

Equivalence of the Traditional and Non-Standard Definitions of Concepts from Real Analysis

John Cowles

University of Wyoming
Laramie, WY, USA
cowles@uwyo.edu

Ruben Gamboa

University of Wyoming
Laramie, WY, USA
ruben@uwyo.edu

ACL2(r) is a variant of ACL2 that supports the irrational real and complex numbers. Its logical foundation is based on internal set theory (IST), an axiomatic formalization of non-standard analysis (NSA). Familiar ideas from analysis, such as continuity, differentiability, and integrability, are defined quite differently in NSA—some would argue the NSA definitions are more intuitive. In previous work, we have adopted the NSA definitions in ACL2(r), and simply taken as granted that these are equivalent to the traditional analysis notions, e.g., to the familiar ϵ - δ definitions. However, we argue in this paper that there are circumstances when the more traditional definitions are advantageous in the setting of ACL2(r), precisely because the traditional notions are classical, so they are unencumbered by IST limitations on inference rules such as induction or the use of pseudo-lambda terms in functional instantiation. To address this concern, we describe a formal proof in ACL2(r) of the equivalence of the traditional and non-standards definitions of these notions.

Keywords: ACL2(r), non-standard analysis, real analysis.

1 Introduction

ACL2(r) is a variant of ACL2 that has support for reasoning about the irrational numbers. The logical basis for ACL2(r) is *non-standard analysis* (NSA), and in particular, the axiomatic treatment of NSA developed as *internal set theory* (IST) [5]. Traditional notions from analysis, such as limits, continuity, and derivatives, have counterparts in NSA

Previous formalizations of NSA typically prove that these definitions are equivalent early on. We resisted this in the development of ACL2(r), preferring simply to state that the NSA notions were the “official” notions in ACL2(r), and that the equivalence to the usual notions was a “well-known fact” outside the purview of ACL2(r). In this paper, we retract that statement for three reasons.

First, the traditional notions from real analysis require the use of quantifiers. For instance, we say that a function f has limit L as x approaches a iff

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } |x - a| < \delta \Rightarrow |f(x) - L| < \epsilon.$$

While ACL2(r) has only limited support for quantifiers, this support is, in fact, sufficient to carry out the equivalence proofs. However, it should be noted that the support depends on recent enhancements to ACL2 that allow the introduction of Skolem functions with non-classical bodies. So, in fact, it is ACL2’s improved but still modest support for quantifiers that is sufficient. That story is interesting in and of itself.

Second, the benefit of formalization in general applies to this case, as the following anecdote illustrates. While trying to update the proof of the Fundamental Theorem of Calculus, we were struggling to formalize the notion of *continuously differentiable*, i.e., that f is differentiable and f' is continuous. To

talk about the class of differentiable functions in ACL2(r), we use an `encapsulate` event to introduce an arbitrary differentiable functions. It would be very convenient to use the existing `encapsulate` for differentiable functions, and prove as a theorem that the derivative was continuous. That is to say, it would be very convenient if all derivative functions were continuous. Note: we mean “derivative” functions, not “differentiable” functions. The latter statement had previously been proved in ACL2(r).

Encouraged by Theorem 5.6 of [5], one of us set out to prove that, indeed, all derivatives of functions are continuous.

Theorem 1 (5.6, [5]). *Let $f : I \rightarrow \mathbb{R}$ where I is an interval. If f is differentiable on I , then f' is continuous on I .*

Nelson’s proof of this theorem begins with the following statement:

We know that

$$\forall^{\text{st}} x \forall x_1 \forall x_2 \left\{ x_1 \approx x \wedge x_2 \approx x \wedge x_1 \neq x_2 \Rightarrow \frac{f(x_2) - f(x_1)}{x_2 - x_1} \approx f'(x) \right\}. \quad (1)$$

This is, in fact, plausible from the definition of continuity, which is similar but with x taking the place of x_2 . The remainder of the proof was “trivially” (using the mathematician’s sense of the word) carried out in ACL2(r), so only the proof of this known fact remained. The hand proof for this fact was tortuous, but eminently plausible. Unfortunately, the last step in the proof failed, because it required that $y \cdot y_1 \approx y \cdot y_2$ whenever $y_1 \approx y_2$ —but this is true only when y is known to be limited.

The other of us was not fooled by Theorem 5.6: What about the function $x^2 \sin(1/x)$? The discrepancy was soon resolved. Nelson’s definition of derivative in [5] is precisely Equation 1. No wonder this was a known fact! And the problem is that Equation 1 is equivalent to the notion of continuously differentiable, and *not equivalent* to the usual notion of differentiability. But in that case, how are we to know if theorems in ACL2(r) correspond to the “usual” theorems in analysis. I.e., what if we had chosen Equation 1 as the definition of derivative in ACL2(r)? Preventing this situation from reoccurring is the second motivator for proving the equivalence of the definitions in ACL2(r) once and for all.

