

Infinitesimal Analysis

Pramook Khungurn

December 11, 2019

This document is written as I read “A Primer to Infinitesimal Analysis” by John L. Bell [1].

1 The Infinitesimal

- Consider the real line. Is it really consist of points? A point is *indivisible*. However, the real line is a *continuum*: it is supposed to be indefinitely divisible. So, no continuum can be composed of an indivisible, right?
- Yet, the set-theoretical formulation of mathematics requires that all entities in mathematics be created from a discrete entity. The continuous nature of the real line comes from the properties of additional structures imposed on the sets of discrete things. Continuity in real analysis does not mean anything if you do not have a metric. This is a rather sneaky way of settling the philosophical question above.
- When a continuum is subdivided further and further indefinitely, you get the *infinitesimal*. It's the “ultimate part” of a continuum in the same sense that we say that the ultimate part of a curve is an infinitesimal stragith line.
- Here are some practical interpretation of the infinitesimal:
 - An infinitesimal quantity is not necessarily coincide with zero, but is smaller than any finite value.
 - An infinitesimal quantity is a quantity so small that its square and all higher powers can be set to zero. (Such a quantity is called a *nilsquare infinitesimal*.)
- A true continuum is indefinitely divisible, so any part of it must also be divisible. So, an infinitesimal must also be divisible. We call an infinitesimal with this property a *nonpunctiform infinitesimal*.
- The infinitesimal played a great role in mathematics and physics before the later half of the nineteenth century. From that point, it has been supplanted by the concept of limits and is treated as non-rigorous.
- However, scientists, engineers, and even differential geometers still use them to solve and reason about problems correctly.

2 Smooth Infinitesimal Analysis

- Nowadays, however, the infinitesimals have been given rigourous basis by mathematical logic.
 - In the 1960s, Abraham Robinson added the infinitesimal to the real number system and used it to form the basis of calculus.
 - In the 1970s, advances in category theory leads to the formation of *smooth infinitesimal analysis* (SIA), which is a rigorous axiomatic theory of the infinitesimals. It can be used to construct calculus and differential geometry.

- To show that the axioms of SIA is consistent, you have to construct a *model* for it.
- A *model* is a mathematical structure containing geometric objects like the real line and the Euclidean spaces, together with transformations between them.
- In models of SIA, all maps between objects are *smooth*. That is, it must be differentiable arbitrary many times. Smoothness implies that the maps are continuous by nature.
- As a result, we call a model of SIA a *smooth world*.
We let \mathbb{S} denote an arbitrary smooth world.
- In a smooth world, however, *the law of excluded middle* is not true in a smooth world.
The law of excluded middle says that every statement is either definitely true or definitely false.
Then, if something is not false, it is not true. If something is not true, it is not false.
- The law of excluded middle cannot hold in any \mathbb{S} because it allows us to construct a discontinuous function in \mathbb{S} .

The construction is very simple. If we assume the law of excluded middle, then each real number is either equal to 0 or not equal to 0. Then, we can define

$$f(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0, \end{cases}$$

which is outright discontinuous.

More precisely, the construction shows that the following statement is false in \mathbb{S} :

for any real number x , either $x = 0$ or *not* $x = 0$.

- Here's another way to show that a statement in \mathbb{S} cannot be regarded as possessing one of just two 'truth value' T or F .

Let Ω be the set of truth values in \mathbb{S} . We assume that T and F are members of this set.

Consider any object X in \mathbb{S} . A map of X to Ω in \mathbb{S} partitions X into parts. A non-constant function splits X into proper non-empty parts. However, there is no non-constant continuous function from X to a two-valued set $\{T, F\}$. Therefore, the set of truth values cannot be reduced to just a set of two elements.

As such, the logic in a smooth world is *many-valued* or *polyvalent* (as opposed to standard logic, which is *bivalent*).

- We say that a part U of an object X is *detachable* if there is a complementary part V such that
 - U and V are disjoint, and
 - U and V together cover X .
- In a smooth world, the only detachable parts of any object X is X itself and the empty part.
If this is not true, then there's a function from X to $\{T, F\}$.
- The absence of the law of excluded middle allows the infinitesimals to exist because it creates a sweet spot between zero and things that are not zero.
- More precisely, if the law of excluded middle is false, then the *law of double negation* ($\neg\neg A \rightarrow A$) is also false.

