Automatic Differentiation
as Nonarchimedean Analysis

M. Berz

Published in
Proceedings International Symposium on
Computer Arithmetic and Scientific Computation
SCAN 91
Elsevier Publishers

Automatic differentiation as nonarchimedean analysis

Martin Berz

Department of Physics and Astronomy and National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, MI 48824, USA

Abstract

It is shown how the techniques of automatic differentiation can be viewed in a broader context as an application of analysis on a nonarchimedean field. The rings used in automatic differentiation can be ordered in a natural way and form finite dimensional real algebras which contain infinitesimals. Some of these algebras can be extended to become a Cauchy-complete real-closed nonarchimedean field, which forms an infinite dimensional real vector space and is denoted by $\mathcal{L}$.

On this field, a calculus is developed. Rules of differentiation and certain fundamental theorems are discussed. A remarkable property of differentiation is that difference quotients with infinitely small differences yield the exact derivative up to an infinitely small error. This is of historical interest since it justifies the concept of derivatives as differential quotients. But it is also of practical relevance; it turns out that the algebraic operations used to compute derivatives in automatic differentiation are just special cases of calculus concepts on $\mathcal{L}$. The arithmetic on $\mathcal{L}$ can be implemented in programming languages, in particular if object oriented features exist, and should provide a useful data type for various applications.

1. INTRODUCTION

The goal of automatic differentiation [1, 2, 3, 4] is the accurate and rapid computation of derivatives of complicated functions in a computer environment. In the forward mode of automatic differentiation, this is achieved by substituting all real arithmetic by arithmetic on certain real algebras. To allow for the differentiation of expressions containing intrinsic functions, it is important to introduce these functions on these algebras as well.

In the simplest case in which only the first derivative with respect to one variable is required, the real arithmetic is replaced by arithmetic on ordered pairs which can be traced back all the way to Veronese [5]. The arithmetic is defined by

$$(a_0, a_1) + (b_0, b_1) := (a_0 + b_0, a_1 + b_1) \tag{1}$$

$$t \cdot (a_0, a_1) := (t \cdot a_0, t \cdot a_1) \tag{2}$$

$$(a_0, a_1) \cdot (b_0, b_1) := (a_0 \cdot b_0, a_0 \cdot b_1 + a_1 \cdot b_0) \tag{3}$$

With this arithmetic, the structure forms a **real algebra**, which we denote by ${}_1D_1$. On ${}_1D_1$ we can introduce an operation $\partial(a_0, a_1) = (0, a_1)$; this operation satisfies $\partial(a \cdot b) = a \cdot \partial b + b \cdot \partial a$, i.e. it is a derivation. With this derivation, the structure ${}_1D_1$ becomes a **differential algebra** in the sense of [6].

The algebra is not a field since $(0, 1)$ has no inverse. This is not surprising because according to the famous theorem of Frobenius, there are only two finite dimensional real vector spaces that are fields, the complex numbers and the quaternions.

We note that $(0, 1)$ is nilpotent: $(0, 1)^2 = (0, 0)$. It is also worth observing that the structure ${}_1D_1$ is a rather unique algebra. It is shown in [5] that up to isomorphisms there are only three algebras on R^2: the complex numbers satisfying $(0, 1)^2 = (-1, 0)$, the so-called dual numbers in which $(0, 1)^2 = (1, 0)$, and the numbers defined in ${}_1D_1$.

Standard functions like sin, exp, log are introduced on this structure in the following way. Let I be a differentiable standard function, then we define

$$I[(a_0, a_1)] := (I(a_0), a_1 \cdot I'(a_0)) \tag{4}$$

The forward mode of automatic differentiation now utilizes that the definition of sum, scalar product, product and functions are just codings of calculus rules for the respective operations. This entails that it is now possible to compute the derivative of a function described by a computational algorithm at x by evaluating it in ${}_1D_1$, beginning with the ordered pair $(x, 1)$. The resulting ordered pair will have the form $(f(x), f'(x))$, and hence the derivative is computed as the second component.

