \pm\!\rightarrow & = & (-\infty,0) \vee (0,\infty)
\end{array}$$

$$\begin{array}{rcl}
\infty & = & [\infty,\infty) \\
-\infty & = & (-\infty,-\infty] \\
\pm\infty & = & (-\infty,-\infty] \vee [\infty,\infty)
\end{array}$$

Now that we have the class of interval filters, we will define the class of connected filters. This definition is much simpler:

Definition: A filter A is called connected when:

$$\forall U \in A.\, \exists U' \in A.\, U' \subseteq U \wedge U' \textit{ is a connected set}$$

Note that each of improper, domain-error and indeterminate is a connected filter.

The next three theorems give the relevant properties of the connected filters. Together they 'explain' why in practice one encounters only joins of interval filters: the filters one starts with are of that kind, and the operations that one applies to them conserve the property.

THEOREM 3.1: *The interval filters are the proper non-empty connected filters.*

Proof: It is easy to verify that every interval filter is proper, non-empty and connected. Therefore all that is needed is to show that if a proper non-empty filter is connected, then it is an interval filter.

Assume A is a proper non-empty connected filter. Define:

$$\begin{array}{rcl} L & = & \{l \in \{-\infty\} \cup \mathbb{R} \mid \quad (l, \infty) \in A\} \\ R & = & \{r \in \mathbb{R} \cup \{\infty\} \quad \mid (-\infty, r) \in A\} \end{array}$$

Both sets are non-empty, because the filter A is non-empty and so $(-\infty, \infty) \in A$. The set L will be closed to the left (if $l \in L$ and $l' < l$ then also $l' \in L$), so L has to have the form $\{-\infty\}$, $[-\infty, a)$, $[-\infty, a]$ or $[-\infty, \infty)$. Depending on this, define α to be respectively $-\infty^+$, a^-, a^+ or ∞^-. Similarly R will be of the form $(-\infty, \infty]$, $[b, \infty]$, $(b, \infty]$ or $\{\infty\}$, which gives β being $-\infty^+$, b^-, b^+ or ∞^-. We have that $L \cap R = \emptyset$ (because else A would be improper), so $\alpha \leq \beta$. We are now going to show that $A = \mathsf{interval}(\alpha, \beta)$.

($\subseteq$) Suppose that $U \in A$. The filter A is proper and connected, so there is a non-empty connected $U' \in A$ with $U' \subseteq U$. A non-empty connected open set U' is always an open interval (l, r), and because $U' \in A$, we have that $l \in L$ and $r \in R$. With the definitions of ϕ_l and ϕ_r it then follows that $(l, r) \in \mathsf{interval}(\alpha, \beta)$. Therefore also $U \in \mathsf{interval}(\alpha, \beta)$.

($\supseteq$) Suppose that $U \in \mathsf{interval}(\alpha, \beta)$. Then from the definition of interval it is clear that there is an open interval (l, r) with $(l, r) \subseteq U$ and $(l, r) \in \mathsf{interval}(\alpha, \beta)$. With the definitions of ϕ_l and ϕ_r it then follows that $l \in L$ and $r \in R$. Therefore $(l, r) = (l, \infty) \cap (-\infty, r) \in A$, and so also $U \in A$. $\square$

So all interval filters are connected, and the only connected filters which are not an interval filter are the 'error filters' improper and domain-error.

THEOREM 3.2: *If f is a function that is continuous on its domain, and A is a connected filter, then $\bar{f}(A)$ is also connected.*

Proof: Suppose that $V \in \bar{f}(A)$, so there is a $U \in A$ with $U \subseteq \mathrm{dom}(f)$ and $f[U] \subseteq V$. We need to find a connected $V' \in \bar{f}(A)$ with $V' \subseteq V$.

Because A is connected, there is a connected $U' \in A$ with $U' \subseteq U$. Take $V' = f[U']$. The image of a connected set under a continuous function is connected, so V' is connected too. Furthermore, V' has the required properties: $U' \subseteq \mathrm{dom}(f)$ from which follows that $V' \in \bar{f}(A)$, and $V' = f[U'] \subseteq f[U] \subseteq V$. □

THEOREM 3.3:

$$\bar{f}(A \vee B) = \bar{f}(A) \vee \bar{f}(B)$$

Proof: ($\sqsubseteq$) Suppose $V \in \bar{f}(A) \vee \bar{f}(B)$, then both $V \in \bar{f}(A)$ and $V \in \bar{f}(B)$, so there are $U \in A$ and $U' \in B$ such that $U, U' \subseteq \mathrm{dom}(f)$ and $f[U], f[U'] \subseteq V$. Then $U \cup U' \in A \vee B$, and furthermore $U \cup U' \subseteq \mathrm{dom}(f)$ and $f[U \cup U'] \subseteq V$, so $V \in \bar{f}(A \vee B)$.

