REAL NUMBERS AS INFINITE DECIMALS — THEORY AND COMPUTATION

NICOLAS FARDIN AND LIANGPAN LI

ABSTRACT. In the 16th century, Simon Stevin initiated a modern approach to decimal representation of measuring numbers, marking a transition from the discrete arithmetic practised by the Greeks to the arithmetic of the continuum taken for granted today. However, how to perform arithmetic directly on infinite decimals remains a long-standing problem, which has seen the popular degeometrisation of real numbers since the first constructions were published in around 1872. Our article is devoted to solving this historical problem. An issue that Hardy called "a fatal defect" is also settled.

1. Introduction

In the 16th century, Simon Stevin ([18, 23, 24]) initiated a modern approach to decimal representation of measuring numbers, marking a transition from the discrete arithmetic practised by the Greeks to the arithmetic of the continuum taken for granted today. However, how to perform arithmetic directly on infinite decimals remains a long-standing problem ([2, p. 97], [3, p. 8], [4, p. 11], [8, p. 10], [13, p. 123], [19, p. 16], [25, p. 80], [28, p. 400], [34, pp. 105–106], [36, p. 739], [37]), which has seen the popular degeometrisation ([14, 23, 24, 27]) of real numbers since the first constructions were published independently by Méray, Heine, Cantor and Dedekind in around 1872. Since then, plenty of attempts have been made to construct the same algebraic structure from various perspectives, such as the least upper bound property, the Archimedean property, equivalence classes, axiomatic approaches, the additive group of integers, continued fractions, harmonic or alternating series, and so on ([5, 36]).

It is well known that any element of the real number system can be identified with an infinite decimal, so why not define arithmetical operations directly on infinite decimals? There is a long list of mathematicians including Weierstrass and Stolz who prioritise the decimal system over other constructions since we all learned at school how to perform arithmetic on terminating decimals. But many decimal approaches lack details, and most of them are essentially not so different from the earliest theories (see [19] for a literature review). Decimal constructions of the real number system are thus rarely seen in modern Mathematical Analysis textbooks.

Our article is devoted to solving this historical problem. Only basic knowledge about elementary arithmetic on terminating decimals, or equivalently on integers, is required. We believe our approach can be used for teaching purposes in universities, and even in schools in some respects.

2010 Mathematics Subject Classification. 97F50.

Key words and phrases. real number system, trigonometric formulas, rectifiable curve, computable number.
The main idea of this paper is as simple as performing arithmetic on integers, but in a slightly different way. We are usually told to add and multiply numbers from right to left. Why not do so from left to right?

Let us consider a slightly strange sum $12 \cdot \cdot + 45 \cdot \cdot$, where each summand is of four digits and the black dots are not specified. Since the sum of digits in the hundreds place is 7 not 9, we get

$$12 \cdot \cdot + 45 \cdot \cdot = 5\frac{7}{8} \cdot \cdot$$

whose thousands place value 5 is independent of the values of black dots, and hundreds place value can only be 7 or 8. Here $\frac{7}{8}$ represents the two possible choices, not the fraction that is equal to 0.875. In exactly the same way, we have

$$0.12 \cdot \cdot + 0.45 \cdot \cdot = 0.5\frac{7}{8} \cdot \cdot.$$ 

Since none of the black dots in the above two examples are deterministic, we can prolong them freely, even to infinity in the second example. Such an observation led Hua ([17]) to define addition on infinite decimals in 1962. Richman ([28]) got the same idea in 1999. To summarise, the principle is to do addition locally from right to left but globally from left to right.

To the best of the authors’ knowledge, we are not aware of any works which define multiplication in a similar way, although Wu ([37]) believes it is doable.

Ahead of stating Hua’s definition and our multiplication proposal, we fix some notations. Our ambient space is $([8, 12, 20])$

$$\mathbb{R} = \{a_0.a_1a_2a_3\cdots \in \mathbb{Z} \times \mathbb{Z}_{10}^\infty : a_k < 9 \text{ for infinitely many } k\},$$

where $\mathbb{Z}_{10}$ denotes the set $\{0, 1, 2, \ldots, 9\}$. Note that we exclude infinite decimals ending in a string of 9s once and for all. Many readers may be more familiar with the classical decimal system ([34])

$$\pm z = \pm a_0.a_1a_2a_3\cdots (z \in \mathbb{R}, z \geq 0),$$

but there is no essential difference between the options as long as we primarily focus on arithmetic on non-negative real numbers, then make a suitable extension. One main reason for choosing $\mathbb{R}$ is because generally given a real number $x$ (or a Dedekind cut, a Cauchy sequence, a map on $\mathbb{Z}$, and so on), we don’t need to know whether it is non-negative or negative in advance, but can find a unique $a_0 \in \mathbb{Z}$ so that $a_0 \leq x < a_0 + 1$, a unique $a_1 \in \mathbb{Z}_{10}$ so that $a_0 + \frac{a_1}{10} \leq x < a_0 + \frac{a_1+1}{10}$, and continue in this way to get $x = a_0.a_1a_2a_3\cdots$. A second reason is that the link between the model $\mathbb{R}$ and the earliest theories of real numbers can be well explained (see Section 4).

An element $x = a_0.a_1a_2a_3\cdots$ is said to be terminating if there exists a non-negative integer $m$ such that $a_k = 0$ for $k > m$. In this case, write $x = a_0.a_1a_2\cdots a_m$ for simplicity. As usual, $10^{-m}$ is the same as 0.00\ldots01 whose last digit 1 is at the $m$-th decimal place. For any element $x = a_0.a_1a_2a_3\cdots$ and any non-negative integer $k$, denote $\theta_k(x) = a_k$, the $k$-th digit of $x$, and $x_k = a_0.a_1a_2\cdots a_k$, the truncation of $x$ up to the $k$-th digit. Defining addition on terminating decimals is rather standard. For example,

$$(-8).765 + 5.678 = (-8) + 0.765 + 5.678 = (-8) + 6.443 = (-2).443.$$ 

\footnote{To compare, the addition on $p$-adic numbers $\cdots b_2b_1b_0.b_{-1}b_{-2}\cdots b_{-m}$ and $\cdots c_2c_1c_0.c_{-1}c_{-2}\cdots c_{-n}$ works both locally and globally from right to left.}