The arithmetic on ordered pairs can be generalized to allow the computation of **higher order derivatives of several variables**. To this end, one arranges the value and all requested derivatives of v variables and up to order n into a vector. A customary way of arranging the vectors is to begin with the value, then list all the first order derivatives $\partial/\partial x_1$,...,$\partial/\partial x_v$, then all second order derivatives $\partial^2/\partial x_1\partial x_1$, $\partial^2/\partial x_1\partial x_2$,...,$\partial^2/\partial x_1\partial x_v$,..., $\partial^2/\partial x_2\partial x_2$,..., $\partial^2/\partial x_v\partial x_v$, and then similarly with third and higher orders.

According to calculus rules, the derivatives of order n of a sum and product apparently only depend on the derivatives of order n of the summands or factors, respectively. Similarly, the derivatives of a standard function applied to an expression depend only on the derivatives of the function and the expression to the same order. The rules to calculate the derivatives of the results from the derivatives of the previous step now define the arithmetic on the vectors. The resulting structure is again a real algebra, denoted by ${}_nD_v$. Similarly, the rules to compute the derivatives of a standard function from the derivatives of its argument define the standard functions on the vectors.

It can be shown [7] that the resulting vector space has dimension $(n + v)!/(n!\ v!)$. Since it is finite dimensional, the resulting structure can not be a field. Again there are **nilpotent elements** and thus **zero divisors**: All elements that contain a zero in the first component vanish if raised to a power exceeding n.

It follows readily that the nilpotent elements actually form an ideal in the algebra which we denote by I_0. Besides I_0, there are n more ideals I_n which contain the vectors whose components vanish up to order n. We observe that these ideals form a tower such that $I_j \subset I_{j-1}$.

The algebras discussed in this introduction were all very practically motivated by the goal of computing derivatives, and they all belong to the so-called hypercomplex numbers discussed in [5]. The structures have certain defects, division is not always possible, and nilpotent elements have no roots. In practice this sometimes entails that automatic differentiation fails for certain differentiable functions. For example, the functions

$$f(x) = \frac{\sin(x)}{x}, \quad f(0) = 1 \quad \text{and} \quad g(x) = \frac{1 - \exp(x^2)}{x}, \quad g(0) = 0 \tag{5}$$

are all differentiable at the origin, yet the attempt to compute their derivatives using automatic differentiation fails.

In the next sections we will give extensions of these simple structures that remedy most of these problems. More importantly (at least from a purists point of view), they provide a very different view of automatic differentiation techniques as an application of analysis on a new system of numbers.

2. ORDERING AND INFINITESIMALS

The first step in the process towards a deeper understanding of the algebraic systems introduced by the automatic differentiation process is to introduce an ordering on the structures. Let us consider the most general structure, ${}_nD_v$. The vectors are to be arranged in the way outlined in the previous section. Then we define **positive numbers** ${}_nD_v^+$ on ${}_nD_v$ as follows: Starting from the left, we find the first nonzero component in the vector. If this component is greater than zero, we say the number is positive.

Such an ordering is called lexicographic, because we start comparing the components with zero from the left, and the first disagreement from zero already determines if an element is positive or not. From this definition, we quickly infer that

$$x = 0 \ \text{ or } \ x \in {}_nD_v^+ \ \text{ or } \ -x \in {}_nD_v^+ \ \text{(exclusive or)} \tag{6}$$

$$x \in {}_nD_v^+, 0 < t \in R \ \Rightarrow \ t \cdot x \in {}_nD_v^+ \tag{7}$$

$$x, y \in {}_nD_v^+ \ \Rightarrow \ x + y \in {}_nD_v^+ \tag{8}$$

$$x, y \in {}_nD_v^+ \ \Rightarrow \ x \cdot y \in {}_nD_v^+ \tag{9}$$

The first and second statements are obvious. The third statement follows because the first nonzero component of $x + y$ contains either the first component of x, the first component of y, or the sum of these. The last result follows because of the particular arrangement, the first nonzero component of $x \cdot y$ contains the product of the first nonzero components of x and y, which are both positive.

