A radical approach to the role of computational thinking in teaching Mathematics

J A Vallejo
Faculty of Sciences, State University of San Luis Potosí, Lomas de Chapultepec 1570, San Luis Potosí (SLP) 78295, México
E-mail: jvallejo@fc.uaslp.mx

Abstract. I present an incomplete, partially biased and highly personal set of reflections about the role of computational thinking in teaching courses on mathematics for students of science (mainly non-mathematicians). My aim is to promote the adoption of some form of strict finitism as a good choice for these courses.

1. Introduction

In this note, I discuss the contents and methods currently in use in courses of mathematics for students of science and engineering. Of course, this requires a previous explanation of the need for such discussion. Are the contents of the current courses inappropriate? Insufficient? Both? To answer these questions, let me remark the fact that every year new generations of engineers, physicists, biologists, etc, come out from our universities, and they develop their professional careers without apparent problems. However, it is a widespread opinion that most of the topics studied in their math courses are useless, too abstract to be connected to their everyday life, and unrelated to the current needs in their respective fields. The reader can check this by herself, doing a quick poll with her own students (asking previously for honesty in answers!), but if in doubt, [26] reports that

All engineering students responding to the questionnaire (15.2% of the variability) feel that mathematics teaching is too theoretical, is not practical enough and does not have enough connection with other sciences and the reality of an engineer’s job.

See also [10], where it is stated that, regarding the standard course of mathematics for engineers,

The lowest agreement (59%) was regarding the success of the course in teaching students how to formulate and solve problems that are directly related to engineering.

Thus, it seems that most students of these scientific fields go ahead despite the mathematics courses they are taught, and not offered by them, as it would be desirable.

In a sense, the situation described in the preceding paragraph was to be expected. If we open any calculus book published in the last, say, 50 years, we will see that taking apart funny photos, grayed text boxes ans supposedly “real life examples”\(^1\), what is left out is, basically, A. Cauchy’s Course d’Analyse\([6]\) which dates back to 1821.

\(^1\) In many texts that claim to contain this kind of examples, one can find statements like the following (taken almost randomly from a very popular calculus text): “You are a commodities trader, and you monitor the price
The mathematics created in the first half of the 19th century were a particular response to the needs of the technical advances that were taking place at that time. The design and control of mechanical and thermal machines required by the Industrial Revolution, led scientists to look at a kind of problems quite different to those to which they were used to (planetary motion, and the like). Modeling the operation of real machines (with heat dissipation, friction and all kinds of physical constraints) urged the development of new mathematical frameworks, which often collided with the mathematical establishment, unable to see the importance of the new ideas. An instructive example of that state of affairs is provided by J. Fourier’s theory of heat [11] which, even being greatly adapted to the problems appearing in nature, giving precise and useful solutions to many practical problems\(^2\), was dismissed as being “not rigorous” [7].

For another example, we can refer to O. Heaviside’s operational calculus, intensively used by physicists and engineers but rejected by many mathematicians until the advent of L. Schwartz’s theory of distributions [25].

These, and other, examples illustrate why so many scientists adhere to the idea that quite often mathematicians arrive late to the party, and then it is only to add a heavy layer of definitions and theorems which are mostly irrelevant to solve practical problems. It is well known, in this regard, the opinion expressed by R. Hamming, who once said [14]

> If whether an airplane would fly or not depended on whether some function that arose in its design was Lebesgue but not Riemann integrable, then I would not fly in it.

Some physicists, like R. Feynman [21], even think that

> If all mathematics disappeared today, physics would be set back exactly one week.

On the other hand, we can consider the situation faced by a biologist studying DNA. Maybe she needs to sequence a certain gene, given by a string of length \(2 \cdot 10^6\) composed of A, C, G, T letters, and she needs to do it quickly. This implies searching for repeating patterns inside that big string. When asking the average mathematician how to do that, it is not clear that she will receive a useful answer, one related to efficient analysis techniques (like shotgun sequencing [29]). What is almost sure, is that she will not find the right answer by looking at the contents of the math courses she studied as an undergraduate.

These considerations suggest that the answer to the questions posed at the beginning of the section is: the contents of the current courses are both inappropriate and insufficient. But, even more, we will argue that it is not the contents what we should change, but the whole approach to the mathematics behind them. The reason is the following: while the vast majority of scientific disciplines relies on big volumes of data and really complex models (when realistic) which can only be treated with the aid of a computer\(^3\), we are still insisting on training our students in the use of good old tools from the Cauchy era. Not only that, we actually impose

of gold on the spot market during an active morning. Suppose you find that the price of an ounce of gold can be approximated by the function \(G(t) = 5t^2 - 85t + 1762\) \((7.5 \leq t \leq 10.5)\), where \(t\) is time in hours.\(^4\) The student is then asked to answer some question that require computing the derivative of \(G(t)\). However, nothing is said about how to obtain that function from the discrete data provided by observations, and the fact that the knowledge of the actual, discrete, data is enough to answer all the really meaningful questions about the price of gold, is carefully hidden.

\(^2\) Fourier stated in the Preliminaries to his work that

> Ces mémoires auront pour objet la théorie de la chaleur rayonnante, la question des températures terrestres, celle de la température des habitations, la comparaison des résultats théoriques avec ceux que nous avons observés dans diverses expériences, enfin la démonstration des équations différentielles du mouvement de la chaleur dans les fluides.

\(^3\) Recall the title of [20]!
on them a certain mindset, appropriate to use these tools, but completely useless for many real
life problems. Thus, we train our students in the delicate intricacies related to the infinite, and
then introduce the notion of limit as a way to tame them. But when facing a real set of data,
there are no infinities, no limiting processes at all. Most of the time, we will use a computer, and
for a computer there is only a finite set of numbers, all of them rationals with a finite decimal
expansion. Given that this will be the final stage in practically all the problems found in real
life, why we do not teach directly how to solve problems in a computer-guided way? Why we
do not teach our students to think in finite (finitistic?) terms, as a more appropriate mindset?
By sticking to the traditional ideas, based on the consideration of a continuum of numbers,
the existence of limiting processes, and the like, we are staying away from the full power that
computers put at our hands. It is as if a biologist insisted on making observations with the
naked eye, dismissing a microscope. A (yet another one!) quotation of P. D. Lax [20] comes to
mind

It is impossible to exaggerate the extent to which modern applied mathematics
has been shaped and fueled by the general availability of fast computers with large
memories. Their impact on mathematics, both applied and pure, is comparable to the
role of telescopes in astronomy and microscopes in biology.