\[ (-8).765 + 5.678 = (-8) + 0.765 + 5.678 = (-8) + 6.443 = (-2).443. \]
Definition 1.1 (addition, Hua [17]). Let \( x, y \) be elements of \( \mathbb{R} \).
Case 1: Suppose there exists a non-negative integer \( m \) such that \( \theta_{k}(x_{k} + y_{k}) = 9 \) for \( k > m \). Then define
\[
x + y = (x_{m} + y_{m})_{m} + 10^{-m}.
\]
Case 2: Suppose there exists a sequence of positive integers \( k_{1} < k_{2} < k_{3} < \cdots \) such that \( \theta_{k_i}(x_{k_i} + y_{k_i}) \neq 9 \) for \( i \in \mathbb{N} \). Then \( x + y \) is defined by setting
\[
(x + y)_{k_{i} - 1} = (x_{k_{i}} + y_{k_{i}})_{k_{i} - 1} \quad (i \in \mathbb{N}).
\]

Note that \( \theta_{k}(x_{k} + y_{k}) = 9 \) if and only if \( \theta_{k}(x) + \theta_{k}(y) = 9 \) and \( (x_{m} + y_{m})_{m} = x_{m} + y_{m} \), and both succinct substitutes were indeed used in the original definition in [17]. But to look for a reasonable multiplication rule, one should not study the analogue \( \theta_{k}(x) \times \theta_{k}(y) \), and the reason will be easily seen later on. An element \( a_{0}a_{1}a_{2}a_{3} \cdots \) is said to be non-negative or negative if \( a_{0} \geq 0 \) or \( a_{0} < 0 \). Everyone knows about performing multiplication on non-negative terminating decimals. Obviously, it suffices to consider the special case \( x + y \leq 1 \) as the general case is linked by \( xy = 10^{2s} \times \left( \frac{x}{10^{s}} \times \frac{y}{10^{s}} \right) \) via a large enough non-negative integer \( s \). Our multiplication proposal is as follows.

Definition 1.2. Let \( x, y \) be non-negative elements of \( \mathbb{R} \) such that \( x + y \leq 1 \).
Case 1: Suppose there exists a non-negative integer \( m \) such that \( \theta_{k}(x_{k}y_{k}) = 9 \) for \( k > m \). Then define
\[
xy = (x_{m}y_{m})_{m} + 10^{-m}.
\]
Case 2: Suppose there exists a sequence of positive integers \( k_{1} < k_{2} < k_{3} < \cdots \) such that \( \theta_{k_i}(x_{k_i}y_{k_i}) \neq 9 \) for \( i \in \mathbb{N} \). Then \( xy \) is defined by setting
\[
(xy)_{k_{i} - 1} = (x_{k_{i}}y_{k_{i}})_{k_{i} - 1} \quad (i \in \mathbb{N}).
\]

The analogy between both definitions supports the validity of the proposal, and a more convincing explanation is as follows. Let \( x, y \) be non-negative such that \( x + y \leq 1 \), and denote \( x = x_{k} + \epsilon_{k}, y = y_{k} + \delta_{k} \). Naturally we expect
\[
xy = (x_{k} + \epsilon_{k})y = y_{k} + \epsilon_{k}y = x_{k}y_{k} + x_{k}\delta_{k} + \epsilon_{k}y,
\]
which implies that
\[
(1.1) \quad 0 \leq xy - x_{k}y_{k} < 10^{-k}
\]
as \( x + y \leq 1 \) and the maximum between \( \epsilon_{k} \) and \( \delta_{k} \) is strictly less than \( 10^{-k} \). Consequently, if \( \theta_{k}(x_{k}y_{k}) \leq 8 \), then \( (xy)_{k-1} = (x_{k}y_{k})_{k-1} \). So the second case of the definition is feasible. To illustrate the first case, we study a special case of \( m = 2 \) and \( (x_{2}y_{2})_{2} = 0.15 \). It follows from (1.1) that
\[
(1.2) \quad (x_{k}y_{k})_{k} \leq xy > (x_{k}y_{k})_{k} + 2 \cdot 10^{-k}.
\]
Letting \( k = 2 \) in (1.2) gives \( 0.15 \leq xy < 0.17 \). Note also \( (x_{2}y_{2})_{2} \leq (x_{3}y_{3})_{3} \). So
\[
0.15 = (x_{2}y_{2})_{2} \leq (x_{3}y_{3})_{3} \leq xy < 0.17.
\]
Considering the assumption \( \theta_{3}(x_{3}y_{3}) = 9 \), we get \( (x_{3}y_{3})_{3} = 0.159 \) or \( (x_{3}y_{3})_{3} = 0.169 \). The second situation actually could not happen, because if it did then
\[
0.169 = (x_{3}y_{3})_{3} \leq (x_{4}y_{4})_{4} \leq xy < 0.17,
\]
\[
0.1699 = (x_{4}y_{4})_{4} \leq (x_{5}y_{5})_{5} \leq xy < 0.17,
\]
\[
0.16999 = (x_{5}y_{5})_{5} \leq (x_{6}y_{6})_{6} \leq xy < 0.17,
\]
\[
\cdots ,
\]
which implies \(0.1699 \cdots \leq xy < 0.17\). This is absurd, so \((x_3y_3)_3 = 0.159\). Similarly, 
\[
0.159 = (x_3y_3)_3 \leq xy < 0.161, \\
0.1599 = (x_4y_4)_4 \leq xy < 0.1601, \\
0.15999 = (x_5y_5)_5 \leq xy < 0.16001, \\
\cdots, 
\]
which implies \(xy = 0.16 = (x_2y_2)_2 + 10^{-2}\).

To summarise, the principle is to do multiplication first locally via elementary arithmetic then globally from left to right. A general definition will be given in the next section.