We now introduce an ordering in the following way: we say $x < y$ iff $y - x \in {}_nD_v^+$. Using the previous equations, we can immediately infer

$$\text{For } x, y \in {}_nD_v, \ \text{exactly one of } \ x < y, \ x > y, \ x = y \ \text{holds} \tag{10}$$

$$x < y \ \Rightarrow \ x + z < y + z \tag{11}$$

$$x < y, z > 0 \;\Rightarrow\; x \cdot z < y \cdot z \tag{12}$$

These conditions mean that the ordering is compatible with the arithmetic and hence is a **total ordering**. We note that the ring homomorphism embedding the reals into ${}_nD_v$ is also order preserving.

Let us now investigate the properties of the ordering. We begin by comparing the basis vectors e_j which have zero components except a 1 in the j-th component. Clearly we have $e_j > 0$. We also note that $e_1 = 1$, and all the other basis vectors are in some nilpotent ideal I_i. Let now $j < k$. We infer that $e_j > e_k$, but also

$$e_j > n \cdot e_k \;\forall\; n \in N. \tag{13}$$

Such a relation can never hold between real numbers; it means that e_k is infinitely small compared to e_j! In particular, we can infer that any e_k, $k > 1$ is infinitely small compared to any positive real number. Number systems in which the ordering allows cases as in (13) are called **nonarchimedean**. So nonarchimedean extensions of R like ${}_nD_v$ contain infinitely small numbers. We note that there is an extensive theory on ordered algebraic structures, and a good summary is contained in [8].

To conclude this section and to provide an outlook, let us now go back to the simple arithmetic on ordered pairs given by ${}_1D_1$. We denote $d = (0, 1)$ and note that d is infinitely small. Then the fact that evaluating a function f in ${}_1D_1$ instead of R yields its value in the first component and its derivative in the second component can be written as

$$f(r + d) \;=\; (f(r), f'(r)) \;=\; f(r) \cdot (1, 0) + f'(r) \cdot (0, 1) \;=\; f(r) + d \cdot f'(r). \tag{14}$$

This resembles $f(x+\Delta x) \approx f(x)+\Delta x \cdot f'(x)$, in which case the approximation becomes better and better for smaller Δx. Here we choose an infinitely small Δx, and the error turns out to be zero.

The following sections will provide a more detailed analysis of this interesting phenomenon and at the same time yield some interesting new calculus. The results obtained are similar to the ones in nonstandard analysis [9, 10, 11, 12, 13]; however, the number systems required here can be constructed directly and described on a computer, while the ones in nonstandard analysis are exceedingly large, non-constructive (in the strict sense that the axiom of choice is used and also in a practical sense), and require quite a machinery of formal logic for their formulation.

3. THE FIELD $\mathcal{L}$

In this section we will provide an extension of the structures encountered in the previous sections. We begin by defining a family of special subsets of the real numbers:

A subset M of the rational numbers is called **almost-finite**, iff below every bound there are only finitely many elements of M. With $\mathcal{F}$ we denote the family of all almost-finite sets. A few basic properties of almost-finite sets are as follows: Let $M, N \in \mathcal{F}$, then

$$M \neq \emptyset \Rightarrow M \text{ has a minimum} \tag{15}$$

$$X \subset M \Rightarrow X \in \mathcal{F} \tag{16}$$

$$M \cup N \in \mathcal{F} \tag{17}$$

$$M \cap N \in \mathcal{F} \tag{18}$$

$$M + N = \{x + y | x \in M, y \in N\} \in \mathcal{F} \tag{19}$$

$$x \in M + N \Rightarrow \ \exists \ \text{ only finitely many } (a, b) \in M \times N \text{ with } x = a + b \tag{20}$$