The rest of this note will be devoted to an exposition of what could be called Practical Strict
Finitism (PSF): I maintain that we can improve a lot students’ understanding of mathematics if
we present them in a constructive manner, avoiding potential entities like infinities, allowing for
an approach which is closer to the way scientists see and use mathematics in their work, based
on an exposition of constructive computational techniques for any mathematical concept that is
introduced in the discourse. These ideas are on solid although not well-known grounds [31], and
one of my goals is to promote them by way of showing some examples that can be interesting
both for students and teachers. If I do not reach it, it will be not because the subject is not
interesting, but due to my lack of competence. Fortunately, there are much more authoritative
voices than mine advocating for similar approaches (this sentence does not imply, in any way,
that they share or support my views!) Besides [31], the reader should consult all the information
available in the personal page of Doron Zeilberger [32] (the main idea for Example 4 came from
[33]), as well as the papers [19, 28], which discuss the relation between finitism and physics
(through the notion of feasible numbers). For a constructive approach of classic (i.e, allowing
infinities) calculus, see [3]. The finite version of the real numbers we will use here is a particular
case of a time scale in the sense of [1, 15, 18], to which we refer for a more detailed treatment
of calculus and differential equations.

2. A proposal based on PSF

As implied by the first adjective in the denomination, Practical Strict Finitism leaves aside all
the foundational questions and focus on the pedagogical use of Strict Finitism. There are two
really simple basic principles in this proposal:

(a) Teach mathematics that are useful.
(b) For any new concept introduced in the exposition, a practical procedure to compute it (in a
computer) must be provided. Considerations about “potential entities”, such as infinities,
must be avoided.

4 Please, note carefully that I do not mean that we simply teach our students how to use a computer! There is
more to it.

5 Thus, the adjective “radical” in the title takes its etymological meaning of “being at the roots”.
That is all. Notice that while the first principle is context-dependent and somewhat ambiguous\(^6\), the second one rules out the use of limits, for instance. This seems to be hard to accept for many people, so let me discuss a couple of arguments I have heard from colleagues opposed to these ideas.

1. **Mathematics should be taught not only for its practical applications, but also for its intrinsic beauty.** A variant of this argument is that non applicable mathematics also have a value in molding students’ logical reasoning and critical thinking. The first variant is simply untenable in front of a physicist or a biologist, as they are not interested in the aesthetical aspects of mathematics but in what can they offer to solve their own problems. Interestingly enough, it seems that the layman can be impressed by what B. Russell described as “a beauty cold and austere, like that of sculpture” \[^{17}\], but this is not the case with many professional scientists\[^{16}\]. The second variant can be applied to many other disciplines: in what respect are mathematics better suited to develop logical thinking than reading Conan Doyle, or a book on compared linguistics?

2. **Excluding infinities also excludes limits and hence calculus\[^{8}\].** This seems to contradict the first principle, as the notion of limit was introduced because of their usefulness; for example, Newton invented calculus to study planetary motion, as contained in his Principia. This sounds good, but it is wrong. To begin with, Newton did not use calculus when writing the Principia. Calculus is not present in any page of the Principia \[^{24}\], and there is evidence that Newton had not used it formerly to write the results contained in this book \[^{9}\], see \[^{8}\]. On the other hand, Newton’s calculus is not the same calculus we use today (which, as stated before, was formulated by Cauchy in the 19th century.) Indeed, Newton introduced a number of computational procedures (divided differences, iterative method for finding roots of \(f(x) = 0\), etc) that fit neatly into finite calculus, as we will see below.

3. **Excluding limits and the like discards many of the advances in mathematics done in the last two centuries, it is a regression to a state of primitive knowledge.** First of all, this is not an excluding proposal: If really needed, one can easily explain what happens in the continuous case and develop all the theorems as usual, starting from the finite setting. What I am saying is that we can recast all that knowledge from the last two centuries in a form which is more suitable for scientists other than mathematicians, and that we can do it taking benefit of the availability of computers. On the other hand, the insistence in the continuous setting can be the origin of some conceptual difficulties. For instance, a student of physics, well used to the phenomenon of chaos, can find it surprising to learn that no chaos is possible in one-dimensional systems described by ordinary differential equations, yet one-dimensional chaos is ubiquitous in physics (you have one such chaotic system in

---

\(^6\) Deliberately ambiguous. I think it is undesirable to set once and for all a standard of what should be taught, but it is apparent that there must be some guiding principle in choosing the topics for a concrete course.

\(^7\) S. Glashow, a Nobel price in Physics, described his feelings about this in the preface of the book \[^{12}\]:

> Here, the reader will find no narcissistic cry of mathematics for mathematics’ sake. To the contrary, we physicist-chauvinist-pigs regard mathematics as the mere handmaiden of physics. Flights of mathematical fancy are tolerated only insofar as they are tethered to observable physical phenomena.

\(^8\) Apparently, see Section 4 below.

\(^9\) It has been suggested that Newton used calculus first and then rewrote his findings in the geometrical language that was fashionable in his times. In this regard, V. I. Arnol’d mentions in \[^{2}\] that

> It must be said that Newton made all the discoveries contained in the Principia without using analysis, although he had command of it at this time. He proved everything that was required by means of direct elementary geometrical arguments more or less equivalent to analysis (and not by translating analytical calculations into geometrical language).

For those interested in, a modern presentation of Newton’s original derivations can be found in \[^{27, 22}\].
your kitchen [5]). In this case, it is the limitations of the continuous approach what is excluding some real physical phenomena!

3. Mathematics with a finite set of numbers

Any rational number $x$ can be uniquely expressed as

$$x = s \cdot m \cdot 10^e,$$

where $s \in \{-1, +1\}$ is its sign, $m \in [1, 10]$ is the mantissa (or significand) of the number, and $e$ is the exponent. For instance, $-237/126$ can be written as

$$-\frac{3237}{126} = -25.69047619047619 = -2.569047619047619 \cdot 10^1,$$

hence $s = -1$, $m = 2.569047619047619 \in [1, 10]$ and $e = 1$. This is the representation in the floating point decimal arithmetic. The number of digits of the mantissa $m$ is the number of decimal digits that we can write within precision using this representation. We also have a floating point binary arithmetic; in this case, the rational $x$ is uniquely expressed in the form

$$x = \sigma \cdot \bar{x} \cdot 2^e,$$

where, again, $\sigma \in \{-1, +1\}$, but now the mantissa satisfies $\bar{x} \in [1, 10_2]$, with the subindex 2 denoting the use of a binary base$^{10}$, and the exponent is biased.

