# Infinitesimals: History & Application

Joel A. Tropp

Plan II Honors Program, WCH 4.104, The University of Texas at Austin, Austin, TX 78712

Abstract. An infinitesimal is a number whose magnitude exceeds zero but somehow fails to exceed *any* finite, positive number. Although logically problematic, infinitesimals are extremely appealing for investigating continuous phenomena. They were used extensively by mathematicians until the late $19^{\text{th}}$ century, at which point they were purged because they lacked a rigorous foundation. In 1960, the logician Abraham Robinson revived them by constructing a number system, the hyperreals, which contains infinitesimals and infinitely large quantities.

This thesis introduces Nonstandard Analysis (NSA), the set of techniques which Robinson invented. It contains a rigorous development of the hyperreals and shows how they can be used to prove the fundamental theorems of real analysis in a direct, natural way. (Incredibly, a great deal of the presentation echoes the work of Leibniz, which was performed in the $17^{\text{th}}$ century.) NSA has also extended mathematics in directions which exceed the scope of this thesis. These investigations may eventually result in fruitful discoveries.

# Contents

To Millie, who sat in my lap every time I tried to work.
To Sarah, whose wonderfulness catches me unaware.
To Elisa, the most beautiful roommate I have ever had.
To my family, for their continuing encouragement.
And to Jerry Bona, who got me started and ensured that I finished.

Traditionally, an infinitesimal quantity is one which,
while not necessarily coinciding with zero,
is in some sense smaller than any finite quantity.

—J.L. Bell [**2**, p. 2]

---

Infinitesimals . . . must be regarded as
unnecessary, erroneous and self-contradictory.

—Bertrand Russell [**13**, p. 345]

# Introduction: Why Infinitesimals?

What is the slope of the curve $y = x^2$ at a given point? Any calculus student can tell you the answer. But few of them understand *why* that answer is correct or *how* it can be deduced from first principles. Why not? Perhaps because classical analysis has convoluted the intuitive procedure of calculating slopes.

One calculus book [**16**, Ch. 3.1] explains the standard method for solving the slope problem as follows.

> Let $P$ be a fixed point on a curve and let $Q$ be a nearby movable point on that curve. Consider the line through $P$ and $Q$, called a *secant line*. The *tangent line* at P is the limiting position (if it exists) of the secant line as $Q$ moves toward $P$ along the curve (see Figure 0.1).
>
> Suppose that the curve is the graph of the equation $y = f(x)$. Then $P$ has coordinates $(c, f(c))$, a nearby point $Q$ has coordinates $(c+h, f(c+h))$, and the secant line through $P$ and $Q$ has slope $m_{\text{sec}}$ given by (see Figure 0.2)
>
> $$m_{\text{sec}} = \frac{f(c+h) - f(c)}{h}.$$
>
> Consequently, the *tangent line* to the curve $y = f(x)$ at the point $P(c, f(c))$—if not vertical—is that

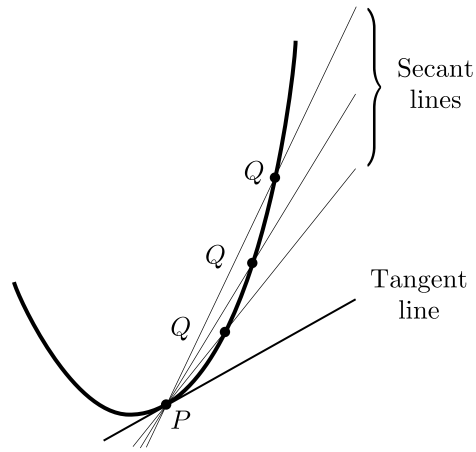


FIGURE 0.1. The tangent line is the limiting position of the secant line.