Third, the NSA definitions are non-classical; i.e., they use notions such as “infinitely close” and “standard.” Indeed, it is these non-classical properties that make NSA such a good fit for the equational reasoning of ACL2(r). However, non-classical functions are severely limited in ACL2(r): Induction can be used to prove theorems using non-classical functions only up to standard values of the free variables, and function symbols may not map to pseudo-lambda expressions in a functional instantiation [3]. As a practical consequence of these restrictions, it is impossible to prove that $\frac{d(x^n)}{dx} = n \cdot x^{n-1}$ by using the product rule and induction in ACL2(r). In [2], for example, this is shown only for standard values of n . However, using the traditional notion of differentiability, the result does follow from induction. This, too, would have been reason enough to undertake this work.

It should be emphasized that the main contribution of this paper is the formalization in ACL2(r) of the results described in this paper. The actual mathematical results are already well-known in the non-standard analysis community. Moreover, some of these equivalence results were formalized mechanically as early as [1]. The novelty here is the formalization in ACL2(r), which complicates things somewhat because of the poor support for (even first-order) set theory.

The rest of this paper is organized as follows. In Section 2, we discuss equivalent definitions regarding convergence of series¹. Section 3 considers the limit of a function at a point. The results in this

¹Readers who attended the ACL2 Workshop in 2013 will recognize many of the results in this section, because they were presented in a Rump Session there.

section are used in Section 4 to show that the notions of continuity at a point are also equivalent. This leads into the discussion of differentiability in Section 5. Finally, Section 6 deals with the equivalent definitions of Riemann integration.

2 Convergence of Series

In this section, we show that several definitions of convergence are in fact equivalent. In particular, we will consider the traditional definitions, e.g., as found in [8], and the corresponding concepts using non-standard analysis, e.g., as found in [7].

We start with the constrained function `Ser1`, which represents an arbitrary sequence; i.e., it is a fixed but arbitrary function that maps the natural numbers to the reals. Moreover, `Ser1` is assumed to be a classical function—otherwise, some of the equivalences do not hold. Similarly, the function `sumSer1-upto-n` defines the partial sum of `Ser1`, i.e., the sum of the values of `Ser1` from 0 to `n`.

The first definition of convergence is the traditional one due to Weierstrass:

$$(\exists L)(\forall \epsilon)(\exists M)(\forall n)(n > M \Rightarrow |\sum_{i=0}^n a_i - L| < \epsilon).$$

In ACL2, we can write the innermost quantified subformula of this definition as follows:

```
(defun-sk All-n-abs-sumSer1-upto-n-L<eps (L eps M)
  (forall n (implies (and (standardp n)
                           (integerp n)
                           (> n M))
    (< (abs (- (sumSer1-upto-n n) L))
        eps))))
```

This version of the definition restricts `n` to be a standard integer, which makes it a non-classical formula. A different version omits this requirement, and it is a more direct translation of Weierstrass's criterion.

```
(defun-sk Classical-All-n-abs-sumSer1-upto-n-L<eps (L eps M)
  (forall n (implies (and (integerp n)
                           (> n M))
    (< (abs (- (sumSer1-upto-n n) L))
        eps))))
```

ACL2 can verify that these two conditions are equal to each other, but only when the parameters `L`, `eps`, and `M` are standard. This follows because `defchoose` is guaranteed to choose a standard witness for classical formulas and standard parameters. More precisely, the witness function is a classical formula is also classical, all all classical functions return standard values for standard inputs [3]. Once this basic equivalence is proved, it follows that both the classical and non-classical versions of Weierstrass's criterion are equivalent. It is only necessary to add each of the remaining quantifiers one by one.

We note in passing that the two versions of Weierstrass's criterion are *not* equivalent for non-standard values of the parameters `L`, `eps`, and `M`. Consider, for example, the case when `eps` is infinitesimally small. It is straightforward to define the sequence $\{a_n\}$ such that the partial sums are given by $\sum_{i=1}^n a_i = 1/n$, clearly converging to 0. Indeed, for any $\epsilon > 0$ there is an N such that for all $m > N$, $\sum_{i=1}^m a_i = 1/m < 1/N < \epsilon$. However, for infinitesimally small ϵ , the resulting N is infinitesimally large. This is fine using the second (classical) version of Weierstrass's criterion, but not according to the first, since for all standard N , $1/N > \epsilon$, so no standard N can satisfy the criterion. However, the two criteria are equivalent when written as sentences, i.e., when they have no free variables.