For instance, the number $-25.69047619047619$ would be written as

$$-\frac{3237}{126} = -11001.10110000110000110001_{2}$$

$$= -1.100110110000110000110001_{2} \cdot 2^4. \quad (1)$$

The kind of precision we use will impose restrictions on: (a) the number of binary digits of the mantissa, (b) the number of binary digits of the exponent.

The double precision is a standard established by the IEEE, consisting in the use of 64 bits (or 8 bytes) to represent numbers in a computer. Of these:

(i) One bit is reserved for the sign (positive if 0, and negative if 1).
(ii) The mantissa can be represented with 53 bits, of which only 52 need to be stored.
(iii) The exponent can be represented by a number $-1022 \leq e \leq 1023$, that is, it needs another 11 bits (including its sign).

**Remark 1.** What is the meaning of the sentence “the 53 bits of the mantissa can be stored in 52 bits”? This is due to a normalization: the mantissa of a binary number, as we have seen, is always a binary fraction comprised between $1_2$ and $10_2$, so its expression will always have the form of a 1 followed by the fractionary dot and then 53 binary digits. Then, it is not necessary at all to physically store in memory the leading 1: it is implicitly understood that it is there, this fact is taken into account when it comes to define operations and algorithms involving the numbers so expressed, but that 1 is never stored. In this way, it is possible to have mantissas with 53 digits stored in 52 bits of memory. Another restriction is that the mantissa cannot have a leading 0 except for 0 itself (whenever that happens, a shift in the biased exponent is applied).

$^{10}$ Another possibility is offered by the theory of computation, where prefixes 0b and 0x are used to denote binary or hexadecimal systems, respectively. Thus, the condition on the mantissa would be written $\bar{x} \in [0b1, 0b10]$. 
Thus, a number in double precision floating point binary arithmetic will be written as

\[ x = \sigma \cdot 1.d_1d_2\ldots d_{52} \cdot 2^e. \]

It is easy to see that the greatest number that can be stored with this method is \( M = 1.797693134862316 \cdot 10^{308} \)

**Remark 2.** The fact that the preferred base is 2 instead of 10 or any other number, has to do with the inner working of a computer: as integrated electronic circuits are used for their function, it is easy to represent 0 by the state given by a certain electric potential \( V_0 \) and 1 by another electric potential \( V_1 \), both values \( V_0, V_1 \) fixed. However, nothing prevents us from creating numerical systems based on floating point arithmetic with other parameters. Thus, by

\[ \mathbb{F}(b, p, e_{\min}, e_{\max}) \]

we will denote a system in which the base (or radix) is \( b \), the precision is determined by the number \( p \) of digits\(^{11}\) of the mantissa, and the minimum and maximum values of the exponent by \( e_{\min}, e_{\max} \), respectively. That implies that, in this system, any number (different from zero) will be written

\[ \sigma \cdot m \cdot b^e \]

with \( \sigma \) representing the sign, \( m \in [b^{-1}, 1 - b^{-p}] \) the mantissa and \( e \in [e_{\min}, e_{\max}] \) the exponent.

A way of measuring the precision with which numbers are stored in a computer is provided by the machine epsilon, denoted \( \text{eps} \). In any precision (simple, double) it is defined as the distance between 1 and the next number (greater than 1) that can be stored in that format. The \( \text{eps} \) in double precision is theoretically determined as \( \text{eps} = 2^{-52} \approx 2.22 \cdot 10^{-16} \), which implies that rational numbers can be stored with precision up to 16 decimal digits. Notice that if \( \tau \) is any number and \( t \) is the representable number closest to it, then, from the definition of \( \text{eps} \), we have

\[ (1 - \text{eps})t \leq \tau \leq (1 + \text{eps})t, \]

hence

\[ -\text{eps} \leq \frac{\tau - t}{t} \leq \text{eps}, \quad (2) \]

so \( \text{eps} \) gives a measure of the relative error committed when rounding.

It is possible, given any other number \( t \), to define its associated epsilon \( \text{eps}_t \) as the distance between that number and its immediate successor (greater in absolute value). In particular, \( \text{eps}_0 \) will be the smallest positive number that can be represented by a given precision. In double precision, this number is \( \text{eps}_0 = 2.225073858507201 \cdot 10^{-308} \). Taking \( \tau = t + \text{eps}_t \) in (2) we get the useful bound

\[ \left| \frac{\text{eps}_t}{t} \right| \leq \text{eps}. \quad (3) \]

The existence of \( \text{eps} \) leads to the following considerations: If in any numerical system representing the real numbers by floating point arithmetic there are a minimal number and a maximal one and, moreover, between a number and its successor there is a finite non-zero distance, then it follows that the quantity of numbers representable in the computer is finite. Those numbers that can be represented in floating point arithmetic are called machine numbers, and they form the discrete version of \( \mathbb{R} \) with which the computer works. Observe that all of them are rationals with finite decimal expansion.

\(^{11}\) This does not mean that \( p = 24 \) amounts to a precision of 24 decimal digits. As we will see in the next section, \( p = 24 \) corresponds to simple precision, and in this case we only have 7 correct decimal digits.
Theorem 1. If $F(b,p,e_{\text{min}},e_{\text{max}})$ is a floating point arithmetic numerical system, the quantity of machine numbers it contains is given by

$$N = 2(b - 1)b^{p-1}(e_{\text{max}} - e_{\text{min}} + 1).$$

Proof. For any given numerical system, $F(b,p,e_{\text{min}},e_{\text{max}})$, the quantity of its machine numbers can be computed as

$$N = \# \text{signs} \cdot \# \text{mantissas} \cdot \# \text{exponents}.$$

The number of possible signs is 2. Notice that in the interval $[-1,1]$ there are 3 integers ($-1$, 0 and 1) and, in general, in $[e_{\text{min}},e_{\text{max}}]$ there are $e_{\text{max}} - e_{\text{min}} + 1$ integers. Regarding the mantissas, there are $b^p$ a priori possibilities, but the floating point format requires that the mantissa does not have a leading zero, ruling out $b^{p-1}$ of them. Thus, there are $b^p - b^{p-1} = (b - 1)b^{p-1}$ possible mantissas.