So, we say that point a and b are *distinguishable* or *distinct* if *not* $a = b$, and we shall denote this by $a \neq b$.

Now, if *not* $a \neq b$, then it does not mean that $a = b$.

- So, let I be the set of points *indistinguishable from 0*. That is, $I = \{x : \neg(x \neq 0)\}$. Now, I is nonpunctiform in the sense that it does not reduce to $\{0\}$. That is,

it is not the case that 0 is the sole member of I .

If we call the members of I the infinitesimals, then this statement must be rephrased as:

it is not the case that all infinitesimals coincide with 0.

On the other hand, we cannot state that

there exists an infinitesimal which is $\neq 0$.

Otherwise, we have something that is distinguishable from 0 while being not distinguishable from it, which is a contradiction.

- A fine point to take from the above discussion is that the *law of non-contradiction* and the *law of excluded middle* are not the same. They are:
 - **Non-contradiction:** a statement and its negation cannot both be true.
 - **Excluded middle:** a statement must either be true or false.

The law of non-contradiction continues to hold in \mathbb{S} as it allows a statement to be some other thing which is not both true and false.

- This implies that infinitesimals can only exist “virtually.” We can say that not all infinitesimals are zero, but we would not be able to pinpoint any infinitesimal that is not zero.

3 Two Concepts of Nonpunctiform Infinitesimals

- There are two concepts of nonpunctiform infinitesimals in traditional mathematics:
 - **Linear infinitesimals:** any curve can be represented as the ‘sum’ of an (infinite) discrete collection of infinitesimally short linear segments.
 - **Indivisibles:** a continuous surface or volume can be conceived as the sum of an indefinitely large, but discrete assemblage of lines or planes.
- These two concepts are related and both give rise to the notion of nilsquare infinitesimal.
- To see this connection, consider the situation where we want to find the area under the curve AB in Figure 1. We want to compute it as sum of thin rectangles $XYRS$.
- If X and S are distinguishable, then Y and R are so distinguishable. As a result, the ‘area defect’ ∇ under the curve is not zero. As such, the area of $ABCO$ cannot be regarded as the sum of area of rectangles such as $XYRS$.
On the other hand, if X and S coincide, then ∇ is zero, but $XYRS$ collapses to a straight line and cannot contribute to any area.
- So, in order to have the area of $ABCO$ be a sum of rectangles like $XYRS$, we require that:
 - X and S be indistinguishable, but not coinciding, and
 - the area defect ∇ is zero.
- The first requirement in the above list necessitates the segment XS to be a linear infinitesimal that is not identical with a single point (aka *non-degenerate*). Isaac Barrow called this a *linelet*.

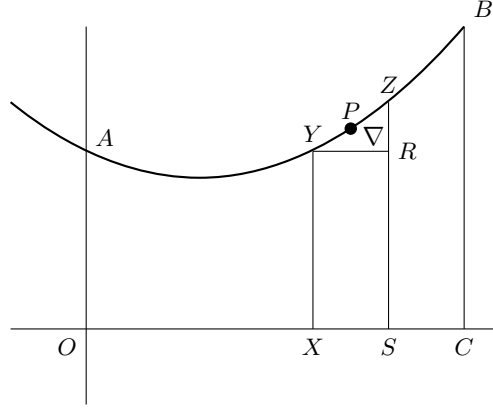


Figure 1: Finding the area under the curve.

- The second requirement can be satisfied by two conditions:
 - The segment YZ around point P is actually straight and non-generate. (That is, it does not reduces to P .)
 - XS has non-degenerate length ε with $\varepsilon^2 = 0$. (That is, the length is a nilsquare infinitesimal.)

With these two conditions, we have that the area deficit ∇ is $\frac{1}{2}(\varepsilon) \cdot (m\varepsilon) = \frac{1}{2}m\varepsilon^2 = 0$.

Now, the rectangle $XYRS$ can be regarded as an indivisible whose sum exhaust the whole area of $ABCO$.