Note that the only difference between the two versions of Weierstrass's criterion is that one of them features only standard variables, whereas the other features arbitrary values for all quantified variables. Using the shorthand $\forall^{\text{st}}$ and $\exists^{\text{st}}$ to introduce quantifiers for standard variables, the two versions of the criteria can be written as follows:

- $(\exists^{\text{st}}L)(\forall^{\text{st}}\epsilon)(\exists^{\text{st}}M)(\forall^{\text{st}}n)(n > M \Rightarrow |\sum_{i=0}^n a_i - L| < \epsilon)$
- $(\exists L)(\forall \epsilon)(\exists M)(\forall n)(n > M \Rightarrow |\sum_{i=0}^n a_i - L| < \epsilon)$

It is obvious that these two statements are extreme variants, and that there are other possibilities mixing the two types of quantifiers. Indeed, we verified with ACL2 that the following versions are also equivalent to the above:

- $(\exists^{\text{st}}L)(\forall^{\text{st}}\epsilon)(\exists^{\text{st}}M)(\forall n)(n > M \Rightarrow |\sum_{i=0}^n a_i - L| < \epsilon)$
- $(\exists^{\text{st}}L)(\forall^{\text{st}}\epsilon)(\exists M)(\forall n)(n > M \Rightarrow |\sum_{i=0}^n a_i - L| < \epsilon)$

The last two versions of Weierstrass's criterion are useful, because they are easier to show equivalent to the typical non-standard criterion for convergence: $(\exists L)(\forall n)(\text{large}(n) \Rightarrow \sum_{i=0}^n a_i \approx L)$, i.e., for large values of n , $\sum_{i=0}^n a_i$ is infinitely close to L . This is the convergence criterion used in [2], for example, where power series are used to introduce functions such as e^x .

There is another statement of the non-standard convergence criterion that appears weaker:

$$(\exists L)(\exists M)(\text{large}(M) \wedge (\forall n)(n > M \Rightarrow \sum_{i=0}^n a_i \approx L)).$$

This version does not require that $\sum_{i=0}^n a_i$ is close to L for all large n , only that this is true for n larger than some large M . We have shown in ACL2 that these statements are in fact equivalent to Weierstrass's criterion for convergence. In fact, since $\{a_n\}$ is a classical sequence, the value of L is guaranteed to be standard, so we can replace $(\exists L)$ with $(\exists^{\text{st}}L)$ in both of the non-classical convergence criteria given above and still retain equivalence.

When the sequence is composed of non-negative numbers, we can make even stronger guarantees. Let $\{b_n\}$ be such a sequence, which we introduce into ACL2 as the constrained function `Ser1a`. All the previous results about `Ser1`—i.e., about $\{a_n\}$ —apply to `Ser1a`, and we can carry over these proofs in ACL2 by using functional instantiation.

Using the non-standard criterion for convergence, we can easily see that if $\sum_{i=0}^{\infty} b_n$ converges, then $\sum_{i=0}^N b_i$ is not infinitely large, where N is a fixed but arbitrary large integer². This simply follows from the facts that $\sum_{i=0}^N b_i \approx L$ and L is standard.

The converse of this fact is also true: if $\sum_{i=0}^N b_i$ is not infinitely large, then $\sum_{i=0}^{\infty} b_n$ converges. This is harder to prove formally. The key idea is as follows. Since $\sum_{i=0}^N b_i$ is not infinitely large, then $\sum_{i=0}^N b_i$ must be close to a unique standard real number, i.e., $\sum_{i=0}^N b_i \approx L$ for some standard L . $\sum b_i$ is monotonic, so for any standard n , $\sum_{i=0}^n b_i \leq \sum_{i=0}^N b_i$. And since L is the unique real number that is close to $\sum_{i=0}^N b_i$, we can conclude that $\sum_{i=0}^n b_i \leq L$ for all standard n . Using the non-standard transfer principle, this is sufficient to conclude that $\sum_{i=0}^n b_i \leq L$ for all n , not just the standard ones. Using monotonicity once more, it follows that whenever $n > N$, $\sum_{i=0}^n b_i \approx L$, which is precisely the (weak) non-standard convergence criterion above. Thus, the series $\sum_{i=0}^{\infty} b_i$ converges, according to any of the criteria above.