Finally, we need to show that the above arithmetical operations, no matter how reasonable they may be, form a field. For whatever reason, many other decimal approaches have stopped here ([19, p. 16]). Actually, the details of proving the field structure of various models have drawn lots of negative feedback in the past. Our method is to first establish
\[
\left| (x + y)_k - x_k - y_k \right| \leq M_1 10^{-k}, \\
\left| (xy)_k - x_k y_k \right| \leq M_2 10^{-k},
\]
then argue by contradiction. Here \(M_1\) and \(M_2\) are positive integers independent of \(k\), and both bounds follow from the corresponding arithmetical definitions in a few lines.

We end this introduction with several citations of impressions of decimal approaches, which can be changed after the real number system is reestablished in Sections 2 and 3.

- “it is not obvious how to perform arithmetical operations” (Brannan [3])
- “any solution involves more and more complications” (Bridger [4])
- “this is not a light task” (Courant [8])
- “simply do not work for infinite decimals” (Gardiner [13])
- “even more tedious to explain multiplication” (Stolz and Gmeiner [33])
- “despite being the most familiar, is actually more complicated” (Tao [34])
- “popular approach by novices but is fraught with technical difficulties” (Weiss [36])

2. Justification of definitions

2.1. Addition. To be precise, by justification of Definition 1.1 we mean that in the first case it is independent of the choices of \(m\), and in the second one \((x_{ki} + y_{ki})_{k_i-1} = (x_n + y_n)_{k_i-1}\) for all \(n > k_i\), and \(x + y\) is an element of \(\mathbb{R}\). As we will do similar work for multiplication, this is left as an exercise. It is easy to check that \(x + 0 = 0 + x = x\) for all \(x \in \mathbb{R}\), so 0 is the unital element of the addition.

2.2. Subtraction. Hua also gave the following definition of subtraction in [17].

**Definition 2.1** (subtraction). Let \(x, y\) be elements of \(\mathbb{R}\).

Case 1: Suppose there exists a non-negative integer \(m\) such that \(\theta_k(x_k) = \theta_k(y_k)\) for \(k > m\). Then define \(x - y = x_m - y_m\).

Case 2: Suppose there exists a sequence of positive integers \(k_1 < k_2 < k_3 \cdots\) such that \(\theta_{k_i}(x_{k_i}) \neq \theta_{k_i}(y_{k_i})\) for \(i \in \mathbb{N}\). Then \(x - y\) is defined by setting
\[
(x - y)_{k_i-1} = (x_{k_i} - y_{k_i})_{k_i-1} \quad (i \in \mathbb{N}).
\]

\(^2\)This is an explanation not a proof, as we exclude infinite decimals ending in a string of 9s. One can add them into the real axis by suitable identification after the field structure is established.
We also leave the justification of the subtraction definition as an exercise. It is easy to check that \( x + (0 - x) = (0 - x) + x = 0 \) for all \( x \in \mathbb{R} \), so \( 0 - x \), denoted simply by \( -x \), is the additive inverse of \( x \). As usual, \( x < y \) means \( x - y \) is negative, or more intuitively, \( x \) appears before \( y \) in the “dictionary” \( \mathbb{R} \) (or \( \mathbb{Z} \times \mathbb{Z}_{10}^N \)). \( x \) is positive means \( 0 < x \), and \( x \leq y \) means \( y - x \) is non-negative, or equivalently, \( x < y \) or \( x = y \).

2.3. Multiplication. The general definition of multiplication is as follows.

**Definition 2.2** (multiplication). Let \( x, y \) be elements of \( \mathbb{R} \).

(1) Suppose \( x, y \) are non-negative. Fix a non-negative integer \( s \) such that \( x + y \leq 10^s \).

Case 1: Suppose there exists a non-negative integer \( m \) such that \( \theta_k(x_{k+s} y_{k+s}) = 9 \) for \( k > m \). Then define \( xy = (x_{m+s} y_{m+s})_m + 10^{-m} \).

Case 2: Suppose there exists a sequence of positive integers \( k_1 < k_2 < k_3 < \cdots \) such that \( \theta_{k_i}(x_{k_i+s} y_{k_i+s}) \neq 9 \) for \( i \in \mathbb{N} \). Then \( xy \) is defined by setting

\[
(xy)_{k_i-1} = (x_{k_i+s} y_{k_i+s})_{k_i-1} \quad (i \in \mathbb{N}).
\]

(2) Suppose \( x, y \) are negative. Then define \( xy = (-x)(-y) \).

(3) Suppose only one of \( x \) and \( y \) is negative. Then define \( xy = -(x(-y)) \).

This section is devoted to justifying Definition 2.2 (1), so its assumptions are followed.

Case 1: First, we claim that for any \( n > m \),

\[
(\dot{x}_{n+s} y_{n+s})_m = (x_{m+s} y_{m+s})_m.
\]

To verify (2.1), it suffices to consider \( n = m + 1 \), and suppose this is the case.

\[
x_{n+s} y_{n+s} = x_{m+s} y_{m+s} + (x_{n+s} - x_{m+s}) y_{m+s} + x_{n+s} (y_{n+s} - y_{m+s})
\]

\[
\leq x_{m+s} y_{m+s} + (x_{n+s} + y_{n+s}) 9 \cdot 10^{n+s} \leq x_{m+s} y_{m+s} + \frac{9}{10^n},
\]

where the last inequality is due to \( x_{n+s} + y_{n+s} \leq 10^s \). Thus

\[
(\dot{x}_{n+s} y_{n+s})_n \leq (x_{m+s} y_{m+s} + \frac{9}{10^n})_n \leq (x_{m+s} y_{m+s})_m + \frac{9}{10^n} + \frac{9}{10^m}.
\]

Considering the assumption \( \theta_n(x_{n+s} y_{n+s}) = 9 \), one gets

\[
(\dot{x}_{n+s} y_{n+s})_n = (x_{n+s} y_{n+s})_m + \frac{9}{10^n}.
\]

Combining (2.2) and (2.3) yields

\[
0 \leq (\dot{x}_{n+s} y_{n+s})_m - (x_{m+s} y_{m+s})_m \leq \frac{9}{10^n} < \frac{1}{10^m},
\]

which proves claim (2.1). Next, we claim that for any \( n > m \),

\[
(\dot{x}_{n+s} y_{n+s})_n + 10^{-n} = (x_{m+s} y_{m+s})_m + 10^{-m}.
\]