($\sqsupseteq$) If $V \in \bar{f}(A \vee B)$, there is a $U \in A \vee B$ (and so $U \in A$ and $U \in B$) with $U \subseteq \mathrm{dom}(f)$ and $f[U] \subseteq V$. This same U shows that $V \in \bar{f}(A)$ and $V \in \bar{f}(A)$, which means that $V \in \bar{f}(A) \vee \bar{f}(B)$. □

Together, these theorems show that if one applies functions that are continuous on their domain to finite joins of interval filters, one always will end up with finite joins of interval filters again.

The final theorem is not needed for actual limit calculations, but it is included for completeness.

THEOREM 3.4:

$$\bar{f}(A \wedge B) \sqsubseteq \bar{f}(A) \wedge \bar{f}(B)$$

Proof: Let $V \in \bar{f}(A) \wedge \bar{f}(B)$, so there are $V' \in \bar{f}(A)$ and $V' \in \bar{f}(B)$ with $V = V' \cap V'$. Then there are $U' \in A$ and $U' \in B$ with $U', U' \subseteq \mathrm{dom}(f)$ and $f[U'] \subseteq V'$, $f[U'] \subseteq V'$. From this $U' \cap U' \in A \wedge B$, $U' \cap U' \subseteq \mathrm{dom}(f)$ and $f[U' \cap U'] \subseteq V' \cap V'$, and so $V' \cap V' \in \bar{f}(A \wedge B)$. □

Note that for $\wedge$ the reverse refinement does not hold. If we take $f(x) = x^2$, then $\bar{f}(\leftarrow \wedge \rightarrow) = \bar{f}(\dagger) = \dagger$, but $\bar{f}(\leftarrow) \wedge \bar{f}(\rightarrow) = \rightarrow \wedge \rightarrow = \rightarrow$, and although $\dagger \sqsubseteq \rightarrow$, it is not the case that $\rightarrow \sqsubseteq \dagger$.

4. Mathematica revisited

Now that we have given a calculus of filters that resembles the way Mathematica calculates with infinity, we will compare the behavior of our calculus and that of Mathematica in detail. This is what the calculations in the example Mathematica from Section 1.1 become when we redo them in our filter calculus:

$$\begin{array}{rclrcl}
1/(\infty + 1) & = & 0^+ & 1/[-1,1] & = & \bot \\
\sqrt{\infty} & = & \infty & [-1,1] \cdot [-1,1] & = & [-1,1] \\
\infty - \infty & = & \leftrightarrow & [-1,1]^2 & = & [0,1] \\
\leftrightarrow + \infty & = & \leftrightarrow & \bar{0} \sin \infty & = & \bar{0} \\
\sin \infty & = & [-1,1] & \infty / \sin \infty & = & \infty/(\sin \infty)^2 = \bot
\end{array}$$

Here are some differences with Mathematica:

- Mathematica does not like to give 'no' for an answer. So it prefers not to complain about undefinedness of a function. According to Mathematica:

  $$\frac{1}{[-1,1]} = (-\infty,-1] \vee [1,\infty)$$

  instead of $\perp$. Our definitions have different behavior because we want the correspondence theorem about limits, Theorem 2.5, to hold. As an example of this difference in attitude consider the limit:

  $$\lim_{x\to 0^+} x \ \arctan(\tan \frac{1}{x})$$

  The graph of x arctan(tan($1/x$)) looks like a 'saw tooth' converging to 0, and it is undefined infinitely often in each neighborhood of 0. Still Mathematica says†:

  ```
  In[12]:= Limit[x*ArcTan[Tan[1/x]], x->0, Direction->-1]
  Out[12]= 0
  ```

  If you ask MathXpert to evaluate this limit, you get the message: *This function is undefined for certain values arbitrarily close to the limit point, so the limit is undefined.*

- Mathematica does not identify as many expressions as it might. For instance, in the example session it might have simplified:

  ```
              Interval[{0, 0}] = 0
    Interval[{Infinity, Infinity}] = Infinity
  Interval[{-Infinity, Infinity}] = Indeterminate
  ```

- Mathematica does not distinguish between open and closed intervals, nor does it have the concept of left and right filters to a point. In order to add this subtlety to its `Interval` calculus all that would be needed is to mark all the endpoints of the intervals with a + or a −.

- We have *two* kinds of 'undefined' in our filter calculus: domain-error $= \perp$ and indeterminate $= \leftrightarrow$. (The third filter, improper $= \dagger$, only occurs as the meet of two disjoint interval filters, and never occurs in practice.) Mathematica only has `Indeterminate`, and does not distinguish between these two kinds of undefinedness.

- Mathematica issues a 'warning' message like:

  Power::infy: Infinite expression $\frac{1}{0}$ encountered.

†In version 3.0. In version 4.1 it leaves the expression unevaluated.

when it gets infinite or indeterminate results. This seems to imply that such results are errors. However, in our theory those results are not errors at all but the correct answers, and they should not generate such a message.