In the case of binary arithmetic with double precision, $F(2,53,-1022,1023)$, the number $N$ in the preceding theorem is

$$N = 2^{43} \cdot 2046 \simeq 1.84 \cdot 10^{19}.$$

The machine numbers are distributed along the real line in a non-uniform way, they accumulate around zero, and become more sparse as we move away from it.

Example 1. We can visualize the distribution of machine numbers by considering the system $F(2,2,-2,3)$. Its machine numbers have the form $\pm 1.d \cdot 2^e$ and, by the preceding theorem, there are $2 \cdot 2 \cdot (3 + 2 + 1) = 24$ of them. In fact, they are the numbers

$$\pm 1.02 \cdot 2^e \text{ and } \pm 1.12 \cdot 2^e.$$

These can be converted to decimal to yield

$$\pm (1 + 0.0) \cdot 2^e \text{ and } \pm (1 + 0.1) \cdot 2^e,$$

that is, the 24 decimal numbers

$$\pm 1.0 \cdot 2^e \text{ and } \pm 1.1 \cdot 2^e,$$

where $e \in [-2,-1,0,1,2,3]$. In a computer algebra system, like Maxima, they can be generated through a command similar to this one:

```maxima
(%i1) lista:append(
create_list([s*2^ee,0],s,[1,-1],ee,makelist(-2+j,j,0,5)),
create_list([s*1.1*2^ee,0],s,[-1,1],ee,makelist(-2+j,j,0,5))
)$
```