To verify (2.4), it suffices to consider \( n = m + 1 \), and suppose this is the case. Recall \( \theta_n(x_{n+s} y_{n+s}) = 9 \), so (2.4) is equivalent to (2.1). Therefore, the definition is independent of the choice of \( m \). On the other hand, it follows from (2.1) that

\[
(\dot{x}_{n+s} y_{n+s})_m + 10^{-m} = (x_{m+s} y_{m+s})_m + 10^{-m},
\]

so the definition is also independent of the choice of \( s \).

\[\text{This is crucial as readers could go to revisit the illustrating example from Definition 1.2.}\]
Case 2: We claim that
\[(x_ny_n)_{k_i-1} = (x_{k_i+s}y_{k_i+s})_{k_i-1}\]
for \(n > k_i + s\). Similar to the verification of the previous case, one gets
\[(x_ny_n) = x_{k_i+s}y_{k_i+s} + \gamma_n\]
for some \(\gamma_n\) with \(0 \leq \gamma_n < \frac{1}{10^{k_i-1}}\). Considering the assumption \(\theta_{k_i}(x_{k_i+s}y_{k_i+s}) \leq 8\), one has
\[x_{k_i+s}y_{k_i+s} = (x_{k_i+s}y_{k_i+s})_{k_i-1} + \epsilon_n\]
for some \(\epsilon_n\) with \(0 \leq \epsilon_n \leq \frac{9}{10^{k_i-1}}\). Combining (2.7) and (2.8) yields
\[0 \leq x_ny_n - (x_{k_i+s}y_{k_i+s})_{k_i-1} < \frac{1}{10^{k_i-1}},\]
which proves claim (2.6). Consequently, \(xy\) is defined as an element of \(\mathbb{Z} \times \mathbb{Z}_{10}^\times\).

Case 2 (continued): In this part we continue to show \(xy \in \mathbb{R}\). Note that this issue is not so important as, even if \(xy \notin \mathbb{R}\), one can identify it with a terminating decimal. We assume \(s = 0\) for simplicity, and leave the general case as an exercise. Suppose \(xy \notin \mathbb{R}\), say for example \(xy = a_0.a_1a_2 \cdots a_{j-1}999 \cdots\), where \(j\) is equal to some \(k_i\). Then, we fix an \(l > j\) so that \(x_n \leq x_j + 10^{-j} - 10^{-l}\) and \(y_n \leq y_j + 10^{-j} - 10^{-l}\) for all \(n \geq l\), and the reason why this is possible will be explained later. Thus considering \(\theta_j(x_jy_j) \leq 8\) and \(x_jy_j\) is of at most \(2j\) decimal places to the right of its integer part, one gets
\[x_ny_n \leq x_jy_j + (x_j + y_j)(10^{-j} - 10^{-l}) + (10^{-j} - 10^{-l})^2 \]
\[\leq ((xy)_{j-1} + 8 \cdot 10^{-j} + (10^{-j} - 10^{-2j})) + (10^{-j} - 10^{-l}) + 10^{-2j} \]
\[= a_0, a_1a_2 \cdots a_{j-1} + 10^{-(j-1)} - 10^{-l}.\]
On the other hand, fixing an \(n > l + 1\) with \(\theta_n(x_ny_n) \leq 8\), one gets
\[x_ny_n \geq (x_ny_n)_{n-1} = (xy)_{n-1} = a_0, a_1a_2 \cdots a_{j-1} + 10^{-(j-1)} - 10^{-(n-1)},\]
which contradicts the above upper bound for \(x_ny_n\). Therefore, we must have \(xy \in \mathbb{R}\). The existence of \(l\) can be seen as follows. One can first pick an \(l_1 > j\) so that \(\theta_{l_1}(x) \leq 8\), then note for any \(n \geq l_1\),
\[x_n \leq x_j + \left(\sum_{i=j+1}^{l_1-1} \frac{9}{10^i}\right) + \frac{8}{10^{l_1}} + \left(\sum_{i=l_1+1}^{n} \frac{9}{10^i}\right) = x_j + 10^{-j} - 10^{-l_1} - 10^{-n} \leq x_j + 10^{-j} - 10^{-l_1}.\]
Similarly, pick an \(l_2 > j\) so that \(\theta_{l_2}(y) \leq 8\), and finally set \(l = \max\{l_1, l_2\}\).

We also use \(x \times y\) to denote \(xy\). It is easy to check that \(x \times 1 = 1 \times x = x\) for all \(x \in \mathbb{R}\), so 1 is the unital element of the multiplication.

2.4. Reciprocal. Stevin’s idea ([15]) works ideally on defining reciprocal operation. Given a positive \(x\), one can find a unique non-negative integer \(a_0\) so that \(x \times a_0 \leq 1 < x \times (a_0 + 1)\), a unique \(a_1 \in \mathbb{Z}_{10}\) so that \(x \times (a_0 + \frac{a_1}{10}) \leq 1 < x \times (a_0 + \frac{a_1 + 1}{10})\), and continue in this way to derive an element \(y = a_0, a_1a_2a_3 \cdots \) of \(\mathbb{Z} \times \mathbb{Z}_{10}^\times\). We leave the verification of the facts \(x \in \mathbb{R}\) and \(xy = yx = 1\) as an exercise. Therefore, the element \(y\), usually denoted by \(x^{-1}\), is the multiplicative inverse (or reciprocal) of \(x\). The reciprocal of a negative element \(z\) is defined to be \(-((z)^{-1})\).
3. **Arithmetical laws**

This section will conclude the proof that \((\mathbb{R}, +, \times)\) is a field. Our strategy agrees with Conway’s suggestion that one should construct the positive reals before constructing any negative ones ([29]).

**Lemma 3.1.** If \(x \neq y\), then there exists an \(l \in \mathbb{N}\) such that \(|x_k - y_k| \geq 10^{-l}\) for \(k > l\).

**Lemma 3.2.** Let \(x, y\) be elements of \(\mathbb{R}\). Then \(|(x + y)_k - x_k - y_k| \leq 4 \cdot 10^{-k}\) for all \(k\).

**Lemma 3.3.** Let \(x, y\) be non-negative elements of \(\mathbb{R}\). Then \(|(xy)_k - x_k y_k| \leq M \cdot 10^{-k}\) for all \(k\), where \(M\) is a positive integer depending only on \(x\) and \(y\).