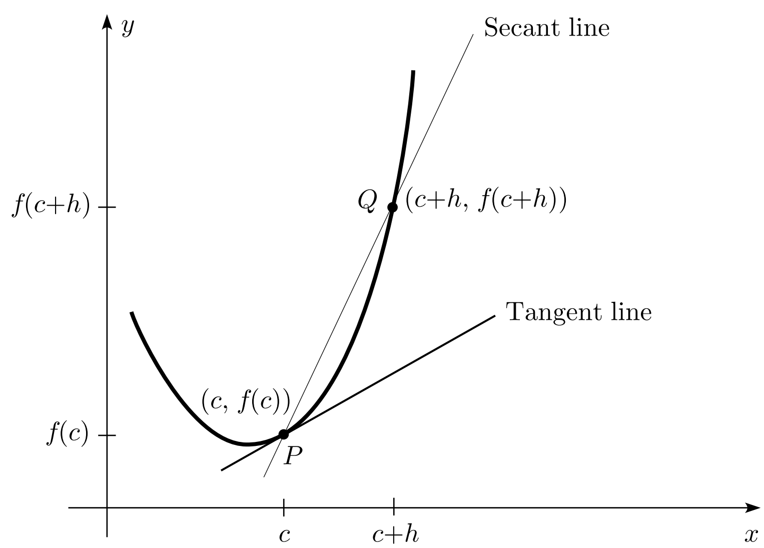


FIGURE 0.2. $m_{\tan} = \lim_{h\to 0} m_{\sec}$

line through $P$ with slope $m_{\tan}$ satisfying

$$m_{\tan} = \lim_{h\to 0} m_{\sec} = \lim_{h\to 0} \frac{f(c+h) - f(c)}{h}.$$

Ignoring any flaws in the presentation, let us concentrate on the essential idea: "The tangent line is the limiting position . . . of the secant line as $Q$ moves toward $P$." This statement raises some serious questions. What does a "limit" have to do with the slope of the tangent line? Why can't we calculate the slope without recourse to this migratory point $Q$? Rigor. When calculus was formalized, mathematicians did not see a better way.

There *is* a more intuitive way, but it could not be presented rigorously at the end of the $19^{\text{th}}$ century. Leibniz used it when he developed calculus in the $17^{\text{th}}$ century. Recent advances in mathematical logic have made it plausible again. It is called *infinitesimal calculus*.

An infinitesimal is a number whose magnitude exceeds zero but somehow fails to exceed *any* finite, positive number; it is infinitely small. (The logical difficulties already begin to surface.) But infinitesimals are extremely appealing for investigating continuous phenomena, since a lot can happen in a finite interval. On the other hand, very little can happen to a continuously changing variable within an infinitesimal interval. This fact alone explains their potential value.

Here is how Leibniz would have solved the problem heading this introduction. Assume the existence of an infinitesimal quantity, $\varepsilon$. We are seeking the slope of the curve $y = x^2$ at the point $x = c$. We will approximate it by finding the slope through $x = c$ and $x = c + \varepsilon$, a point infinitely nearby (since $\varepsilon$ is infinitesimal). To calculate slope, we divide the change in $y$ by the change in $x$. The change in $y$ is given by $y(c+\varepsilon) - y(c) = (c+\varepsilon)^2 - c^2$; the change in $x$ is $(c+\varepsilon) - c = \varepsilon$. So we form the quotient and simplify:

$$\begin{aligned}
\frac{(c+\varepsilon)^2 - c^2}{\varepsilon} &= \frac{c^2 + 2c\varepsilon + \varepsilon^2 - c^2}{\varepsilon} \\
&= \frac{2c\varepsilon + \varepsilon^2}{\varepsilon} \\
&= 2c + \varepsilon.
\end{aligned}$$

Since $\varepsilon$ is infinitely small in comparison with $2c$, we can disregard it. We see that the slope of $y = x^2$ at the point $c$ is given by $2c$. This is the correct answer, obtained in a natural, algebraic way without any type of limiting procedure.

We can apply the infinitesimal method to many other problems. For instance, we can calculate the rate of change (i.e. slope) of a sine curve at a given point $c$. We let $y = \sin x$ and proceed as before. The quotient becomes

$$\frac{\sin(c+\varepsilon) - \sin c}{\varepsilon} = \frac{\sin c \cdot \cos\varepsilon + \sin\varepsilon \cdot \cos c - \sin c}{\varepsilon}$$

by using the rule for the sine of a sum. For any infinitesimal $\varepsilon$, it can be shown geometrically or algebraically that $\cos\varepsilon = 1$ and that $\sin\varepsilon = \varepsilon$. So we have

$$\begin{aligned}\frac{\sin c \cdot \cos\varepsilon + \sin\varepsilon \cdot \cos c - \sin c}{\varepsilon} &= \frac{\sin c + \varepsilon\cos c - \sin c}{\varepsilon}\\ &= \frac{\varepsilon\cos c}{\varepsilon}\\ &= \cos c.\end{aligned}$$

Again, the correct answer.