The output already has the form of a list of points $(x_i,0)$, ready to be displayed:

```maxima
(%i2) plot2d([discrete,lista],[style,[points,1.5,1]]);
(%o2)```

7
Notice how the machine numbers are not uniformly distributed, and how they accumulate around zero.

Once the manner in which numbers are stored in the computer is decided, there is a whole floating point arithmetic available to implement the basic operations, which is well known in computer science [23]. We will not deepen into this topic in this introductory work.

4. Finite Calculus

As stated in the Introduction, the finite version of the real numbers provided by floating point arithmetic in double precision is an example of a time scale 12. In order to be self-contained, I will present here the basics of calculus on time scales adapted to our case, for a full treatment the reader can consult the references cited in the Introduction.

To conform to the notation of the theory of time scales, we will denote by $\mathbb{T}$ (a finite version of $\mathbb{R}$) the set formed by all the machine numbers, so $\mathbb{T} = \mathbb{F}(b, p, e_{\min}, e_{\max})$. If $n_{\min} = \min(\mathbb{T})$, and $n_{\max} = \max(\mathbb{T})$, we will write $\mathbb{T}^* = \mathbb{T} - \{n_{\min}, n_{\max}\}$. The set $\mathbb{T}$ inherits the notion of distance in $\mathbb{R}$ originated from the absolute value. Due to the discrete nature of our number system, with that distance all functions $\mathbb{T} \rightarrow \mathbb{C}$ are automatically continuous.

To construct a calculus on $\mathbb{T}$, we start from the classical notion of derivative, but taking into account the finiteness of elements in $\mathbb{T}$. Thus, for any $t \in \mathbb{T}^*$ and function $f : \mathbb{T} \rightarrow \mathbb{C}$, we define the forward derivative

$$\Delta f(t) := \frac{f(t + \text{eps}_t) - f(t)}{\text{eps}_t}. \quad (4)$$

Notice that no indeterminacies can appear in this expression. An equivalent form is given by the condition that $f'(t)$ be the only number such that

$$\frac{f(t + \text{eps}_t) - f(t) - \Delta f(t) \cdot \text{eps}_t}{\text{eps}_t} = 0,$$

which in turn (because 0 is any number smaller than $\text{eps}_0$) is equivalent to

$$|f(t + \text{eps}_t) - f(t) - \Delta f(t) \cdot \text{eps}_t| \leq \text{eps}_0 \cdot \text{eps}_t.$$

12 Technically, what we do have is a time scale with a left scattered maximum and a right scattered minimum, where every non-maximal element is right scattered, and the forward jump operator is given by $\sigma(t) = t + \text{eps}_t$. 
Example 2. If \( f(t) = t^2 \), then for any \( t \in \mathbb{T} \) we have
$$
\frac{(t + \epsilon s t)^2 - t^2}{\epsilon s t} = t^2 + 2t\epsilon s t + \epsilon s t^2 - t^2
$$

so
$$
\Delta f(t) = 2t + \epsilon s t.
$$

This illustrates the general result that the difference between the (forward) discrete derivative \( \delta f(t) \) and the usual derivative is of order \( O(\epsilon s t) \).

Remark 3. It can be easily proved that any function \( f : \mathbb{T} \to \mathbb{C} \) admits a derivative in the sense of (4), at any point of \( \mathbb{T} \) (see Theorem 1.2.2 in [18]).

The above definition is based on a forward difference scheme. Alternatively, we could use a backward difference scheme, as in the backward derivative
$$
\nabla f(t) := \frac{f(t - \epsilon s t) - f(t)}{\epsilon s t},
$$

and its equivalent form
$$
|f(t - \epsilon s t) - f(t) - \nabla f(t) \cdot \epsilon s t| \leq \epsilon s_0 \cdot \epsilon s t,
$$
or even a central difference scheme as in the centered derivative
$$
\delta f(t) := \frac{f(t + \epsilon s t) - f(t - \epsilon s t)}{2\epsilon s t},
$$
equivalently
$$
\frac{1}{2} f(t + \epsilon s t) - \frac{1}{2} f(t - \epsilon s t) - \delta f(t) \cdot \epsilon s t \leq \epsilon s_0 \cdot \epsilon s t.
$$

In particular, the centered derivative leads to the usual expression for the product rule. First, write
$$
\frac{(fg)(x + h) - (fg)(x - h)}{2h} = \frac{f(x + h)g(x + h) - f(x - h)g(x - h)}{2h}
$$

$$
= \frac{f(x + h) + f(x - h)}{2h} \frac{g(x + h) - g(x - h)}{2h}
$$

$$
= \text{av}_h(f(x)) \frac{g(x + h) - g(x - h)}{2h} + \frac{f(x + h) - f(x - h)}{2h} \text{av}_h(g(x)),
$$

where the \( h \)-average of a function \( f \) at a point \( x \) is defined by
$$
\text{av}_h(f(x)) := \frac{f(x + h) - f(x - h)}{2}.
$$

It can be checked that
$$
\text{av}_h(f(x)) = f(x) + O(h^2).
$$

Hence, upon evaluation\(^{13}\) at \( h = \epsilon s_0 \) in the above expression (taking into account that \( \epsilon s_0^2 < \epsilon s_0 \) is zero):
$$
\delta(fg)(x) = \frac{(fg)(x + h) - (fg)(x - h)}{2h} \bigg|_{h=\epsilon s_0} = f(x) \cdot \delta g(x) + \delta f(x) \cdot g(x).
$$

\(^{13}\)Notice that \( h = \epsilon s_0 \) is a special value in this regard. In general, for any other \( t \neq 0 \), we can not discard terms \( O(\epsilon s t^2) \).
Example 3. The centered derivative coincides with the exact derivative on polynomials. To see it, write $x^k = x \cdot x^{k-1}$ for any integer $k > 1$ and apply the product rule:

$$\delta(x^k) = \delta(x) \cdot x^{k-1} + x \cdot \delta(x^{k-1}).$$

A trivial calculation yields $\delta(x) = 1$, so iterating the procedure we get

$$\delta(x^k) = x^{k-1} + x \cdot \delta(x^{k-1}) = x^{k-1} + x \cdot (\delta(x^{k-2}) + x \cdot \delta(x^{k-3})).$$

An easy recursion on $k$ steps, leads us to the final expression

$$\delta(x^k) = k \cdot x^{k-1}.$$

\[ \triangle \]

Example 4. Consider the case of the logarithm function. In a usual-calculus based course, one proves that given $f(x) = a^x$ with $a > 0$, then

$$f'(x) = \ln(a) \cdot a^x = \lim_{\lambda \to 0} \frac{a^{x+\lambda} - a^x}{\lambda},$$

that is,

$$\ln(a) = \left. \frac{d}{dx} \right|_{x=0} (a^x).$$

We take this expression as our starting point and define (using a central difference scheme)

$$\ln(a) := \left. \frac{a^h - a^{-h}}{2h} \right|_{h=\epsilon h_0} = \delta(a^x)|_{x=0}. \quad (5)$$

The particular value at $a = 1$ is readily computed:

$$\ln(1) = \left. \frac{1^h - 1^{-h}}{2h} \right|_{h=\epsilon h_0} = \left. \frac{1 - 1}{2h} \right|_{h=\epsilon h_0} = 0.$$

Also, the basic property of logarithms follows from the definition and the product rule:

$$\ln(ab) = \delta((ab)^x)|_{x=0} = \delta(a^x \cdot b^x)|_{x=0} = a^x|_{x=0} \delta(b^x)|_{x=0} + \delta(a^x)|_{x=0} b^x|_{x=0} = \ln(a) + \ln(b).$$

The definition (5) is quite impractical because it involves arithmetic operations with close numbers, something prone to numerical instabilities. In the next Section, we will see how to compute concrete values of the ln function, learning on the way some interesting results. \[ \triangle \]