A proof of Lemma 3.1 is as follows. Assume without loss of generality that 

\[ x = a_0.a_1a_2a_3 \cdots < y = b_0.b_1b_2b_3 \cdots. \]

Take first a non-negative integer \(m\) such that \(x_m < y_m\), then a positive integer \(l > m\) so that \(a_l \leq 8\). For \(k > l\), we have

\[
y_k - x_k \geq y_m - \left( x_m + \left( \frac{1}{10^k} \sum_{i=m+1}^{l} a_i \right) + 10^{-l} \right) \geq y_m - \left( x_m + \frac{1}{10^k} \sum_{i=m+1}^{l} 9 \right)
= (y_m - x_m) - 10^{-m} + 10^{-l} \geq 10^{-l},
\]

which finishes the proof. Tracing the justification of Definition 2.2 can yield a proof of Lemma 3.3, so we omit the details. Lemma 3.2 can be dealt with in a similar way.

**Commutative laws:** \(x + y = y + x, \ xy = yx\).

These laws are self-evident.

**Associative laws:** \((x + y) + z = x + (y + z), \ (xy)z = x(yz)\).

It follows from Lemma 3.2 that

\[
|(x + y + z)_k - x_k - y_k - z_k| = |((x + y) + z)_k - (x + y)_k - z_k + (x + y)_k - x_k - y_k| \leq 8 \cdot 10^{-k}.
\]

Similarly,

\[
|(x + (y + z))_k - x_k - y_k - z_k| \leq 8 \cdot 10^{-k},
\]

so

\[
|((x + y) + z)_k - (x + (y + z))_k| \leq 16 \cdot 10^{-k}.
\]

If \((x + y) + z\) and \(x + (y + z)\) are not the same, then there exists an \(l \in \mathbb{N}\) such that

\[
|((x + y) + z)_k - (x + (y + z))_k| \geq 10^{-l}
\]

for \(k > l\). Consequently, \(10^{-l} \leq 16 \cdot 10^{-k}\) for \(k > l\), which is absurd if we let \(k = l + 2\). This proves the associative law for addition. In much the same way, one can establish the associative law for multiplication between three non-negative elements. The general case is left as an exercise.

**Distributive law:** \(x(y + z) = xy + xz\).

Case 1: Suppose \(x, y, z\) are non-negative. One can provide a proof that is similar to that of the associative law for addition.

Case 2: Suppose \(y\) and \(z\) are of the same sign. Then the law follows from Case 1.

Case 3: Suppose \(y\) and \(z\) are not of the same sign. We can assume without loss of generality that \(y + z, -y, \) and \(z\) are of the same sign. According to Case 2, \(x(y + z) + x(-y) = xz\), which yields \(x(y + z) = xy + xz\).
To conclude, \((\mathbb{R}, +, \times)\) is a field.

4. Classical theories

From now on we will discuss various issues related to real numbers. Many introductory analysis books ([12, 16, 20, 30]) choose Dedekind’s approach, some ([34]) prefer Cantor’s theory, but most avoid a detailed construction. Note that Cantor’s work is essentially the same as those of Méray and Heine ([24]). We should understand that an object could have various characterizations or disguises, and so do real numbers. Below we discuss Dedekind and Cantor’s theories from Stevin’s viewpoint of infinite decimal expansions.

4.1. Dedekind’s theory. A Dedekind cut \((A \mid B)\) is formed of two subsets \(A, B\) of \(\mathbb{Q}\) such that \((\mathbb{Q})\) \(A \cup B = \mathbb{Q}\), \(a < b\) for any \(a \in A\) and \(b \in B\), and \(B\) is of no smallest element\(^4\).

Given a Dedekind cut \((A \mid B)\), there exists a unique integer \(a_0\) such that \(a_0 \in A\), \(a_0 + 1 \in B\). Similarly, there exists a unique integer \(a_1 \in \mathbb{Z}_{10}\) such that \(a_0 + \frac{a_1}{10} \in A\), \(a_0 + \frac{a_1+1}{10} \in B\). Continuing in this way yields an element \(a_0,a_1a_2a_3\cdots\) of \(\mathbb{R}\), which is naturally identified with the cut \((A \mid B)\). One can easily show such an identification is a bijection.

The biggest disadvantage of Dedekind’s approach may be that this language is rarely used in advanced courses and research activities. According to Gamelin’s viewpoint ([14]), “It is not clear that even Dedekind grasped the import of what he had done”.

4.2. Cantor’s theory. A sequence of rational numbers \(\{q_n\}_{n=1}^{\infty}\) is said to be Cauchy if for any \(\varepsilon > 0\), there exists a natural number \(N\) (depending on \(\varepsilon\)) such that \(|q_m - q_n| < \varepsilon\) for all \(m, n > N\). First, show that the given sequence is bounded. By the pigeonhole principle, we then pick an \(a_0 \in \mathbb{Z}\) so that infinitely many elements of the sequence lie in \(\{a_0,a_0+1\}\), an \(a_1 \in \mathbb{Z}_{10}\) so that infinitely many elements of the sequence lie in \(\{a_0 + \frac{a_1}{10}, a_0 + \frac{a_1+1}{10}\}\). Continuing in this way yields an element \(a_0,a_1a_2a_3\cdots\) of \(\mathbb{Z} \times \mathbb{Z}_{10}\). In most cases this procedure outputs a unique identification element of \(\mathbb{R}\). Sometimes\(^5\) it could also provide two “different” elements such as 0.999\cdots and 1.000\cdots so we identify them, but three or more identification elements could never exist all together. To verify this claim, one needs to explore what Cauchy sequence really means, but this is not so difficult. Once again, one can easily show that the above identification is a bijection.