- We gave the details of the filter theory for the space of real numbers. However, the expression language of Mathematica is about the complex numbers. This is clear, for example, from the results of applying `Sqrt` and `Log` to negative numbers. It is therefore strange that Mathematica gives answers involving intervals to limit questions, since such answers are appropriate to *real* limits.

In any case, our filter theory can be adapted to the complex numbers. For example, complex infinity is represented by the filter generated by the exteriors of disks centered at 0 (i.e., 'neighborhoods of infinity'). The 'one-sided' filters a^+ and a^- are replaced by a wide variety of other filters representing different ways in which a complex number z can 'approach' a limit point a: for example, in complex analysis it is common to consider a limit restricted to an angular sector, such as $|\theta| < \pi/4$. It is easy to define a 'sector filter' generated by such a sector. Our theorems that do not involve interval filters carry over to the complex setting: pushing filter limits inside functions, the method of limit evaluation by refinement, etc. We have not given a characterization of the connected filters in the complex case. For example, there are more than just the sector filters: the filter generated by $|\theta| < r^2$ is not refined by any sector filter.

5. Conclusion and future directions

We have presented the filter approach to evaluating limits involving infinity. The usual way of calculating with infinities is not rigorous; indeed the central concept *infinity* is never defined in calculus textbooks. The issue is skirted by such statements as: 'The symbol ∞ does not represent a real number and we cannot use it in arithmetic in the usual way.' (8), p. 112.

Consider a student who says that $\lim_{x\to 0^+} 1/x$ is 'undefined', while the teacher says that ∞ is a better answer. 'But', says the student, 'you said ∞ *is* undefined.' Such dialogues do occur regularly in classrooms and teachers are unable to answer these questions on any rigorous basis. We have now, at least in principle, provided a remedy for this situation, since our theory of infinite limits is completely rigorous. Questions at the student level in our theory can usually be proved or refuted.

When computer algebra systems make use of a set of calculation rules, there should ideally be a semantics according to which these calculation rules are correct. Even for ordinary algebra, this is not usually the case. But it is usually the case that the rules are correct *except* that the system fails to check the preconditions. That is, the semantics of algebra is understood – but systems fail to implement the rules in a semantically correct way. Up until now, the semantics of limits has not been properly understood, and so the behavior of computer algebra systems did not even have a standard against which implementations could be measured. In using intervals as answers to limits, Mathematica has ventured into uncharted territory. We are now providing maps.

Our work, being completely rigorous, and based on simple set theory, is also completely formal.[‡] Therefore computer-checking the theory from this paper is possible, and the resulting formalization would not only be an interesting exercise, but also probably could be used to make the prover automatically evaluate more limits. In another direction, this material is suitable for inclusion in an undergraduate real-analysis course, and the distinctions between different types of limits that it makes are suitable for inclusion in calculus books. In particular, calculus books need no longer steer away from calculations involving infinity. Simple rules for manipulating infinity can be given and the justifications omitted, as is usually the case now when the justifications involve epsilon-delta arguments.

References

[1] Beeson, M., *Mathpert Calculus Assistant*. This software program (now known as MathXpert) was published in July, 1997 by Mathpert Systems, Santa Clara, CA, and is commercially available from `<http://www.mathxpert.com/>`.

[2] Beeson, M., Design Principles of Mathpert: Software to support education in algebra and calculus, in: Kajler, N. (ed.) *Computer-Human Interaction in Symbolic Computation*, Texts and Monographs in Symbolic Computation, Springer-Verlag, Berlin/Heidelberg/New York (1998), pp. 89-115.

[3] Beeson, M. Beeson, M. MathXpert: un logiciel pour aider les élèves à apprendre les mathématiques par l'action, *Sciences et techniques éducatives*, **9** (1) 37-62, 2002. An English translation of this article under the title *MathXpert: learning mathematics in the twenty-first century* is available at `<http://www.mathcs.sjsu.edu/faculty/beeson/Pubs/pubs.html>`.

[4] Beeson, M., Using nonstandard analysis to verify the correctness of computations, *International Journal of Foundations of Computer Science*, **6** (3) (1995), pp. 299-338.

[5] K. Kuratowski. *Topology*, volume 1. Academic Press, New York, London, 1966.

[6] M. Monagan, K. Geddes, K. Heal, G. Labahn, and S. Vorkoetter. *Maple V Programming Guide for Release 5*. Springer-Verlag, Berlin/Heidelberg, 1997.

[‡] *Rigorous* implies that the concepts and theorems have a clear meaning and the theorems can be correctly proved. *Formal* implies that the concepts can be defined and the theorems proved in terms of set theory (or some other foundational theory of mathematics).

[7] B. Sims. *Fundamentals of Topology.* MacMillan, New York, 1976.

[8] Stewart. *Calculus, 3rd edition*, Brooks-Cole, Pacific Grove, CA 1995.

[9] S. Wolfram. *The Mathematica book.* Cambridge University Press, Cambridge, 1996.