This method even provides more general results. Leibniz determined the rate of change of a product of functions like this. Let $x$ and $y$ be functions of another variable $t$. First, we need to find the infinitesimal difference between two "successive" values of the function $xy$, which is called its *differential* and denoted $\mathrm{d}(xy)$. Leibniz reasoned that

$$\mathrm{d}(xy) = (x + \mathrm{d}x)(y + \mathrm{d}y) - xy,$$

where $\mathrm{d}x$ and $\mathrm{d}y$ represent infinitesimal increments in the values of $x$ and $y$. Simplifying,

$$\begin{aligned}\mathrm{d}(xy) &= xy + x\,\mathrm{d}y + y\,\mathrm{d}x + \mathrm{d}x\,\mathrm{d}y - xy\\ &= x\,\mathrm{d}y + y\,\mathrm{d}x + \mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

Since $(\mathrm{d}x\,\mathrm{d}y)$ is infinitesimal in comparison with the other two terms, Leibniz concluded that

$$\mathrm{d}(xy) = x\,\mathrm{d}y + y\,\mathrm{d}x.$$

The rate of change in $xy$ with respect to $t$ is given by $\mathrm{d}(xy)/\mathrm{d}t$. Therefore, we determine that

$$\frac{\mathrm{d}(xy)}{\mathrm{d}t} = x\frac{\mathrm{d}y}{\mathrm{d}t} + y\frac{\mathrm{d}x}{\mathrm{d}t},$$

which is the correct relationship.

---

At this point, some questions present themselves. If infinitesimals are so useful, why did they die off? Is there a way to resuscitate them? And how do they fit into modern mathematics? These questions I propose to answer.

CHAPTER 1

# Historical Background

Definition 1.1. An *infinitesimal* is a number whose magnitude exceeds zero yet remains smaller than *every* finite, positive number.

## 1.1. Overview

Infinitesimals have enjoyed an extensive and scandalous history. Almost as soon as the Pythagoreans suggested the concept 2500 years ago, Zeno proceeded to drown it in paradox. Nevertheless, many mathematicians continued to use infinitesimals until the end of the $19^{\text{th}}$ century because of their intuitive appeal in understanding continuity. When the foundations of calculus were formalized by Weierstrass, *et al.* around 1872, they were banished from mathematics.

As the $20^{\text{th}}$ century began, the mathematical community officially regarded infinitesimals as numerical chimeras, but engineers and physicists continued to use them as heuristic aids in their calculations. In 1960, the logician Abraham Robinson discovered a way to develop a rigorous theory of infinitesimals. His techniques are now referred to as Nonstandard Analysis, which is a small but growing field in mathematics. Practioners have found many intuitive, direct proofs of classical results. They have also extended mathematics in new directions, which may eventually result in fruitful discoveries.

## 1.2. Origins

The first deductive mathematician, Pythagoras (569?–500? B.C.), taught that all is Number. E.T. Bell describes his fervor:

> He ... preached like an inspired prophet that all nature, the entire universe in fact, physical, metaphysical, mental, moral, mathematical—*everything*—is built on the *discrete* pattern of the integers, $1, 2, 3, \ldots$ [**1**, p. 21].

Unfortunately, this grand philosophy collapsed when one of his students discovered that the length of the diagonal of a square cannot be written as the ratio of two whole numbers.