Cantor’s approach is the first example of completing metric spaces in functional analysis. Although he calls Cauchy sequences real numbers, his greatest mathematical contributions were inspired by decimal expansions. For example, Cantor’s idea of proving \(\mathbb{R}^2\) being the same size as \(\mathbb{R}\), which he called continuum, can be described as follows. Given any two real numbers \(x = a_0.a_1a_2a_3\cdots\) and \(y = b_0,b_1b_2b_3\cdots\), define

\[
\Psi(x, y) = 0.a_1b_1\boxplus a_2b_2\boxplus a_3b_3\boxplus \cdots,
\]

where the sequence of empty boxes is left to encode the integer parts of \(x\) and \(y\), and there are plenty of ways to do so. Obviously, \(\Psi\) is injective, so the size of \(\mathbb{R}^2\) is not greater than that of \(\mathbb{R}\). Clearly, the continuum is not greater than the size of \(\mathbb{R}^2\) either, hence\(^6\) they must be the same.

---

\(^4\)Many authors replace this uniqueness condition with \(A\) having no greatest element (see e.g. [11, 16, 20]). If so, then the identification of \(\{\{x \in \mathbb{Q} : x < 1\}\} \cup \{\{x \in \mathbb{Q} : x \geq 1\}\}\) is 0.999\cdots, which does not belong to \(\mathbb{R}\).

\(^5\)For example, study the sequence \(\{1 + \frac{(−1)^n}{n}\}_{n=1}^{\infty}\).

\(^6\)This is due to the Cantor-Schröder-Bernstein theorem which states that if there exist injections \(f : A \to B\) and \(g : B \to A\), then there exists a bijection \(h : A \to B\) ([9, 11]).
5. Completeness and its application to “a fatal defect” in Analysis

In general, completeness means the real axis has no gaps. There are several equivalent ways to characterize the completeness of $\mathbb{R}$, depending on whether it is regarded as a metric space or a totally ordered set. If $\mathbb{R}$ is viewed as a metric space, then Cauchy’s criterion for convergence is a completeness property; if it is treated as a totally ordered set, then the least upper bound property plays the same role. At the end of this section, we will include a standard proof of the greatest lower bound property, the dual of the least upper bound property.

Next, we discuss an issue about the concept of angle that many authors have overlooked. Without a rigorous definition of the length of arcs, any use of the sine and cosine functions could be flawed. A common strategy in many books ([3, 6, 12, 17, 32, 38]) is to give a definition of the length of smooth curves as an application of the integration theory. The trouble is that they have already used $\sin x$ or $\cos x$ as basic examples before integration.

Kodaira ([20]) and Rudin ([30]) observed this issue, but their solutions involve the complex-valued exponential function. Courant ([8, pp. 44–45]) gave a traditional definition of the length of smooth curves as an application of the integration theory. The last inequality is due to the sum of two sides of a triangle is not smaller than $\sqrt{a^2 + b^2}$. Without a rigorous definition of the length of arcs, any use of derived trigonometric formulas ([10, pp. 81–85]) could be flawed. Hardy called the issue “a fatal defect” in his course of Pure Mathematics ([16, p. 316]). In the following, a class of curves is introduced to solve this issue.

Let $[a, b] \subset \mathbb{R}$ be a bounded interval. A curve $F = (f_1, f_2) : [a, b] \to \mathbb{R}^2$ is said to be monotone if its coordinate functions $f_1$ and $f_2$ are monotone (increasing or decreasing). Since the “shortest” curve connecting two points is a straight line, we expect the “length” of $F$, denoted by $\mathcal{L}(F)$, to be an upper bound for that of $F(a)F(b)$. Suppose we divide $[a, b]$ into two small pieces $[a, c] \cup [c, b]$. Clearly,

$$\mathcal{L}(F) = \mathcal{L}(F|_{[a,c]}) + \mathcal{L}(F|_{[c,b]}) \geq \mathcal{L}(F(a)F(c)) + \mathcal{L}(F(c)F(b)) = \mathcal{L}(F(a)F(b)),$$

where the last inequality is due to the sum of two sides of a triangle is not smaller than the third part. Continuing in this procedure, the quantity closest to $\mathcal{L}(F)$ we can find is

$$\sup \left\{ \sum_{i=1}^{n} \mathcal{L}(F(c_{i-1})F(c_i)) : a = c_0 < c_1 < c_2 < \cdots < c_n = b, \quad n \in \mathbb{N} \right\},$$

where $\sup(\cdot)$ denotes the least upper bound for a given bounded-above subset of $\mathbb{R}$. Note

$$\sum_{i=1}^{n} \mathcal{L}(F(c_{i-1})F(c_i)) = \sum_{i=1}^{n} \sqrt{(f_1(c_{i-1}) - f_1(c_i))^2 + (f_2(c_{i-1}) - f_2(c_i))^2}$$

$$\leq \sum_{i=1}^{n} \left( |f_1(c_{i-1}) - f_1(c_i)| + |f_2(c_{i-1}) - f_2(c_i)| \right)$$

$$= |f_1(a) - f_1(b)| + |f_2(a) - f_2(b)|,$$

where the inequality is due to $\sqrt{x^2 + y^2} \leq |x| + |y|$, and the last equality owes to the monotonicity of $f_1$ and $f_2$. So the existence of (5.1) is guaranteed by the least upper bound property.

**Definition 5.1.** Given a monotone curve $F : [a, b] \to \mathbb{R}^2$, define its length $\mathcal{L}(F)$ to be the supremum (5.1).
More generally, a curve \( F : [a, b] \to \mathbb{R}^2 \) is said to be rectifiable ([8, 30]) if (5.1) exists (as an element of \( \mathbb{R} \)), or equivalently if its coordinate functions are of bounded variation. Jordan’s decomposition theorem ([9, p. 173]) states that any function of bounded variation can be written as the difference between two monotone functions, hence Definition 5.1 is very close to the ultimate scenario of rectifiable curves.

Since the semicircle \( S : t \mapsto (t, \sqrt{1 - t^2}) \) \((-1 \leq t \leq 1)\) is two-piece monotone, its length exists and is denoted by \( \pi \). In exactly the same way, one can define the length of subarcs of \( S \), which is regarded the same as the concept of the angle of the corresponding sectors. One can easily extend the length concept to all arcs, and show that the length of the curve concatenating two concentric arcs is the sum of those of two pieces. So we are free to use any trigonometric formulas taught in school.