We now define the new set of numbers $\mathcal{L}$. These numbers were probably first studied by Levi-Civita [14, 15, 16], and their nice algebraic properties have been rediscovered many times, for example in [17, 18, 19, 20]:

$$\mathcal{L} = \{f : Q \to R | \{x | f(x) \neq 0\} \in \mathcal{F}\} \tag{21}$$

So $\mathcal{L}$ contains the functions from the rational numbers into R whose support is almost-finite. For the sake of clarity from now on we denote the function in $\mathcal{L}$ with x, y, ... and their values at $q \in Q$ with $x[q]$ etc. This is helpful to avoid confusion when functions on $\mathcal{L}$ are discussed. On the set $\mathcal{L}$ we now define **addition**:

$$(x + y)[q] \ = \ x[q] + y[q] \tag{22}$$

We note that the support of $x + y$ is contained in the union of the supports of x and y and is thus also almost-finite. We now define a **multiplication** in the following way: Let N_x and N_y be the supports of x and y. We set $(x \cdot y)[q] = 0$, if $q \notin N_x + N_y$. In case $q \in N_x + N_y$, we set

$$(x \cdot y)[q] \ = \sum_{\substack{q_x \in N_x, \ q_y \in N_y, \\ q_x + q_y = q}} x[q_x] \cdot y[q_y]$$

Since the support of $x \cdot y$ is contained in $N_x + N_y$, it is almost-finite. Furthermore, the sum in the definition of the product contains at most finitely many contributions.

On $\mathcal{L}$ one can introduce an **ordering** in a similar way as in the last section: we define the set $\mathcal{L}^+$ to be the set of all elements of $\mathcal{L}$ that have positive value at the smallest support point. Then again $\mathcal{L}^+ + \mathcal{L}^+ \subset \mathcal{L}^+$ and $\mathcal{L}^+ \cdot \mathcal{L}^+ \subset \mathcal{L}^+$, and we again say $x < y$ iff $y - x \in \mathcal{L}^+$. Altogether, the ordering is **total**.

With this ordering, $\mathcal{L}$ becomes nonarchimedean. It turns out that there are now both **infinitely small** and **infinitely large** numbers. In fact, all positive numbers whose first support point is positive are infinitely small, whereas those with negative first support point are infinitely large. Of particular interest are the elements d^r defined by

$$d^r[q] = \begin{cases} 1 & \text{if } q = r \\ 0 & \text{else} \end{cases} \tag{23}$$

In particular, d^r is infinitely small for $r > 0$ and infinitely large for $r < 0$. We note that the real numbers can be embedded in an order preserving way by mapping

$$r \in R \to x_r, \quad x_r[q] = \begin{cases} r & \text{if } q = 0 \\ 0 & \text{else} \end{cases} \tag{24}$$

The Veronese numbers (a_0, a_1) can also be mapped into $\mathcal{L}$ in an order preserving way:

$$(a_0, a_1) \in {}_1D_1 \to x_{(a_0,a_1)}, \quad \text{where} \quad x_{(a_0,a_1)}[q] = \begin{cases} a_0 & \text{if } q = 0 \\ a_1 & \text{if } q = 1 \\ 0 & \text{else} \end{cases} \tag{25}$$

In the latter case, after any algebraic operation defined in ${}_1D_1$ is performed in $\mathcal{L}$, the values at 0 and 1 agree with the corresponding values in ${}_1D_1$. Usually additional terms occur at the values of $q = 2, 3, 4, ...$ related to the fact that the element $(0, 1)$ is no longer nilpotent. In a very similar way, ${}_nD_1$ can be mapped into $\mathcal{L}$.

We also introduce an absolute value:

$$|x| = \begin{cases} x & \text{if } x > 0 \\ -x & \text{else} \end{cases} \tag{26}$$

It is now relatively simple to conclude that $\mathcal{L}$ is **Cauchy-complete** with this absolute value; to find the value of the limit at $q \in Q$, simply choose an n such that the terms of the Cauchy sequence do not differ by more than $\epsilon = d^{q+1}$ from n on, and define the value of the limit as $x_n[q]$. Since the limit agrees to an element of the sequence to the left of any q, its support is almost-finite and it is thus in $\mathcal{L}$.

A very important concept for the study of $\mathcal{L}$ is the **fixed point theorem** [20]: Let f be a function defined in an interval M around the origin, and let f be contracting with **infinitely small contraction factor** q, then f contains a unique fixed point in M.

The proof is very similar to the Banach space case. One begins with an arbitrary element x_0, and defines $x_{i+1} = f(x_i)$. The resulting sequence is Cauchy. Let x denote its limit; it follows that x is the fixed point.