The argument was simple. If a square has sides of unit length, then its diagonal has a length of $\sqrt{2}$, according to the theorem which bears Pythagoras' name. Assume then that $\sqrt{2} = p/q$, where $p$ and $q$ are integers which do not share a factor greater than one. This is a reasonable assumption, since any common factor could be canceled immediately from the equation. An equivalent form of this equation is

$$p^2 = 2q^2.$$

We know immediately that $p$ cannot be odd, since $2q^2$ is even. We must accept the alternative that $p$ is even, so we write $p = 2r$ for some whole number $r$. In this case, $4r^2 = 2q^2$, or $2r^2 = q^2$. So we see that $q$ is also even. But we assumed that $p$ and $q$ have no common factors, which yields a contradiction. Therefore, we reject our assumption and conclude that $\sqrt{2}$ cannot be written as a ratio of integers; it is an *irrational number* [**1**, p. 21].

According to some stories, this proof upset Pythagoras so much that he hanged its precocious young author. Equally apocryphal reports indicate that the student perished in a shipwreck. These tales should demonstrate how badly this concept unsettled the Greeks [**3**, p. 20]. Of course, the Pythagoreans could not undiscover the proof. They had to decide how to cope with these inconvenient, non-rational numbers.

The solution they proposed was a crazy concept called a *monad*. To explain the genesis of this idea, Carl Boyer presents the question:

> If there is no finite line segment so small that the diagonal and the side may both be expressed in terms of it, may there not be a monad or unit of such a nature that an indefinite number of them will be required for the diagonal and for the side of the square [**3**, p. 21]?

The details were sketchy, but the concept had a certain appeal, since it enabled the Pythagoreans to construct the rational and irrational numbers from a single unit. The monad was the first infinitesimal.

Zeno of Elea (495–435 B.C.) was widely renowned for his ability to topple the most well-laid arguments. The monad was an easy target. He presented the obvious objections: if the monad had any length, then an infinite number should have infinite length, whereas if the monad had no length, no number would have any length. He is also credited with the following slander against infinitesimals:

> That which, being added to another does not make it greater, and being taken away from another does not make it less, is nothing [**3**, p. 23].

The Greeks were unable to measure the validity of Zeno's arguments. In truth, ancient uncertainty about infinitesimals stemmed from a greater confusion about the nature of a *continuum*, a closely related question which still engages debate [**1**, pp. 22–24].

### 1.3. Continuity

Zeno propounded four famous paradoxes which demonstrate the subtleties of continuity. Here are the two most effective.

> The *Achilles*. Achilles running to overtake a crawling tortoise ahead of him can never overtake it, because

> he must first reach the place from which the tortoise started; when Achilles reaches that place, the tortoise has departed and so is still ahead. Repeating the argument, we see that the tortoise will always be ahead.
>
> The *Arrow*. A moving arrow at any instant is either at rest or not at rest, that is, moving. If the instant is indivisible, the arrow cannot move, for if it did the instant would immediately be divided. But time is made up of instants. As the arrow cannot move in any one instant, it cannot move in any time. Hence it always remains at rest.

The *Achilles* argues that the line cannot support infinite division. In this case, the continuum must be composed of finite atomic units. Meanwhile, the *Arrow* suggests the opposite position that the line *must* be infinitely divisible. On this second view, the continuum cannot be seen as a set of discrete points; perhaps infinitesimal monads result from the indefinite subdivision.

Taken together, Zeno's arguments make the problem look insoluble; either way you slice it, the continuum seems to contradict itself [**1**, p. 24]. Modern mathematical analysis, which did not get formalized until about 1872, is necessary to resolve these paradoxes [**3**, pp. 24–25].

Yet, some mathematicians—notably L.E.J. Brouwer (1881–1966) and Errett Bishop (1928–1983)—have challenged the premises underlying modern analysis. Brouwer, the founder of Intuitionism, regarded mathematics "as the formulation of mental constructions that are governed by self-evident laws" [**4**]. One corollary is that mathematics must develop from and correspond with physical insights.

Now, an intuitive definition of a continuum is "the domain over which a continuously varying magnitude *actually varies*" [**2**, p. 1]. The phrase "continuously varying" presumably means that no jumps or breaks occur. As a consequence, it seems as if a third point must lie between *any* two points of a continuum. From this premise, Brouwer concluded that a continuum can "never be thought of as a mere collection of units [i.e. points]" [**2**, p. 2]. Brouwer might have imagined that the discrete points of a continuum cohere due to some sort of infinitesimal "glue."