Similar results hold for divergence to positive infinity. Let $\{c_n\}$ be an arbitrary sequence. Weierstrass's criterion is given by $(\forall^{\text{st}}B)(\exists^{\text{st}}M)(\forall^{\text{st}}n)(n > M \Rightarrow \sum_{i=0}^n c_i > B)$. As before, for classical $\{c_n\}$ this

²The ACL2 constant `(i-large-integer)` is often used to denote an otherwise unspecified large integer, and that is what we use in this case.

is equivalent to a criterion with quantifiers over all reals, not just the standard ones: $(\forall B)(\exists M)(\forall n)(n > M \Rightarrow \sum_{i=0}^n c_i > B)$. And just as before, other variants (with B and M standard or just B standard) are also equivalent. Moreover, these are equivalent to the non-standard criterion for divergence to positive infinity, namely that $(\forall n)(large(n) \Rightarrow large(\sum_{i=0}^n c_i))$. A seemingly weaker version of this criterion is also equivalent, where it is only necessary that c_n is large for all n beyond a given large integer: $(\exists M)(large(M) \wedge (\forall n)(n > M \Rightarrow large(\sum_{i=0}^n c_i)))$. Finally, if the sequence $\{c_n\}$ consists of non-negative reals, then it is even easier to show divergence. It is only necessary to test whether $large(\sum_{i=0}^N c_i)$ where N is an arbitrary large integer, and as before we choose the ACL2 constant `i-large-integer` for this purpose.

3 Limits of Functions

In this section, we consider the notion of limits. In particular, we show that the following three notions are equivalent (for standard functions and parameters):

- The non-standard definition (for standard parameters a and L):

$$\lim_{x \rightarrow a} f(x) = L \Leftrightarrow ((\forall x)(x \approx a \wedge x \neq a \Rightarrow f(x) \approx L)).$$

- The traditional definition over the classical reals:

$$\lim_{x \rightarrow a} f(x) = L \Leftrightarrow ((\forall^{\text{st}} \epsilon > 0)(\exists^{\text{st}} \delta > 0)(0 < |x - a| < \delta \Rightarrow |f(x) - L| < \epsilon)).$$

- The traditional definition over the hyperreals:

$$\lim_{x \rightarrow a} f(x) = L \Leftrightarrow ((\forall \epsilon > 0)(\exists \delta > 0)(0 < |x - a| < \delta \Rightarrow |f(x) - L| < \epsilon)).$$

We begin by assuming the non-standard definition, which can be introduced in ACL2(r) by encapsulating the function f , its domain, and the limit function L , so that $\lim_{x \rightarrow a} f(x) = L(a)$. The first step is to observe that $a \approx b$ is a shorthand notation for the condition that $|a - b|$ is infinitesimally small. Moreover, if $\epsilon > 0$ is standard, then it must be (by definition) larger than any infinitesimally small number. Thus, we can prove that

$$(\forall^{\text{st}} \epsilon > 0)((\forall x)(x \approx a \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

Similarly, if $\delta > 0$ is infinitesimally small, then $|x - a| < \delta$ implies that $x \approx a$. It follows then that

$$(\forall^{\text{st}} \epsilon > 0)(\forall \delta > 0)(small(\delta) \Rightarrow (\forall x)(0 < |x - a| < \delta \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

It is an axiom of ACL2(r) that there exists a positive infinitesimal, namely $(/ \text{ (i-large-integer)})$. Consequently, we can specialize the previous theorem with the constant δ_0 (i.e., $(/ \text{ (i-large-integer)})$).

$$(\forall^{\text{st}} \epsilon > 0)(0 < \delta_0 \wedge small(\delta_0) \wedge (\forall x)(0 < |x - a| < \delta_0 \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

Using ACL2 terminology, the specific number δ_0 can be generalized to yield the following theorem:

$$(\forall^{\text{st}} \epsilon > 0)(\exists \delta > 0)((\forall x)(0 < |x - a| < \delta \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

Note that the statement inside the $\forall^{\text{st}}$ is classical; i.e., it does not use any of the notions from NSA, such as standard, infinitesimally close, infinitesimally small, etc. Consequently, we can use the transfer principle so that the quantifier ranges over all reals instead of just the standard reals. This results in the traditional definition of limits over the hyperreals:

$$(\forall \epsilon > 0)(\exists \delta > 0)((\forall x)(0 < |x - a| < \delta \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

The transfer can also be used in the other direction. The introduction of the existential quantifier is done via `defun-sk`, and `ACL2(r)` introduces such quantifiers by creating a Skolem choice function $\delta(a, \epsilon)$ using `defchoose`. Since the criteria used to define this Skolem function are classical, `defchoose` introduces the Skolem function itself as classical. That means that when a and ϵ are standard, so is $\delta(a, \epsilon)$. This observation is sufficient to show that $\lim_{x \rightarrow a} f(x) = L(a)$, using the traditional definition over the classical reals:

$$(\forall^{\text{st}} \epsilon > 0)(\exists^{\text{st}} \delta > 0)((\forall x)(0 < |x - a| < \delta \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

It is worth noting that this last theorem is not obviously weaker or stronger than the previous one, where the quantifiers range over all reals, not just the standard ones. The reason is that the $\forall$ quantifier ranges over more values than $\forall^{\text{st}}$, so it would appear that using $\forall$ instead of $\forall^{\text{st}}$ yields a stronger result. However, this advantage is lost when one considers the $\exists$ quantifier, since $\exists^{\text{st}}$ gives an apparently stronger guarantee. In actual fact, the two statements are equivalent, since the transfer principle can be used to guarantee that the value guaranteed by $\exists$ can be safely assumed to be standard.