This fixed point theorem allows to show several important properties of $\mathcal{L}$. In particular, we can show that $\mathcal{L}$ **is a field**. To prove this, assume we are interested in the inverse of a nonzero $X \in \mathcal{L}$. We write $X = x_0 \cdot d^r \cdot (1 + x)$, where x_0 is real and x is infinitely small. Since $x_0 \cdot d^r$ has the inverse $x_0^{-1} \cdot d^{-r}$, it suffices to find an inverse to $(1+x)$. Write this inverse as $(1 + y)$, and conclude

$$\begin{aligned} 1 &= (1 + x) \cdot (1 + y) \Leftrightarrow \\ 1 &= 1 + x + y + x \cdot y \Leftrightarrow \\ y &= -x - x \cdot y \end{aligned} \tag{27}$$

This is a fixed point problem for y with the function $f(y) = -x - x \cdot y$. Since x is infinitely small, f is contracting on $M = \mathcal{L}$ with infinitely small contraction factor, and thus there is a unique fixed point.

Note also that not only does the fixed point theorem guarantee the existence of an inverse, it also allows to compute it in a rather direct way. Indeed, to determine the inverse up to a certain depth, all that is required is to iterate f sufficiently often.

In a similar way, we can show the existence of **roots of positive elements** and provide an algorithm to compute them. Again we write $X = x_0 \cdot d^r \cdot (1+x)$ and note that $x_0 \cdot d^r$ has the root $\sqrt{x_0} \cdot d^{r/2}$. For the root of $1 + x$ we try $1 + y$ and obtain

$$
\begin{aligned}
(1+x) &= (1+y)\cdot(1+y) \Leftrightarrow \\
x &= 2y+y^2 \Leftrightarrow \\
y &= x/2 - y^2/2 = f(y)
\end{aligned} \tag{28}
$$

Limiting ourselves to the interval around the origin that contains all the numbers whose square is not infinitely much larger than x, we again obtain that f is contracting. Hence there is a root; and furthermore, iteration of f provides an elegant way to compute it to any depth q.

Clearly the algorithm can be modified to compute higher roots. The fixed point theorem can also be used to prove that the structure obtained from $\mathcal{L}$ by **adjoining the imaginary unit** is **algebraically closed**, i.e. all polynomials have roots. The proof of this theorem is much more involved than the two cases discussed here and certainly goes beyond the scope of this paper. For details, we refer to [19, 20].

4. FUNCTIONS ON $\mathcal{L}$

In this section we want to introduce functions on $\mathcal{L}$. The algebraic properties of $\mathcal{L}$ already allow for a **direct introduction of polynomials, rational functions, roots**, and any combination thereof. Besides these conventional functions, $\mathcal{L}$ readily contains **delta functions**. For example,

$$f(x) = \frac{d}{d^2+x^2} \tag{29}$$

assumes the infinitely large value d^{-1} at the origin, falls off as $|x|$ gets larger, is infinitely small for any real x, and becomes smaller yet for infinitely large x.

For the scope of this paper, however, it is more important to study the extendability of standard functions, in particular all **power series**. To this end, it is helpful to study two kinds of convergence. The first type of convergence is the one defined by the topology induced by the ordering. It is called **strong convergence**, and above we saw that $\mathcal{L}$ is complete under strong convergence.

Besides this kind of convergence, there is another kind, the so-called **weak convergence**. We say that the sequence x_n converges weakly to the limit $x \in \mathcal{L}$ if for all $q \in Q$, $x_n[q] \to x[q]$ for $n \to \infty$. So weak convergence is coordinatewise. It follows rather directly that strong convergence implies weak convergence, but not vice versa. It will be the weak convergence that will allow us to generalize power series to $\mathcal{L}$.

To this end, we first make an observation about almost finite sets. Let M be almost finite, and if $M \neq \emptyset$ let the minimum of M be non-negative. Define

$$M_\sigma = \{x | \exists k \in N, m_1, ..., m_k \in M \text{ with } x = m_1 + ... + m_k\} \tag{30}$$

So M_σ is the set of all rational numbers which can be obtained by adding finitely many elements of M. Then M_σ **is almost finite**.