Some philosophers would extend Brouwer's argument even farther. The logician Charles S. Peirce (1839–1914) wrote that

> [the] continuum does not consist of indivisibles, or points, or instants, and does not contain any except insofar as its continuity is ruptured [**2**, p. 4].

Peirce bases his complaint on the fact that it is impossible to single out a point from a continuum, since none of the points are distinct.[1] On this view, a line is *entirely* composed of a series of indistinguishable overlapping infinitesimal units which flow from one into the next [**2**, Introduction].

Intuitionist notions of the continuum resurface in modern theories of infinitesimals.

### 1.4. Eudoxus and Archimedes

In ancient Greece, there were some attempts to skirt the logical difficulties of infinitesimals. Eudoxus (408–355 B.C.) recognized that he need not assume the existence of an infinitely small monad; it was sufficient to attain a magnitude *as small as desired* by repeated subdivision of a given unit. Eudoxus employed this concept in his *method of*

[1] More precisely, all points of a continuum are topologically identical, although some have algebraic properties. For instance, a small neighborhood of zero is indistinguishable from a small neighborhood about another point, even though zero is the unique additive identity of the field $\mathbb{R}$.

*exhaustion* which is used to calculate areas and volumes by filling the entire figure with an increasingly large number of tiny partitions [**1**, pp. 26–27].

As an example, the Greeks knew that the area of a circle is given by $A = \frac{1}{2}rC$, where $r$ is the radius and $C$ is the circumference.[2] They probably developed this formula by imagining that the circle was composed of a large number of isosceles triangles (see Figure 1.1). It is important to recognize that the method of exhaustion is strictly geometrical, not arithmetical. Furthermore, the Greeks did not compute the limit of a sequence of polygons, as a modern geometer would. Rather, they used an indirect *reductio ad absurdem* technique which showed that any result other than $A = \frac{1}{2}rC$ would lead to a contradiction if the number of triangles were increased sufficiently [**7**, p. 4].

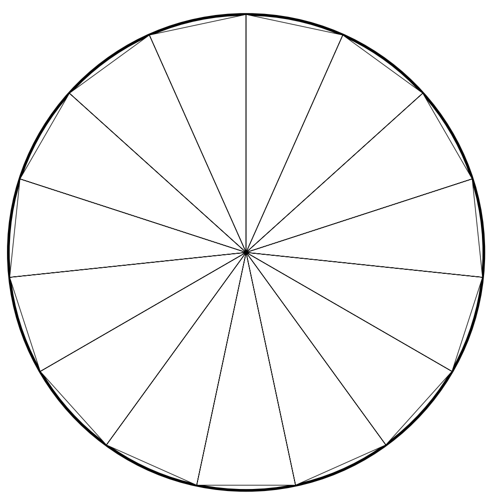

FIGURE 1.1. Dividing a circle into isosceles triangles to approximate its area.

Archimedes (287–212 B.C.), the greatest mathematician of antiquity, used another procedure to determine areas and volumes. To measure an unknown figure, he imagined that it was balanced on a

[2]The more familiar formula $A = \pi r^2$ results from the fact that $\pi$ is *defined* by the relation $C = 2\pi r$.

lever against a known figure. To find the area or volume of the former in terms of the latter, he determined where the fulcrum must be placed to keep the lever even. In performing his calculations, he imagined that the figures were comprised of an indefinite number of *laminae*—very thin strips or plates. It is unclear whether Archimedes actually regarded the laminae as having infinitesimal width or breadth. In any case, his results certainly attest to the power of his method: he discovered mensuration formulae for an entire menagerie of geometrical beasts, many of which are devilish to find, even with modern techniques. Archimedes recognized that his method did not *prove* his results. Once he had applied the mechanical technique to obtain a preliminary guess, he supplemented it with a rigorous proof by exhaustion [**3**, pp. 50–51].