Newton’s algorithm for computing the roots of an equation like $f(x) = 0$ follows now quite naturally. Thus, the iterations given by

$$x_{n+1} = x_n - \frac{f(x_n)}{\Delta f(x_n)},$$

with the usual caveats about the right choice of the initial value of $x_0$ and the non vanishing of the derivative will give an excellent approximation to the roots. In particular, if we want to determine the $k$th root of $a \in T$ we can consider $f(x) = x^k - a.$
It is clear that even for those $s \in \mathbb{T}$ such that $\varepsilon s_t < s$, but sufficiently small, we get a good approximation of the derivative $f'(t)$ by putting

$$f'(t) \simeq \frac{f(t + s) - f(t)}{s}. \quad (6)$$

This leads to an elementary version of the Mean Value Theorem, namely,

$$f(t + s) \simeq f(t) + \Delta f(t) \cdot s,$$

for $s$ sufficiently small. More accurate versions can be found in [1, 15, 18] and [13].

Just with the very basic results we have presented, many interesting things can be studied, certainly many of the 17th century mathematics, which include such topics as the determination of planetary orbits\(^{14}\). The proofs are identical in many cases to those of a usual-calculus based course, if simpler, as the technicalities involving pathologies such as continuous but non-differentiable functions do not appear in this context. Among the many choices available, in the next section I present a straightforward application to the effective computation of a transcendental function, the natural logarithm\(^{15}\).

### 5. How does a calculator compute ln(114.23)?

Continuing with the example of the ln function, we ask ourselves how can we compute its numerical values for generic cases, such as ln(114.23). This is a question recurrently asked in internet forums\(^{16}\). If you have a pocket calculator at hand, you can check that it returns the value 4.73821 very quickly, and one is tempted to think that the really fast algorithm behind the calculations must be somehow related to the magical powers of calculus. Indeed it is, but not in the manner many people expect: no limits are involved here, and everything takes place in the realm of finite calculus!

The algorithms implemented in many calculators are actually some kind of variation of the original one devised by H. Briggs in 1624 [4], which we now describe in an adapted version\(^{17}\). It is based on a linear approximation of the ln function, which in turn depends on the following result (a rational form of the binomial theorem).

**Lemma 1.** Let $\alpha = p/q$ and $x$ be numbers in $\mathbb{T}$ such that $|x| < 1$. Then, there exists an $N \in \mathbb{N} \cap \mathbb{T}$ such that

$$(1 + x)\alpha = 1 + \alpha x + \frac{\alpha(\alpha - 1)}{2} x^2 + \cdots + \sum_{k=0}^{N} \frac{\alpha^k}{k!} x^k, \quad (7)$$

where the $k$–falling power of $\alpha$ is defined as

$$\alpha^k = \alpha(\alpha - 1) \cdots (\alpha - k + 1).$$

**Proof.** As we have $(1 + x)^{p/q} = ((1 + x)^{1/q})^p$, we only need to prove the statement for powers of the form $(1 + x)^{1/q}$ and then apply the usual binomial theorem repeatedly. Thus, assume $\alpha = 1/q$ and write

$$(1 + x)^{1/q} = a_0 + a_1 x + a_2 x^2 + \cdots + a_M x^M,$$

\(^{14}\) If you were to ask a 17th century mathematician where to look in the sky tonight at 21:30 in order to find Venus, sure he would have knew what computations to do. I bet you can not.

\(^{15}\) Similar algorithms, fully describable in the framework of finite calculus, are available for other functions, such as the trigonometric ones. In this case, the most popular is the family of CORDIC algorithms [30].

\(^{16}\) I thank my colleague J. Medina, responsible for the Youtube channel lasmatematicas.es, for suggesting this example.

\(^{17}\) Briggs introduced logarithms in base 10, while we will work with natural logarithms.
for some $M \in \mathbb{N} \cap \mathbb{T}$ which we will determine. We can compute the coefficients in the expansion recursively, by comparing both sides of the equation that results when taking powers of $q$:

$$1 + x = (a_0 + a_1 x + a_2 x^2 + \cdots + a_M x^M)^q$$

$$= (a_0 + a_1 x + a_2 x^2 + \cdots + a_M x^M) \cdots (a_0 + a_1 x + a_2 x^2 + \cdots + a_M x^M)$$

$$= a_0^q + q a_0^{q-1} a_1 x + \left( \frac{q}{2} a_0^{q-2} a_1^2 + \cdots \right) x^2 + \cdots$$

We obtain

$$\begin{cases} a_0 = 1 \\ qa_0^{q-1} a_1 = 1 \\ qa_0^{q-1} a_2 + \left( \frac{q}{2} a_0^{q-2} a_1^2 \right) x^2 = 0 \\ \vdots \end{cases}$$

whose solution is

$$\begin{cases} a_0 = 1 \\ a_1 = 1/q \\ a_2 = -\frac{1}{q} (\frac{q}{2} \frac{1}{q^2} = -\frac{1}{2!} \frac{q(q-1)}{q^2} = \frac{1}{2!} \frac{1}{q} \left( \frac{1}{q} - 1 \right) = \frac{q^2}{2q} \\ \vdots \end{cases}$$

Notice that the process is finite: it eventually stops. To see this, we only need to bound the terms of the sum in the right hand side of (7), which can be done as follows. As $\alpha = 1/q$ with $q \in \mathbb{Z} \cap \mathbb{T}$ satisfies

$$|\alpha| < 1, \ |\alpha - 1| < |\alpha| + 1 < 2, \ldots, |\alpha - k + 1| < k,$$

we get

$$\left| \frac{\alpha(\alpha - 1) \cdots (\alpha - k + 1)}{k!} \right| < \frac{1 \cdot 2 \cdots k}{1 \cdot 2 \cdots k} = 1,$$

hence

$$\left| \frac{\alpha^k}{k!} x^k \right| < |x|^k.$$ 

Now, as $|x| < 1$, for great enough $k$ we have $|x|^k < \epsilon s_0$, so it is indistinguishable from 0 and the sum only contains a finite number of terms.

Here is the promised approximation.

**Proposition 1.** The $\ln$ function satisfies

$$\Delta \ln(1) = 1.$$ 

As a consequence, we have the linear approximation (valid for small $|x| < 1$ at order $O(\epsilon)$)

$$\ln(1 + x) \simeq x.$$  

(8)
Proof. Using the definition of derivative and the preceding lemma, we can compute
\[
\Delta \ln(1) = \frac{1}{h} (\ln(1 + h) - \ln(1)) \bigg|_{h=\text{eps}} = \frac{1}{k} \left( (1 + h)^k - 1 \right) \bigg|_{k=\text{eps}_0} - 0 \bigg|_{h=\text{eps}} = \frac{1}{kh} \left( (1 + h)^k - 1 \right) \bigg|_{k=\text{eps}_0} - 0 \bigg|_{h=\text{eps}} = \left( 1 + \frac{k^2}{2!} h + \cdots \right) \bigg|_{k=\text{eps}_0} \bigg|_{h=\text{eps}} = \left( 1 + \frac{k}{2} h + \cdots \right) \bigg|_{k=\text{eps}_0} \bigg|_{h=\text{eps}} = 1.
\]
Where, in the last step, we have used that \( \text{eps}_0 \cdot \text{eps} \ll \text{eps}^2 \ll \text{eps}_0 \) (and a similar result for higher-order products). Finally, the replacement
\[
\Delta \ln(1) = 1 = \frac{1}{h} (\ln(1 + h) - \ln(1)) \bigg|_{h=\text{eps}} \approx \frac{1}{x} (\ln(1 + x) - \ln(1)) = \ln(1 + x)