To complete the proof, we need to show that if $\lim_{x \rightarrow a} f(x) = L(a)$, using the traditional definition over the standard reals, then $\lim_{x \rightarrow a} f(x) = L(a)$ using the non-standard definition. To do this, we introduce a new `encapsulate` where f is constrained to have a limit using the traditional definition over the standard reals. We then proceed as follows. First, fix ϵ so that it is positive and standard. From the (standard real) definition of limit, it follows that

$$(\exists^{\text{st}} \delta > 0)(\forall x)(0 < |x - a| < \delta \Rightarrow |f(x) - L(a)| < \epsilon).$$

Now suppose that δ_0 is a positive, infinitesimally small number. It follows that $\delta_0 < \delta$ for any positive, standard δ . In particular, this means that

$$0 < \delta_0 \wedge (\forall x)(0 < |x - a| < \delta_0 \Rightarrow |f(x) - L(a)| < \epsilon).$$

Since δ_0 is an arbitrary positive infinitesimal, we can generalize it as follows:

$$(\forall \delta > 0)(\text{small}(\delta) \Rightarrow (\forall x)(0 < |x - a| < \delta \Rightarrow |f(x) - L(a)| < \epsilon)).$$

Next, we remove the universal quantifier on x . This step does not have a dramatic impact on the mathematical statement, but it is more dramatic in `ACL2(r)`, since it opens up a function introduced with `defun-sk`:

$$(\forall \delta > 0)(\text{small}(\delta) \Rightarrow (0 < |x - a| < \delta \Rightarrow |f(x) - L(a)| < \epsilon)).$$

Recall that $x \approx a$ is a shorthand for $|x - a|$ is infinitesimally small. Thus, the theorem implies the following

$$(\forall \delta > 0)(\text{small}(\delta) \Rightarrow (x \approx a \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon)).$$

At this point, the variable δ is unnecessary, so we are left with the following:

$$x \approx a \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon.$$

Now, recall that we fixed ϵ to be an arbitrary, positive, standard real. This means that what we have shown is actually the following:

$$(\forall^{\text{st}} \epsilon)(x \approx a \wedge x \neq a \Rightarrow |f(x) - L(a)| < \epsilon).$$

To complete the proof, it is only necessary to observe that if $|x - y| < \epsilon$ for all standard ϵ , then $x \approx y$. We prove this in ACL2(r) by finding an explicit standard ϵ_0 such that if $x \not\approx y$, then $|x - y| > \epsilon_0$. The details of that proof are tedious and not very elucidating, so we omit them from this discussion³. Once that lemma is proved, however, it follows that $\lim_{x \rightarrow a} f(x) = L(a)$ using the non-standard definition:

$$x \approx a \wedge x \neq a \Rightarrow f(x) \approx L(a).$$

These results show that the three definitions of limit are indeed equivalent, at least when f and L are classical, and a is standard.

4 Continuity of Functions

Now we consider the notion of continuity. The function f is said to be continuous at a if $\lim_{x \rightarrow a} f(x) = f(a)$. Since this uses the notion of limit, it is no surprise that there are three different characterizations which are equivalent (for standard functions and parameters):

- The non-standard definition (for standard parameter a):

$$f \text{ is continuous at } a \Leftrightarrow ((\forall x)x \approx a \wedge x \neq a \Rightarrow f(x) \approx f(a)).$$

- The traditional definition over the classical reals:

$$f \text{ is continuous at } a \Leftrightarrow ((\forall^{\text{st}} \epsilon > 0)(\exists^{\text{st}} \delta > 0)(0 < |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon)).$$

- The traditional definition over the hyperreals:

$$f \text{ is continuous at } a \Leftrightarrow ((\forall \epsilon > 0)(\exists \delta > 0)(0 < |x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon)).$$

What this means is that the notion of continuity can be completely reduced to the notion of limits. In particular, the results from Section 3 can be functionally instantiated to derive the results for continuity. It is only necessary to instantiate both functions $f(x)$ and $L(x)$ to the same function $f(x)$.

³The interested reader can consult the definition of `standard-lower-bound-of-diff` which produces the constant ϵ_0 mentioned above, and the lemmas `standards-are-in-order-2`, `standards-are-in-order`, `rlfn-classic-has-limits-step-3`, and the more trivial lemmas leading up to the main theorem `rlfn-classical-has-a-limit-using-nonstandard-criterion`.