The proof is simpler than it may appear; first note that if $0 \in M$, it does not contribute to any sum, so we may actually assume that M has only positive elements. Since it is almost finite, it has a smallest element q_M, and $q_M > 0$. Let now $q \in Q$ be given; let n such that $n \cdot q_M > q$. If it is our goal to obtain elements of M_σ that are less than q, we have to restrict ourselves to the finitely many elements of M that are less then q. But we can also never add more than n of these, because by doing so q is exceeded. So altogether we only have finitely many ways of writing sums that do not exceed q, and thus M_σ is almost finite.

Now let $\sum a_n x^n$ be a **power series** with real coefficients and conventional radius of convergence r. Let $x \in \mathcal{L}$, $x < r$, and let M be the support of x. One can show [19, 20] that the partial sum $x_m = \sum_{n=1}^{m} a_m x^m$ actually **converges weakly** inside the classical radius of convergence. Further we note that the support points of $\sum a_n x^n$ are all in M_σ, and hence the weak limit is actually an element of $\mathcal{L}$.

This procedure allows the automatic generalization of any power series within its radius of convergence to the field $\mathcal{L}$. So in a simple way, we have a very large class of functions readily available. In particular, this includes all the conventional intrinsic functions of a computer environment.

5. DIFFERENTIATION

In this section we will define differentiation on $\mathcal{L}$, which will allow us to view the algorithms of automatic differentiation in a different light. Furthermore, we will provide a method to perform automatic differentiation in cases when the conventional methods fail. We begin with the definition of differentiability.

Let f be a function on a subset of $\mathcal{L}$. We say f is **differentiable** with derivative $f'(x)$ at the point x in M, if for any $\epsilon > 0 \in \mathcal{L}$ there is a $\delta > 0 \in \mathcal{L}$ with δ/ϵ not infinitely small such that

$$|\frac{f(x+\Delta x) - f(x)}{\Delta x} - f'(x)| < \epsilon \tag{31}$$

for all Δx with $x + \Delta x \in M$ and $|\Delta x| < \delta$. So this definition very much resembles the conventional differentiability; an important difference being the restriction imposed by requiring δ not to become too small. This restriction, which is automatically satisfied in archimedean structures, was first studied in [19] and will prove crucial to making the concept of differentiation useful.

It turns out that the usual **rules for sums and products** hold in the same way as in the real case, with the only exception that factors are not allowed to be infinitely large. Furthermore, it follows readily that if f coincides with a real function on all real numbers and is differentiable, then so is the real function and the derivatives agree at the real point up to an infinitely small error. This will allow the computation of derivatives of real functions using techniques of $\mathcal{L}$.

A very important consequence of the definition of derivatives is the Fundamental Theorem: **Derivatives are differential quotients** up to an infinitely small error.

Let $\Delta x \neq 0$ be a differential, i.e. infinitely small. Choose $\epsilon > 0$ infinitely small such that $|\Delta x|/\epsilon$ is also infinitely small. Because of differentiability, there is $\delta > 0$ with δ/ϵ finite such the difference quotient differs from the derivative by an infinitely small error of less than ϵ for all ordinate differences less than δ. But since δ/ϵ is finite and $|\Delta x|/\epsilon$ is infinitely small, we have $|\Delta x|/\delta$ infinitely small, and in particular $|\Delta x| < \delta$. Thus Δx yields an infinitely small error in the difference quotient.

This elegant method now allows to compute the real derivative of any real function that has been extended to the field $\mathcal{L}$ and is differentiable there. In particular, all real functions that can be expressed in terms of power series functions combined in finitely many operations can be conveniently differentiated in this way. But note that it also works for the cases discussed above where automatic differentiation fails. Furthermore, it is of historical interest since it retroactively justifies the ideas of the fathers of calculus of derivatives being differential quotients. It is worth pointing out that the computation of exact derivatives as real parts of difference quotients corresponds to the result in Eq. (14), except that there division by d is impossible leading to the different form of the expression.