\]
gives the statement. \(\square\)

We come now to the idea of Briggs’ algorithm. Suppose we want to calculate \( \ln(a) \), with \( a > 0 \) arbitrary in \( \mathbb{T} \). From the property \( \ln(a \cdot a) = \ln(a) + \ln(a) \), it follows that, for any \( k \in \mathbb{N} \cap \mathbb{T} \),
\[
\ln(a) = 2^k \ln \left( a^{\frac{1}{2^k}} \right).
\]
(9)

When \( k \) is sufficiently large \( a^{1/2^k} \) is very close to 1 and can be written as \( a^{1/2^k} = 1 + x \), where \( |x| \ll 1 \). From the preceding Proposition, we get
\[
\ln(a) \approx 2^k \left( a^{\frac{1}{2^k}} - 1 \right).
\]
(10)

Briggs, in his treatise [1], chose \( k = 54^{18} \). The problem with (10) is that, for large \( k \in \mathbb{N} \cap \mathbb{T} \), \( a^{\frac{1}{2^k}} - 1 \) is the difference between close numbers, and it is well known that this operation leads to numerical instability. This is exemplified in the following Maxima session (identical results are obtained in Mathematica 12.0):

\[
\begin{align*}
\text{(i1)} & \quad \text{approx}(k):=2^k((114.23)^{(1/2^k)}-1) \\
\text{(i2)} & \quad \text{approx}(6) \\
\text{(i2)} & \quad 4.918019599915098 \\
\text{(i3)} & \quad \text{approx}(10) \\
\text{(i3)} & \quad 4.749193129303649 \\
\text{(i4)} & \quad \text{approx}(54) \\
\text{(i4)} & \quad 4.0 \\
\text{(i5)} & \quad \text{approx}(100) \\
\text{(i5)} & \quad 0.0
\end{align*}
\]

The modification of Briggs’ algorithm that many calculators use [9] consists in reducing the calculation of \( \ln(a) \) to the sum of known logarithms of certain numbers. To this end, first write the number \( a \in \mathbb{T} \) in scientific notation as
\[
a = m \cdot 10^K
\]

18 Actually, once aware of property (9) he started to compute values of the form \( \log \left( 10^{\frac{1}{2^k}} \right) \) for \( k = 1, 2, 3, \ldots \), and very quickly realized that the linear approximation (8) was valid.
with mantissa $0 \leq m < 1$ and exponent $K \in [e_{\text{min}}+1, e_{\text{max}}+1]$ (notice the shift in the exponents). Then

$$\ln(a) = \ln(m) + K \ln(10),$$

(11)

so we need to compute separately $\ln(10)$. It remains to calculate $\ln(m)$.

Suppose we can find $a_0, a_1, \ldots, a_n \in T$, a finite set of numbers whose logarithms $\ln(a_i)$ ($i \in \{0, 1, \ldots, n\}$) are known, and integers $k_i$ such that

$$r = a_0^{k_0} a_1^{k_1} \cdots a_n^{k_n} \cdot m$$

(12)

is very close to 1. In that case, from the properties of $\ln$ we get

$$\ln(r) = \sum_{i=0}^{n} k_i \ln(a_i) + \ln(m).$$

Being $r \simeq 1$, we can approximate $\ln(r) = \ln(1 + (r - 1)) \simeq r - 1$, obtaining

$$\ln(m) = r - 1 - \sum_{i=0}^{n} k_i \ln(a_i).$$

(13)

As $0 \leq m < 1$, the numbers $a_i$ must be greater than 1. To make computations easy, a good choice consists in taking, for $i \in \{0, 1, \ldots, n\}$,

$$a_i = 1 + 10^{-i},$$

so that the first four values are $a_0 = 2, a_1 = 1.1, a_2 = 1.01, a_3 = 1.001$. The values of the logarithms of these constants, along with $\ln(10)$, are computed separately and stored in a hash table like this:

| Value | $\ln$   |
|-------|---------|
| 10    | 2.30258509299 |
| 2     | 0.69314718056  |
| 1.1   | 0.0953101798   |
| 1.01  | 0.009995033085 |
| 1.001 | 0.00099950033  |

Finally, to find the constants $K_i$ a recursive construction is used. Define $P_0 = ma_0^{k_0}$ and, for $j \geq 1$,

$$P_j = P_{j-1} a_j^{k_j} = mP_{j-1}(1 + 10^{-j})^{k_j}. $$

Then, choose the value of $k_j$ as the largest integer such that $mP_j \leq 1$. Notice that this choice means that the successive approximations to $r$ in (12) are all less than 1.

**Remark 4.** During the computation of $k_j$ we must evaluate several intermediate products of the form $mP_{j-1}(1 + 10^{-j})^0$, $mP_{j-1}(1 + 10^{-j})^1$, $mP_{j-1}(1 + 10^{-j})^2$, etc, and check whenever any of them is greater than 1. Each one of these is obtained from the preceding one simply by shifting the digits to the right the same number of positions indicated by its exponent, and adding the number so obtained to the original value. This is important to handle the exceptional case in the algorithm in which $mP_{j-1}a_j^{k_j} = mP_{j-1}(1 + 10^{-j})^{k_j}$ exceeds 1 for any value $k_j \geq 1$. Then, we skip entirely the factor $a_j$, increase in one unit the digit in the position $j + 1$ of $mP_{j-1}$ and continue the algorithm with $a_{j+1}$. 

14
Example 5. Let us compute \( \ln(114.23) \), as promised. First, write \( a = 114.23 = 0.11423 \cdot 10^3 \), so \( m = 0.11423 \) and \( \ln(114.23) = \ln(0.11423) + 3\ln(10) \). Let us focus on \( \ln(0.11423) \). We start the algorithm with \( a_0 = 2 \) and consider \( mP_0 = ma_0^{k_0} = 0.11423 \cdot 2^{k_0} \), where \( k_0 \) must be chosen as the largest integer such that \( 0.11423 \cdot 2^{k_0} < 1 \). We have the following values:

| \( k_0 \) | \( mP_0 \) |
|-------|---------|
| 0     | 0.11423 |
| 1     | 0.22864 |
| 2     | 0.45692 |
| 3     | 0.91384 |
| 4     | 1.82768 |

so we must take \( k_0 = 3 \) and \( mP_0 = 0.11423 \cdot 2^3 = 0.91384 \). Continue with \( mP_1 = mP_0(1 + 10^{-1})^{k_1} = 0.91384 \cdot (1.1)^{k_1} \). Here, \( k_1 \) must be chosen as the largest integer such that \( mP_1 = 0.91384 \cdot (1.1)^{k_1} < 1 \), but this we are in the exceptional case, since

| \( k_1 \) | \( mP_1 \) |
|-------|---------|
| 0     | 0.91384 |
| 1     | 1.005224|

Thus, we change the value \( mP_0 = 0.91384 \) to 0.92384 and proceed to consider \( mP_2 = 0.92384 \cdot (1.01)^{k_2} \). This time, we compute

| \( k_2 \) | \( mP_2 \) |
|-------|---------|
| 0     | 0.92384 |
| 1     | 0.9330784|
| \vdots| \vdots |
| 8     | 0.9904815236905402 |
| 9     | 1.000386338927446 |

so \( k_2 = 8 \). The algorithm continues in this way until we reach the desired degree of approximation. In this example, we have arrived at

\[
r = mP_2 = 0.11423 \cdot 2^3 \cdot (1.01)^8 = 0.9895577718711648
\]

that is, \( r - 1 = -0.0104422812883522 \). Going back to (11) and (13), we find

\[
\ln(114.23) \simeq -0.01044222812883522 - (3 \ln(2) + 8 \ln(1.01)) + 3 \ln(10) = 4.7382688862348122,
\]
which is to be compared with the exact value to 15 decimal places:

\[ \ln(114.23) = 4.738213959745856. \]

The power of the algorithm is quite evident. In fact, only a few values \( a_i \) need to be computed separately and stored (the first desktop calculators only used half a dozen values), a good example of ingenuity.

6. Conclusions