We include a proof of the greatest lower bound property\(^7\) whose dual plays a vital role for many authors ([1, 15, 19, 21]) in constructing the real number system. Let \( A = \{a_0^{(\lambda)}, a_1^{(\lambda)}, a_2^{(\lambda)}, a_3^{(\lambda)} \ldots \in \mathbb{R} : \lambda \in \Lambda \} \) be a non-empty subset that is bounded from below. Denote \( z^{(\lambda)} = a_0^{(\lambda)}, a_1^{(\lambda)}, a_2^{(\lambda)}, a_3^{(\lambda)} \ldots \). Pick first the smallest integer \( b_0 \) from \( \{a_0^{(\lambda)} : z^{(\lambda)} \in A \} \); then the smallest integer \( b_1 \) from \( \{a_1^{(\lambda)} : z^{(\lambda)} \in A, a_0^{(\lambda)} = b_0 \} \), and continue in this way to get an element \( b_0, b_1, b_2, b_3 \ldots \), which belongs to \( \mathbb{R} \) and is the greatest lower bound for \( A \).

6. Computational discussions

In the previous section, we discussed the concept of rectifiable curves, but did not explain how to do practical calculation, which is a standard topic in integration. Considering the field of real numbers has been reestablished in Sections 2 and 3, we study some relevant computational issues.

6.1. General principles. Readers may ask which case of the addition (or multiplication) definition happens more frequently. To answer this question, knowledge about Cantor’s continuum or Lebesgue’s measure theory ([32]) is required. (1) Let \( x \in \mathbb{R} \) be fixed, and let \( y \in \mathbb{R} \) be arbitrary. Since the set of all terminating decimals is countable, its complement must be uncountable because \( \mathbb{R} \) is uncountable. So in this sense, the second case is more likely to be applied to compute \( x + y \). (2) Let \( x \) and \( y \) be arbitrary. If the cardinality criterion is replaced by measure, then the answer is the same; if not, we first need go back to Cantor’s paradise and understand \( \mathbb{R}^2 \) as being the same size as \( \mathbb{R} \) (see Section 4.2), then give a reasonable explanation.

6.2. Optimal bounds. Hua’s definition of addition implies that

\[
(x_{k+1} + y_{k+1})_k \leq (x + y)_k \leq (x_{k+1} + y_{k+1})_k + 10^{-k},
\]

which is tighter than the more frequently used

\[
x_k + y_k \leq (x + y)_k \leq x_k + y_k + 2 \cdot 10^{-k}.
\]

Thus \((x + y)_k\) has only two choices: one is \((x_{k+1} + y_{k+1})_k\); the other is \((x_{k+1} + y_{k+1})_k + 10^{-k}\). The chance (or probability) of getting the first one is 0.95, and its proof is left as an exercise.

To be clear, given the integer parts of \( x \) and \( y \) have nothing to do with the purpose, the sample space is taken to be \([0,1) \times [0,1)\).

\(^7\)Why not the least upper bound property? Everything is essentially the same except one more identification procedure needs to be included.
Next, consider the multiplication and let \( \Omega = \{ (x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + y \leq 1 \} \) be the sample space with doubled Lebesgue measure. According to Definition 1.2, \((xy)\) has only two choices: one is \((x_{k+1}y_{k+1})k\); the other is \((x_{k+1}y_{k+1})k + 10^{-k}\). Since the chance of getting the first choice on each vertical segment \(\{(x, y) \in \Omega : 0 \leq y \leq 1 - x\}\) with \(x\) being fixed is not less than 0.9, so is the chance of the same choice on \(\Omega\) by Fubini’s theorem.

6.3. Turing’s computable numbers. A real number consists of an infinitely long string of elements of \(\mathbb{Z}_{10}\), so in principle its digits carry an infinite amount of information. According to Turing ([35], see also [22]), a real number is said to be computable if there exists a finite, terminating algorithm such that given any positive integer \(k\), it outputs the first \(k\) digits of that number\(^8\). To be clear, we understand that the given object is not explicitly stated as its identification element of \(\mathbb{R}\), but means a representation or disguise that is implicitly determined as a root of an equation (such as \(\sqrt{2}\), the positive root of \(x^2 - 2 = 0\)), an arithmetical operation of numbers (such as \(\pi + \sqrt{2}\)), a Dedekind cut, a supremum, the limit of a convergent sequence or series (such as Euler’s number \(e\), see [8, p. 43]), and so on.

We claim that if it is known in advance that the sum, subtraction, product, or division of two computable numbers does not terminate, then the arithmetical output is computable. The first three cases follow on from the corresponding addition, subtraction and multiplication definitions. To study the division case, we can assume without loss of generality that \(0 < x \leq 1 \leq y\). Note for any positive integer \(k\), \((x/y)k = (x_{2k}/y_{2k})k\) and

\[
0 \leq \frac{x}{y} - \frac{x_{2k}}{y_{2k}} < 10^{-k}.
\]

Thus if \(\theta_k(x_{2k}/y_k) > 0\), then \((x/y)k-1 = (x_{2k}/y_{2k})k-1\). We now have two cases to consider. In the first case, suppose there exists an increasing sequence of positive integers \(\{k_i\}_{i=1}^{\infty}\) such that \(\theta_k(x_{2k}/y_{k_i}) > 0\) for all \(i \in \mathbb{N}\). According to the above analysis, \(x/y\) is computable. In the second case, suppose there exists a non-negative integer \(m\) such that \(\theta_k(x_{2k}/y_k) = 0\) for all \(k > m\). Similar to the illustrative example of Definition 1.2 or the rigorous justification of Definition 2.2, one can show that \(x/y\) is a terminating decimal. This suffices to conclude the proof of the claim.

As an application, \(\pi + \sqrt{2}, \pi - \sqrt{2}, \pi \times \sqrt{2}\) and \(\pi/\sqrt{2}\) are computable because they are transcendental. Brannan asked a similar question about computing \(\pi + \sqrt{2}\), and his solution relies on the least upper bound property ([3, p. 30]).

It is not clear which case of Hua’s definition applies to the sum \(\pi + e\). In fact, it is widely believed that \(\pi + e, \pi - e, \pi e, \pi/e\) are all irrational ([31]), and every element of \(\mathbb{Z}_{10}\) appears infinitely often in the digit sequences of \(\pi\) and \(e\), but a proof remains elusive.