To continue our overview over analysis on $\mathcal{L}$, the details of which can be found in [20], we present an **intermediate value theorem**: Let f be defined on the interval $[a, b] \subset \mathcal{L}$ and let f be differentiable there. Let $f(x)$ be finite and $f'(x)$ be nonzero and finite in the interval. Then f assumes every intermediate value between $f(a)$ and $f(b)$.

It turns out that the proof can be obtained in a rather elegant way from the fixed point theorem. We assume that S lies between $f(a)$ and $f(b)$. Let S_R be the real part of S. Consider now the function f_R obtained by restricting f to R. Then f_R is continuous as a real function, and thus assumes S_R as a real intermediate value. Let X be the real point at which the real intermediate value is assumed. Then we have

$$s = S - f(X) = (S - S_R) + (S_R - f_R(X)) + (f_R(X) - f(X)) \tag{32}$$

and hence s is infinitely small. We now search for an infinitely small x such that $S = f(X + x)$. Because of differentiability it follows that

$$S = f(X + x) = f(X) + f'(X) \cdot x + r(x) \cdot x^2, \tag{33}$$

where $r(x)$ is actually at most finite [19], and by assumption $f'(X)$ is finite. Combining the last two equations yields

$$f'(X) \cdot x + r(x) \cdot x^2 = s \tag{34}$$

In case one can find an infinitely small x satisfying this equation, $X + x$ is the desired point where the intermediate value is assumed. We rewrite the equation as a fixed point problem:

$$x = \frac{s}{f'(X)} - \frac{r(x)}{f'(X)} \cdot x^2 \tag{35}$$

We now choose M to be the set of numbers which are not infinitely much larger than s. Since by assumption $f'(X)$ is finite and $r(x)$ is at most finite, the function on the right hand side is contracting with an infinitely small contraction factor; thus there is a fixed point and hence an intermediate value.

The intermediate value theorem is only the beginning of an analysis on $\mathcal{L}$, but this is not the place to present more advanced results. We just want to mention that it is possible to prove an equivalent of **Rolle's theorem**. We also obtain an equivalent of **Taylor's theorem**.

The last result we want to mention here is called **Cauchy's Pointformula**. Let $f = \sum_{i=0}^{\infty} a_i (z-z_0)^i$ be a power series with real coefficients. Then the function is **uniquely determined** by its value at a point z_0+h, where h is an arbitrary nonzero infinitely small number.

For the proof note that $f(z_0+h) = \sum_{i=0}^{\infty} a_i h^i$ Let $h \approx h_0 d^r$, $h_0 \in R, r \in Q^+$, then we conclude

$$\begin{aligned} a_0 &= (f(z_0+h))[0], \\ a_1 &= (f(z_0+h))[r]/h_0, \\ a_2 &= (f(z_0+h) - a_1 h)[2r]/h_0^2, \\ a_3 &= (f(z_0+h) - a_1 h - a_2 h^2)[3r]/h_0^3, \\ &\dots \end{aligned} \tag{36}$$

hence the coefficients can be computed. Choosing $h = d$ yields the particularly simple form $a_i = f(z_0+d)[i]$.

This formula allows the computation of derivatives of any function which can be written as a power series with nonzero radius of convergence; this includes all differentiable functions obtained in finitely many steps using arithmetic and intrinsic functions.

6. IMPLEMENTATION

Besides allowing illuminating theoretical conclusions, , the strength of the Levi-Civita numbers is that they can be used in practice, and even in a computer environment. In this respect, they differ from the non-constructive structures in Nonstandard Analysis.

An implementation of the Levi-Civita numbers is not as direct as one of the algebras of automatic differentiation since the Levi-Civita field is infinite dimensional. However, as we shall see now, it is still possible to implement the structure in a very useful way. Since there are only finitely many support points below every bound, it is possible to pick any such bound and store all the values of a function to the left of it. So each "number" is represented by these values as well as the value of the bound.