It should be possible to present the mathematics needed by a general scientist in a setting that avoids the difficulties related to infinities. The usual notions of calculus can be expressed using a particular time scale modeled on floating point arithmetic, and the properties of transcendental functions such as logarithms or trigonometric ones can be well understood, leading to computational algorithms to determine their values. Many possibilities open up if these ideas are expanded. For instance, we have not touched upon the topic of integration, which in the finite context is just summation; there is a fundamental theorem of calculus available, and using it we can prove Stirling’s approximation. In a course of mathematics for biologists, this can lead to a discussion of useful probability distributions converging to a normal one, in a course for physicists it can be used to discuss the computation of entropy or to understand Einstein’s model of a solid. The purpose of this note is just to point out that these examples exist, can be taught on solid grounds, and that it is worthwhile to give them a try.

Acknowledgments

This paper is an expanded version of some of the ideas I presented at the ISIMMED 2019 (Universitas Negeri Yogyakarta, Indonesia). The stay there was a wonderful experience thanks to Dr. Kismiantini and Dr. Karyati, who took care of every detail and did every effort to create a nice environment. Also, I thank many discussions with my colleagues and friends Dr. F. Periago and J. Medina, that helped to sharpen many aspects of the paper. Prof. Dr. Zeilberger deserves special thanks for providing invaluable insight for all his readers.

References

[1] R. Agarwal, M. Bohner, D. O'Regan and A. Peterson: *Dynamic equations on time scales: a survey*. Journal of Computational and Applied Mathematics 141 (2002) 1–26.
[2] V. I. Arnol’d: *Huygens and Barrow, Newton and Hooke*. Birkaüser, Basel, 1990.
[3] E. Bishop and D. S. Bridges: *Constructive Analysis*. Springer-Verlag, 1985.
[4] H. Briggs: *Arithmetica Logarithmica*. London, 1624. A modern reconstruction of Briggs’ tables (with useful comments) is freely available at [https://locomat.loria.fr/briggs1624/briggs1624doc.pdf](https://locomat.loria.fr/briggs1624/briggs1624doc.pdf).
[5] R. F. Cahalan, H. Leidecker and G. D. Cahalan: *Chaotic Rhythms of a Dripping Faucet*. Computers in Physics 4 (1990) 368–382.
[6] A. L. Cauchy: *Cours d’Analyse*. L’Imprimerie Royale, Debure frères, Libraires du Roi et de la Bibliothèque du Roi, Paris, 1821. Freely available at [http://sites.mathdoc.fr/cgi-bin/oieitn?id=OE_CAUCHY_2_3_PS_0](http://sites.mathdoc.fr/cgi-bin/oieitn?id=OE_CAUCHY_2_3_PS_0).
[7] L. Debnath: *A short biography of Joseph Fourier and historical development of Fourier series and Fourier transforms*. International Journal of Mathematical Education in Science and Technology 43(5) (2012) 589–612.
[8] F. de Gandt: *Force and Geometry in Newton's Principia*. Princeton University Press, 1995.
[9] W. E. Egbert: *Personal Calculator Algorithms IV: Logarithmic Functions*. Hewlett Packard Journal, April 1978, 29–32.
[10] J. Flegg, D. G. Mallet and M. Lupton: *Students’ perceptions of the relevance of mathematics in engineering*. International Journal of Mathematical Education in Science and Technology 43(6) (2012) 717–732.
[11] J. B. J. Fourier: *Théorie Analytique de la Chaleur*. Firmin Didot, Père et Fils, Paris, 1822. Freely available at [https://gallica.bnf.fr/ark:/12148/bpt6k1045508v.texteImage](https://gallica.bnf.fr/ark:/12148/bpt6k1045508v.texteImage).
[12] H. Georgi: *Lie Algebras in Particle Physics*. Addison-Wesley, Redwood City (California), 1982.
[13] S. Gulsan-Topal: *Rolle’s and generalized mean value theorems on time scales*. Journal of Difference Equations and Applications 8(4) (2002) 333–343.
[14] R. Hamming: *Mathematics on a Distant Planet*. The American Mathematical Monthly 105(7) (1998) 640–650.
[15] S. Hilger: *Analysis on measure chains, a unified approach to continuous and discrete calculus*. Results in Mathematics 18 (1990) 18–56.
[16] S. Hossenfelder: *Lost in Math: How beauty leads Physics astray*. Basic Books, New York, 2018.
[17] S.G.B. Johnson and S. Steinerberger: *Intuitions about mathematical beauty: A case study in the aesthetic experience of ideas*. Cognition 189 (2019) 242–259.
[18] B. Kaymakçalan, V. Lakshmikantham and S. Sivasundaram: *Dynamical Systems on Measure Chains*. Kluwer Academic Publishers, Boston, 1996.
[19] A. Kornai: *Explicit Finitism*. International Journal of Theoretical Physics 42(2) (2003) 301–307.
[20] P. D. Lax: *The flowering of Applied Mathematics in America*. SIAM Review 31(4) (1989) 533–541.
[21] J. Lörinczi, F. Hiroshima and V. Betz: *Feynman-Kac-Type formulae and Gibbs measures on path space: with applications to rigorous Quantum Field Theory*. De Gruyter Studies in Mathematics 34, Berlin, 2011.
[22] G. Markowsky: *A retelling of Newton’s work on Kepler’s Laws*. Expositiones Mathematicae 293 (2011) 253–282.
[23] J. M. Muller: *Handbook of Floating-Point Arithmetic*. Birkhäuser Basel, 2010.
[24] I. Newton: *The Principia: Mathematical Principles of Natural Philosophy*. Translation and annotations by I. Bernard Cohen and A. Whitman. University of California Press, Berkeley, 1999.
[25] S. S. Petrova: *Heaviside and the development of the symbolic calculus*. Archive for History of Exact Sciences, 37(1) (1987) 1–23.
[26] S. Pohjolainen et al.: *Modern Mathematics Education for Engineering Curricula in Europe*. Birkhäuser Basel, 2018.
[27] J. P. Provost and C. Bracco: *A simple derivation of Kepler’s laws without solving differential equations*. European Journal of Physics 30(3) (2009) 581–586.
[28] V. Y. Sazonov: *On feasible numbers*. In: D. Leivant (ed), Logic and Computational Complexity (LCC 1994). Lecture Notes in Computer Science 960. Springer, Berlin, 1995.
[29] R. Staden: *A strategy of DNA sequencing employing computer programs*. Nucleic Acids Research, 6(70) (1979) 2601–2610.
[30] J. S. Walther: *A unified algorithm for elementary functions*. Proceedings of the Spring Joint Computer Conference. Atlantic City, New Jersey, USA. Hewlett-Packard Company 38 (1971) 379–385.
[31] F. Ye: *Strict Finitism and the Logic of Mathematical Applications*. Springer Verlag, Dordrecht, 2011.
[32] Doron Zeilberger’s web page at http://sites.math.rutgers.edu/~zeilberg/.
[33] D. Zeilberger: “*Real* analysis is a degenerate case of discrete analysis.” In Proceedings of the Sixth International Conference on Difference Equations, Augsburg (Germany) 2001. CRC Press, Boca Raton, 2004.