5 Differentiability of Functions

Next, we consider differentiability. At first sight, it appears that we can also define differentiability in terms of limits. After all, f' is the derivative of f iff

$$\lim_{\epsilon \rightarrow 0} \frac{f(x+\epsilon) - f(x)}{\epsilon} = f'(x).$$

The problem, however, is that the difference quotient on the left of the equation is a function of both x and ϵ , and having free variables complicates functional instantiation when non-classical functions are under consideration. So we chose to prove this result essentially from scratch, although the pattern is very similar to the equivalence of limits.

Before proceeding, however, it is worth noting one other equivalence of interest. The non-standard definition of differentiability is as follows:

$$\text{standard}(a) \wedge x_1 \approx a \wedge x_1 \neq a \wedge x_2 \approx a \wedge x_2 \neq a \Rightarrow \\ \left(\text{-large} \left(\frac{f(x_1) - f(a)}{x_1 - a} \right) \wedge \frac{f(x_1) - f(a)}{x_1 - a} \approx \frac{f(x_2) - f(a)}{x_2 - a} \right).$$

The form of this definition was chosen because it does not have a dependency on f' , so it can be applied to functions even when their derivative is unknown. However, when f' is known, a simpler definition can be used:

$$\text{standard}(a) \wedge x \approx a \wedge x \neq a \Rightarrow \left(\frac{f(x) - f(a)}{x - a} \approx f'(a) \right).$$

In fact, this latter form is the definition of differentiability that was used in [6]. In that context, ACL2(r) was able to automatically define f' from the definition of f , so f' was always known and the simpler definition was appropriate.

So the first result we show is to relate the definitions of differentiable and derivative. To do so, we can begin with a differentiable function f and define f' (for standard a) as follows:

$$f'(a) \equiv \text{standard part} \left(\frac{f(a + \epsilon) - f(a)}{\epsilon} \right)$$

where ϵ is a fixed but arbitrary, positive, small real, e.g., $(/ \text{ (i-large-integer)})$. By assumption, the difference quotient at a is not large for $x_1 = a + \epsilon$. Since $f'(a)$ is defined as the standard part of the difference quotient, it follows that it really is close to the difference quotient, so f' really is the derivative of f .

Conversely, suppose f' is the derivative of f . Since f' is classical and a is standard, it follows that $f'(a)$ is standard, and in particular it is not large. Therefore, for any x_1 such that $x_1 \approx a$ and $x_1 \neq a$, the difference quotient at x_1 must be close to $f'(a)$ (by definition of derivative). It follows then that the difference quotient at x_1 is not large, since it's close to something that is not large. Moreover, since $\approx$ is transitive, if x_2 is also such that $x_2 \approx a$ and $x_2 \neq a$, then the difference quotients at x_1 and x_2 are both close to $f'(a)$, so they must also be close to each other. Thus, f is differentiable according to the non-standard criterion. This simple argument is sufficient to combine the results of differentiability in ACL2(r) with the automatic differentiator described in [6], making the automatic differentiator much more useful, since the notion of differentiability it uses is now consistent with the main definition in ACL2(r).

Next, we show that the non-standard definition of derivative is equivalent to the traditional definition (both for the hyperreals and for the standard reals). The proof is nearly identical to the corresponding proof about limits, so we omit it here.

Discussion

There is a possible misconception that needs to be corrected. We have shown that the three different notions of differentiability are equivalent in principle. However, this is far from sufficient in practice.

To understand the problem, consider a function such as x^n , which may be represented in ACL2(r) as `(expt x n)`. In a real application of analysis, we may want to show that $f(x) = x - x^{2^n}$ achieves its maximum value at $x = 1/\sqrt[2^{n-1}]{2n}$. ACL2(r) has the basic lemmas that are needed to do this:

- $\frac{d(x^n)}{dx} = n \cdot x^{n-1}$ (at least for standard n)
- Chain rule
- Extreme value theorem (EVT)
- Mean value theorem (MVT)

But these lemmas cannot be used directly. Consider the chain rule, for example. Its conclusion is about the differentiability of $f \circ g$, and the notion of differentiability is the non-standard definition. What this means is that the functions f and g cannot be instantiated with pseudo-lambda expressions, so f and g must be unary, and that rules out x^n which is formally a binary function, even if we think of it as unary because n is fixed.

Moreover, suppose that we have a stronger theorem, namely that

$$\frac{d(x^n)}{dx} = n \cdot x^{n-1}$$

for all n , not just the standard ones. It's possible to prove this using induction and the hyperreal definition of differentiability (since it's a purely classical definition, so induction can be used over all the naturals, not just the standard ones). Suppose we want to invoke the MVT on x^n over some interval $[a, b]$. It is not possible to use the equivalence of the hyperreal and non-standard definitions. The reason, again, is that the non-standard definition is non-classical, so we cannot use pseudo-lambdas in functional instantiations. Even though the two definitions of differentiability are equivalent for arbitrary (unary) $f(x)$, they are not equivalent for the function x^n (which is binary).