- From the above result, we may conclude that *an indivisible is a rectangle whose base is a linelet, and a linelet's length is nilsquare*.
- Let us generalize the two conditions above to arbitrary curves and give them names:
 - **Principle of microstraightness:** For any smooth curve C and any point P on the curve, there is a small, non-degenerate segment of C around P that is straight. We call this segment a *microsegment*, and we say that C is *microstraight* around P .
 - **Existence of nilsquare infinitesimals:** The set Δ of magnitudes ε such that $\varepsilon^2 = 0$ does not reduces to $\{0\}$.
- We can see that the existence of nilsquare infinitesimals is a direct consequence of the principle of microstraightness.

To see this, consider the curve $y = x^2$. Let U be the microsegment of the curve around the point $(0,0)$. This microsegment is the intersection between the curve $y = x^2$ and the line $y = 0$. Note that, for any point x that is in U , we have that $x^2 = 0$, so $x \in \Delta$. However, since we asserted that U is non-degenerate, we have that Δ must also be non-degenerate.
- The principle of microstraightness is related to two other principles:
 - **Leibniz's principle of continuity:** Processes in nature occur continuously.
 - **Principle of microuniformity:** Processes in nature can be consider to occur at a constant rate over any sufficiently small period of time (a *timelet*).

For example, movement over a short enough period of time can be said to have constant speed.

- Principle of continuity is a consequence of principle of microuniformity. (Proof is not given in the book.)
- Principle of microuniformity, when applied to the movement of a particle along a curve, becomes the principle of microstraightness.
- The principle of microstraightness solves Zeno's paradox of arrow. It says that the smallest unit of movement is not a point in time where things are steady, but rather a very small movement over a very short, undetectable time, but the time is long enough to render some undetectable change into the position.
- In other words, a curve can be thought of consisting of short rigid rods that is so short it cannot said to have finite, non-zero length. The segment is long enough to have some direction to indicate some motion, but not long enough to bend.
- The principle of microstraightness assigns a quantitative meaning to the *instataneous rate of change*: it is just the slope of the microsegment around a point.

4 Properties of Smooth Worlds

4.1 The Real Line

- A fundamental object in a smooth world \mathbb{S} is a straight, indefinitely divisible line R . This is called the *smooth line*, the *affine line*, or the *real line*.
- There is a set of locations on R . The location comes with a relation $=$, which denotes *coincidence* or *identify* of locations.

We write $a \neq b$ for *not* $a = b$. This is read as a is *distinct* or *distinguishable* from b .

We do not assume that the $=$ relation is decidable: it is not true that, for any two locations a and b , either $a = b$ or $a \neq b$.

- There are two distinct points on R , which we call 0 and 1. They are called *zero* and *unit*, respectively.

4.2 Addition

- There's a unary operator $-$ such that, for each location a , it assigns a point $-a$ to a where $-a$ is the mirror image of a around 0.

We require that the $-$ operator satisfies the following properties:

- $-0 = 0$, and
- for all a , $-(-a) = a$.

- For each pair of points a and b , let $a \hat{=} b$ denoted the *oriented* line segment from a to b .

We say that $a \hat{=} b = c \hat{=} d$ if $a = b$ and $c = d$.

- We denote the segment $0 \hat{=} a$ as a^* . We call a^* the *segment of length a* .

The function $a \mapsto a^*$ is a bijection.

A segment of length a can be thought of as “oriented magnitude” because $(-a)^*$ has the same length as a^* but points in the opposite direction.

- Given two segments a^* and b^* , we let $a^* : b^*$ to be the segment obtained by juxtaposing a^* and b^* in that order.

- If $a^* : b^*$ is of the form c^* for a unique point c , then we call c the *sum* of a and b , and we denote c as $a + b$.

We assume that the sum forms an Abelian group structure on the positions with 0 being the identity element, and $-a$ acting as a 's inverse. That is, it follows the following rules:

- $0 + a = a$,
- $a + (-a) = 0$,
- $a + b = b + a$, and
- $a + (b + c) = (a + b) + c$.
- For brevity, we let $a - b$ denote $a + (-b)$.