7. Conclusion

From Dedekind and Cantor’s era to the present day, numerous mathematicians have continuously called for a convincing decimal construction of the real number system. This article, focusing on terminating decimals and elementary arithmetical rather than rational numbers and derived properties, together with theoretical comparison and computational discussions, is bound to have accomplished their wish, and provides an ideal way for younger generations to understand real numbers and their properties in the future.

---

\(^8\) We remark that the modern definition of computable numbers ([26]) differs from Turing’s original one.
Acknowledgements. Both authors would like to thank Prof. Jean-Pierre Demailly for encouraging comments and helpful suggestions. The second listed author would also like to thank Dr. Jiyou Li and Prof. Yaokun Wu for useful discussions.

References

[1] A. Abian, Calculus must consist of the study of real numbers in their decimal representation and not of the study of an abstract complete ordered field or nonstandard real numbers, Int. J. Math. Educ. Sci. Technol. 12 (1981), 465–472.
[2] G. Birkhoff, S. Mac Lane, A Survey of Modern Algebra, fifth edition, A K Peters/CRC Press, 2008.
[3] D. Brannan, A First Course in Mathematical Analysis, Cambridge Univ. Press, 2006.
[4] M. Bridger, Real Analysis A Constructive Approach, Jon Wiley & Sons, 2007.
[5] R. P. Burn, Irrational numbers in English language textbooks, 1890–1915: Constructions and Postulates for the completeness of the real numbers, Historica Mathematica 19 (1992), 158–176.
[6] G.-Z. Chang, J.-H. Shi, A Course in Mathematical Analysis, Volume 1, in Chinese, Higher Education Press, 2003.
[7] K.-L. Chung, A Course in Probability Theory, third edition, Elsevier, 2001.
[8] R. Courant, Differential and Integral Calculus, Volume 1, Blackie & Sons Limited, Glasgow and London, translated by E. J. McShane, 1937.
[9] E. DiBenedetto, Real Analysis, Birkhäuser Boston, 2002.
[10] J.-P. Demailly, Perspectives pour une renaissance de l’enseignement des mathématiques dans le primaire et le seconnaire, available at https://www-fourier.ujf-grenoble.fr/~demailly/, 2010.
[11] H. B. Enderton, Elements of Set Theory, Elsevier Science (USA), 1977.
[12] G. M. Fichtenholz, Differential and Integral Calculus, Volume 1, eighth edition, in Russia, Fizmatlit Publishers Russia, 2003.
[13] A. Gardiner, Infinite Processes: Background to Analysis, Springer-Verlag New York, 1982.
[14] T. W. Gamelin, What really are real numbers? available at http://www.math.ucla.edu/~twg, 2006.
[15] T. Gowers, What is so wrong with thinking of real numbers as infinite decimals? available at https://www.dpmms.cam.ac.uk/~wtg10/.
[16] G. H. Hardy, A Course of Pure mathematics, Centenary Edition, Cambridge Univ. Press, 2008.
[17] L.-K. Hua, An Introduction to Higher Mathematics, Volume 1, first published in 1962, in Chinese by Higher Education Press in 2009, in English by Cambridge University Press in 2012 (translated by Peter Shiu).
[18] K. U. Katz, M. G. Katz, Stevin numbers and reality, Foundations of Science 17 (2012), 109–123.
[19] M. Klazar, Real numbers as infinite decimals and irrationality of $\sqrt{2}$, arXiv:0910.5870, 2009.
[20] K. Kodaira, An Introduction to Calculus, in Japanese, Iwanami Shoten, Publishers, Tokyo, 2003.
[21] L. Li, A new approach to the real numbers, arXiv:1101.1800, 2011.
[22] M. L. Minshy, Computation: Finite and Infinite Machines, Prentice Hall, 1967.
[23] J. J. O’Connor, E. Robertson, The real numbers: Pythagoras to Stevin, available at http://www-history.mcs.st-and.ac.uk/history, 2005.
[24] J. J. O’Connor, E. Robertson, The real numbers: Stevin to Hilbert, available at http://www-history.mcs.st-and.ac.uk/history, 2005.
[25] A. C. Patton, The End of Infinity: Where Mathematics and Philosophy Meet, Algora Publishing, 2018.
[26] H. G. Rice, Recursive real numbers, Proc. Amer. Math. Soc. 5 (1954), 784–791.
[27] F. Richman, Interview with a constructive mathematician, Modern Logic 6 (1996), 247–271.
[28] F. Richman, Is 0.999... = 1? Mathematics Magazine 72 (1999), 396–400.
[29] F. Richman, Real numbers and other completions, Math. Log. Quart. 54 (2008), 98–108.
[30] W. Rudin, Principles of Mathematical Analysis, third edition, McGraw-Hill, 1976.
[31] J. H. Silverman, A Friendly Introduction to Number Theory, third edition, Pearson Education, 2006.
[32] M. Stoll, Introduction to Real Analysis, second edition, Addison-Wesley Longman, 2001.
[33] O. Stolz, J. A. Gmeiner, Theoretische Arithmetik, Part 2, in German, B.G. Teumber, 1915.
[34] T. Tao, Analysis I, Hindustan Book Agency, 2006.
[35] A. M. Turing, On computable numbers, with an application to the Entscheidungsproblem, Proc. London. Math. Soc. 42 (1936), 230–265.
[36] I. Weiss, The real numbers - a survey of constructions, Rocky Mountain J. Math. 45 (2015), 737–762.
[37] Wen-Tsun Wu, Loo-Keng Hua: people’s mathematician, available at http://news.sciencenet.cn/html-news/2010/10/238974.shtm, 2010.
[38] V. A. Zorich, Mathematical Analysis I, fourth edition, Springer-Verlag Berlin Heidelberg, 2004.

LGT DE LATRE DE TASSIGNY, 85000 LA ROCHE-SUR-YON, FRANCE

E-mail address: Nicolas.Fardin@ac-nantes.fr

DEPARTMENT OF MATHEMATICAL SCIENCES, LOUGHBOURGH UNIVERSITY, LE11 3TU, UK

E-mail address: liliangpan@gmail.com