It may seem that this is an unnecessary limitation in the part of ACL2(r). But actually, it's just part of the definition. The non-standard definition says that the difference quotient of f is close to f' at standard points x . It says nothing about non-standard points. But when a binary function is considered, e.g., x^n , what should happen when x is standard but n is not? In general, the difference quotient need *not* be close to the derivative.

This fact can be seen quite vividly by fixing $x = 2$ and N an arbitrary (for now), large natural number. Is the derivative with respect to x of x^n close to the difference quotient when $x = 2$ and $n = N$? The answer can be no, as the following derivation shows:

$$\begin{aligned} \frac{(2+\epsilon)^N - 2^N}{\epsilon} &= \frac{(2^N + N\epsilon 2^{N-1} + \binom{N}{2}\epsilon^2 2^{N-2} + \dots + \epsilon^N) - 2^N}{\epsilon} \\ &= \frac{N\epsilon 2^{N-1} + \binom{N}{2}\epsilon^2 2^{N-2} + \dots + \epsilon^N}{\epsilon} \\ &= \frac{\epsilon(N 2^{N-1} + \binom{N}{2}\epsilon 2^{N-2} + \dots + \epsilon^{N-1})}{\epsilon} \\ &= N 2^{N-1} + \binom{N}{2}\epsilon 2^{N-2} + \dots + \epsilon^{N-1} \end{aligned}$$

All terms except the first have a factor of ϵ , so if N were limited, those terms would be infinitesimally small, and thus the derivative would be close to the difference quotient. But if N is large, $\binom{N}{2} = \frac{N(N-1)}{2}$ is also large. And if $N = \lceil 1/\epsilon \rceil$, then $\binom{N}{2}\epsilon$ is roughly $N/2$, which is large. So the difference between the difference quotient and the derivative is arbitrarily large!

This shows that it is not reasonable to expect that we can convert from the traditional to the non-standard definition of derivative in all cases. Therefore, we cannot use previously proved results, such as the MVT directly.

A little subterfuge resolves the practical problem. What must be done is to prove a new version of the MVT (and other useful theorems about differentiability) for functions that are differentiable according to the ϵ - δ criterion for reals or hyperreals, as desired. Of course, the proofs follow directly from the earlier proofs. For instance, suppose that $f(x)$ is differentiable according to the hyperreal criterion. Then, we can use the equivalence theorems to show that $f(x)$ is differentiable according to the non-standard criterion. In turn, this means that we can prove the MVT for $f(x)$ using functional instantiation. Now, the MVT is a classical statement, so we instantiate it functionally with pseudo-lambda expressions. E.g., we can now use the MVT on $f(x) \rightarrow (\lambda(x)x^n)$. So even though we cannot say that x^n satisfies the non-standard criterion for differentiability, we can still use the practical results of differentiability, but only after proving analogues of these theorems (e.g., IVT, MVT, etc.) for the classical versions of differentiability. The proof of these theorems is a straightforward functional instantiation of the original theorems. We have done this for the key lemmas about differentiation (e.g., MVT, EVT, Rolle's Theorem, derivative composition rules, chain rule, derivative of inverse functions). We have also done this for some of the other equivalences, e.g., the Intermediate Value Theorem for continuous functions.

6 Integrability of Functions

The theory of integration in ACL2(r) was first developed in [4], which describes a proof of a version of the Fundamental Theorem of Calculus (FTC). The version of the FTC presented there is sometimes called the First Fundamental Theorem of Calculus, and it states that if f is integrable, then a function g can be defined as $g(x) = \int_0^x f(t)dt$, and that $g'(x) = f(x)$. As part of this proof effort, we redid the proof in [4], and generalized the result to what is sometimes called the Second Fundamental Theorem of Calculus. This more familiar form says that if $f'(x)$ is continuous on $[a, b]$, then $\int_a^b f'(x)dx = f(b) - f(a)$.

The integral formalized in [4] is the Riemann integral, and the non-standard version of integrability is as follows:

$$\int_a^b f(x)dx = L \Leftrightarrow (\forall P)(P \text{ is a partition of } [a, b] \wedge \text{small}(\|P\|) \Rightarrow \sum_{x_i \in P} (f(x_i)(x_i - x_{i-1})) \approx L)$$

P is a monotonically increasing partition of $[a, b]$ if P is given by a list $P = [x_1, x_2, \dots, x_n]$ with $x_1 = a$ and $x_n = b$. The term $\|P\|$ denotes the maximum value of $x_i - x_{i-1}$ in the partition P .