4.3 Multiplication

- We assume that we can form Cartesian product $R \times R$, and also Cartesian powers $R^n = R \times R \times \cdots \times R$.
- Each point in R^n is an n -tuple (a_1, a_2, \dots, a_n) where each a_i is a point in R .
- We say that two points $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$ are *distinct*, and write $\mathbf{a} \neq \mathbf{b}$, if $a_i \neq b_i$ for some i .
- A product of two points a and b , denoted $a.b$, can be constructed geometrically in $R \times R$ as follows.

Given two magnitude a and b , construct three points on the plane $R \times R$:

- $I = (1, 0)$,
- $B = (0, b)$, and
- $A = (a, 0)$.

Because I and B are distinct points, they determine a unique line IB . The line through A that is parallel to IB intersects the y -axis at a point C . The y -coordinate of this point is defined to be the product $a.b$. The situation is depicted in the following figure.

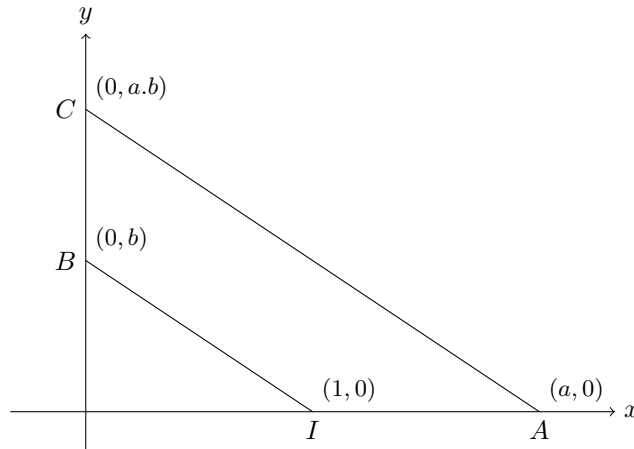


Figure 2: Geometric definition of the product of two real numbers.

- In a smooth world, we do not assume that, if $ab = 0$, then either $a = 0$ or $b = 0$. The reason is that we do not want to exclude the possibility that $\varepsilon^2 = 0$ while it is not true that $\varepsilon = 0$.

- Given $a \neq 0$, we can construct the multiplicative inverse a^{-1} of a geometrically as follows. We construct 3 points on the plane:

- $A = (a, 0)$,
- $B = (0, 1)$, and
- $I = (1, 0)$,

Since point A and B are distinct, they determine a unique line segment AB . Construct a line segment passing through I , and let the line segment intersect the y -axis at point C . The y -coordinate of C is defined to be the multiplicative inverse of a . The situation is depicted in the figure below.

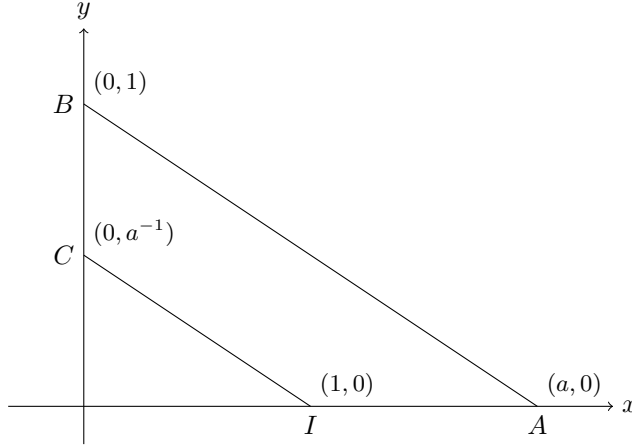


Figure 3: Geometric definition of the multiplicative inverse of a non-zero real number.

- We assume that the product and the multiplicative inverses satisfy the following rules:

- $0.a = 0$,
- $1.a = a$,
- $a.b = b.a$,
- $a.(b.c) = (a.b).c$,
- $a.(b + c) = a.b + a.c$, and
- $a \neq 0$ implies $a/a = 1$.

- That is, the operator $+$ and $.$ form a field on the set of points in R .

References

- [1] John L Bell. *A Primer of Infinitesimal Analysis; 2nd ed.* Cambridge Univ. Press, Leiden, 2008.