The sum of two such functions can then be computed for all values to the left of the minimum of the bounds; so the bound of the sum is the minimum of the bounds. In a similar way it is possible to find a bound below which the product of two numbers can be computed from the bounds of the two numbers. Altogether, the bound to which each individual variable is known is carried along through all arithmetic.

There is actually an illuminating similarity to the implementation of the real numbers in floating point format on a computer. In their decimal representations, the reals cannot be represented exactly because they have infinitely many digits. Instead, we store the digits to a certain depth, the mantissa length, and ignore the rest. Finding the mantissa of a sum or product of reals can be done by just manipulating the digits of the mantissas of the two operands. In the case of the real numbers, this arithmetic can obviously introduce errors because of the carry operation which produces a feed up from lower to higher mantissa places. Interestingly, since there is no equivalent to the carry in the addition or multiplication in $\mathcal{L}$, this problem does not exist. So there is no loss of accuracy over the course of computation, except the one in the real number values at the support points of the functions.

References

[1] A. Griewank, G. F. Corliss, and Eds. *Automatic Differentiation of Algorithms.* SIAM, Philadelphia, 1991.

[2] L.B. Rall. The arithmetic of Differentiation. *Mathematics Magazine*, 59:275, 1986.

[3] G. Corliss and L. B. Rall. Automatic generation of taylor series in PASCAL-SC: Basic applications to ordinary differential equations. In *Transactions of the First Army Conference on Applied Mathematics and Computing*, volume ARO Report 84-1, 1984.

[4] A. Griewank. On automatic differentiation. Technical Report MCS-P10-1088, Argonne National Laboratory, 1988.

[5] G. Kantor and I. Lovovich. *Hyperkomplexe Zahlen.* Teubner.

[6] J. F. Ritt. *Differential Algebra.* American Mathematical Society, Washington, D.C., 1950.

[7] M. Berz. Differential algebraic description of beam dynamics to very high orders. *Particle Accelerators*, 24:109, 1989.

[8] L. Fuchs. *Partially Ordered Algebraic Systems.* Pergamon Press, Addison Wesley, 1963.

[9] C. Schmieden and D. Laugwitz. Eine Erweiterung der Infinitesimalrechnung. *Mathematische Zeitschrift*, 69:1–39, 1958.

[10] A. Robinson. Non-standard analysis. In *Proceedings Royal Academy Amsterdam, Series A*, volume 64, page 432, 1961.

[11] A. Robinson. *Non-Standard Analysis.* North-Holland, 1974.

[12] M. Davies. *Applied Nonstandard Analysis.* John Wiley and Sons, 1977.

[13] K. D. Stroyan and W. A. J. Luxemburg. *Introduction to the Theory of Infinitesimals.* Academic Press, 1976.

[14] Tullio Levi-Civita. Sugli infiniti ed infinitesimi attuali quali elementi analitici. *Atti Ist. Veneto di Sc., Lett. ed Art.*, 7a, 4:1765, 1892.

[15] Tullio Levi-Civita. Sui numeri transfiniti. *Rend. Acc. Lincei*, 5a, 7:91,113, 1898.

[16] D. Laugwitz. Tullio Levi-Civita's work on nonarchimedean structures (with an Appendix: Properties of Levi-Civita fields). In *Atti Dei Convegni Lincei 8: Convegno Internazionale Celebrativo Del Centenario Della Nascita De Tullio Levi-Civita*, Academia Nazionale dei Lincei, Roma, 1975.

[17] L. Neder. Modell einer Leibnizschen Differentialrechnung mit aktual unendlich kleinen Größen. *Mathematische Annalen*, 118:718–732, 1941-1943.

[18] D. Laugwitz. Eine nichtarchimedische Erweiterung angeordneter Körper. *Mathematische Nachrichten*, 37:225–236, 1968.

[19] M. Berz. Analysis auf einer nichtarchimedischen Erweiterung der reellen Zahlen. Report (in German) MSUCL-753, Department of Physics, Michigan State University, 1990.

[20] M. Berz. Analyis on a nonarchimedean extension of the real numbers. Technical Report MSUCL-753, National Superconducting Cyclotron Laboratory, Michigan State University, East Lansing, MI 48824, 1990.