The traditional definition uses limits instead of the notion of infinitesimally close. It can be written as follows:

$$\int_a^b f(x)dx = L \Leftrightarrow \lim_{\|P\| \rightarrow 0} (\sum_{x_i \in P} (f(x_i)(x_i - x_{i-1}))) \approx L.$$

The notion of limit is strange here, because what approaches 0 is $\|P\|$. Many partitions can have the same value of $\|P\|$, so this limit ranges over all such partitions at the same time.

Opening up the definition of limits, integrals can be expressed as follows:

$$\int_a^b f(x)dx = L \Leftrightarrow (\forall \epsilon > 0)(\exists \delta > 0)(\forall P) \\ (P \text{ is a partition of } [a, b] \wedge \|P\| < \delta \Rightarrow |\sum_{x_i \in P} (f(x_i)(x_i - x_{i-1})) - L| < \epsilon).$$

Once integrals are viewed in this way, the remainder of the proof is clear. Specifically, it follows the same line of reasoning as in Section 3. First, the δ that exists depends on a , b , and ϵ , so it is standard when those are standard. Second, since there is a standard δ that is sufficient, any infinitesimal can take the place of δ , and then the condition $\|P\| < \delta$ can be recast as *small*($\|P\|$). Finally, since the Riemann sum is within ϵ of L , for an arbitrary, positive, standard ϵ , it must be that the Riemann sum is infinitesimally close to L . So the two definitions are, in fact, equivalent.

7 Conclusions

In this paper, we showed how the non-standard definitions of traditional concepts from analysis are in fact equivalent to the traditional ϵ - δ definitions. The results are especially important in ACL2(r) because the non-standard definitions feature non-classical notions, such as “infinitely close” and “infinitely small.” Consequently, they are limited in the use of induction and functional instantiation. However, the traditional notions are (by definition) classical, so they are unencumbered by such limitations.

This presents an interesting dilemma. In our experience, analysis style proofs are much easier to do and automate using non-standard analysis. However, *using* those results in subsequent proof attempts is much easier to do with the traditional (i.e., classical) statements. The distinction we’re making is between *proving* the correctness of Taylor’s Theorem, say, and actually *using* Taylor’s Theorem in a larger verification effort. For example, the formalization of Taylor’s Theorem in [9] took extreme care to push free variables (including what were really summation indexes for the series) all the way into the original `encapsulate` introducing the function to be approximated. However, now that the equivalences are proved, a more elegant approach can be followed: First, prove a “clean” version of Taylor’s Theorem using NSA, then use that result to show that Taylor’s Theorem also holds using the traditional definition of derivative. The “traditional” version of Taylor’s Theorem would then be used with no restrictions during functional instantiation, so free variables would no longer present a problem. We plan to pursue this idea for Taylor’s Theorem in the near future, as part of a comprehensive verification effort into the implementation of hardware algorithms for square root and various trigonometric and exponential functions.

References

- [1] A.M. Ballantyne & W. W. Bledsoe (1977): *Automatic Proofs of Theorems in Analysis Using Non-Standard Techniques*. *Journal of the Association for Computing Machinery (JACM)* 24(3), pp. 353–371, doi:10.1145/322017.322018.
- [2] R. Gamboa (1999): *Mechanically Verifying Real-Valued Algorithms in ACL2*. Ph.D. thesis, The University of Texas at Austin.
- [3] R. Gamboa & J. Cowles (2007): *Theory Extension in ACL2(r)*. *Journal of Automated Reasoning*, doi:10.1007/s10817-006-9043-0.

- [4] M. Kaufmann (2000): *Modular Proof: The Fundamental Theorem of Calculus*. In M. Kaufmann, P. Manolios & J S. Moore, editors: *Computer-Aided Reasoning: ACL2 Case Studies*, chapter 6, Kluwer Academic Press, doi:10.1007/978-1-4615-4449-4.
- [5] E. Nelson (1977): *Internal Set Theory: A New Approach to Nonstandard Analysis*. *Bulletin of the American Mathematical Society* 83, pp. 1165–1198, doi:10.1090/S0002-9904-1977-14398-X.
- [6] P. Reid & R. Gamboa (2011): *Automatic Differentiation in ACL2*. In: *Proc of the Second Conference on Interactive Theorem Proving (ITP-2011)*, doi:10.1007/978-3-642-22863-6_23.
- [7] A. Robinson (1996): *Non-Standard Analysis*. Princeton University Press.
- [8] W. Rudin (1976): *Principles of Mathematical Analysis*, third edition. McGraw-Hill.
- [9] J. Sawada & R. Gamboa (2002): *Mechanical Verification of a Square Root Algorithm using Taylor's Theorem*. In: *Formal Methods in Computer-Aided Design (FMCAD'02)*, doi:10.1007/3-540-36126-X_17.