## 1.5. Apply when Necessary

All the fuss about the validity of infinitesimals did not prevent mathematicians from working with them throughout antiquity, the Middle Ages, the Renaissance and the Enlightenment. Although some people regarded them as logically problematic, infinitesimals were an effective tool for researching continuous phenomena. They crept into studies of slopes and areas, which eventually grew into the differential and integral calculi. In fact, Newton and Leibniz, who independently discovered the Fundamental Theorem of Calculus near the end of the 17$^{\text{th}}$ century, were among the most inspired users of infinitesimals [**3**].

Sir Isaac Newton (1642–1727) is widely regarded as the greatest genius ever produced by the human race. His *curriculum vitae* easily supports this claim; his discoveries range from the law of universal gravitation to the method of fluxions (i.e. calculus), which was developed using infinitely small quantities [**1**, Ch. 6].

Newton began by considering a variable which changes continuously with time, which he called a *fluent.* Each fluent $x$ has an associated rate of change or "generation," called its *fluxion* and written $\dot{x}$. Moreover, any fluent $x$ may be viewed as the fluxion of another fluent, denoted $\overset{|}{x}$. In modern terminology, $\dot{x}$ is the *derivative* of $x$, and $\overset{|}{x}$ is the *indefinite integral* of $x$.[3] The problem which interested Newton was, given a fluent, to find its derivative and indefinite integral with respect to time.

Newton's original approach involved the use of an infinitesimal quantity $o$, an infinitely small increment of time. Newton recognized that the concept of an infinitesimal was troublesome, so he began to focus his attention on their ratio, which is often finite. Given this ratio, it is easy enough to find two finite quantities with an identical quotient. This realization led Newton to view a fluxion as the "ultimate ratio" of finite quantities, rather than a quotient of infinitesimals. Eventually, he disinherited infinitesimals: "I have sought to demonstrate that in the method of fluxions, it is not necessary to introduce into geometry infinitely small figures." Yet in complicated calculations, $o$ sometimes resurfaced [**3**, Ch. V].

The use of infinitesimals is even more evident in the work of Gottfried Wilhelm Leibniz (1646–1716). He founded his development of calculus on the concept of a *differential*, an infinitely small increment in the value of a continuously changing variable. To calculate the rate of change of $y = f(x)$ with respect to the rate of change of $x$, Leibniz formed the quotient of their differentials, $\mathrm{d}y/\mathrm{d}x$, in analogy to the formula for computing a slope, $\Delta y/\Delta x$ (see Figure 1.2). To find the area under the curve $f(x)$, he imagined summing an indefinite number of

[3] Newton's disused notation seems like madness, but there is method to it. The fluxion $\dot{x}$ is a "pricked letter," indicating the rate of change at a *point.* The inverse fluent $\overset{|}{x}$ suggests the fact that it is calculated by summing thin rectangular strips (see Figure 1.3).

rectangles with height $f(x)$ and infinitesimal width d$x$ (see Figure 1.3). He expressed this sum with an elongated $s$, writing $\int f(x)\,\mathrm{d}x$. Leibniz's notation remains in use today, since it clearly expresses the essential ideas involved in calculating slopes and areas [**3**, Ch. V].

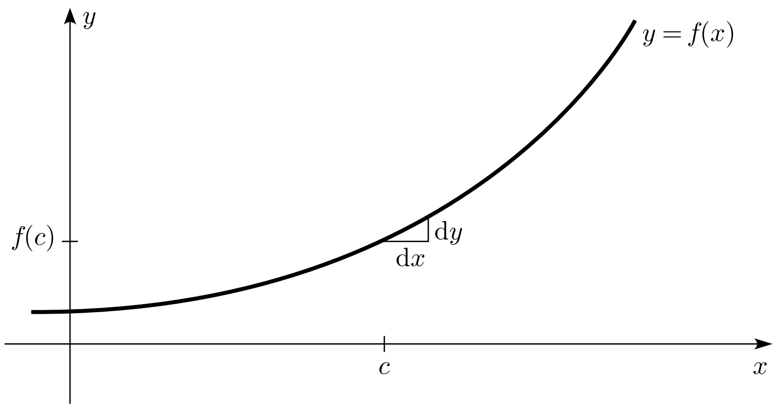


FIGURE 1.2. Using differentials to calculate the rate of change of a function. The slope of the curve at the point $c$ is the ratio d$y$/d$x$.

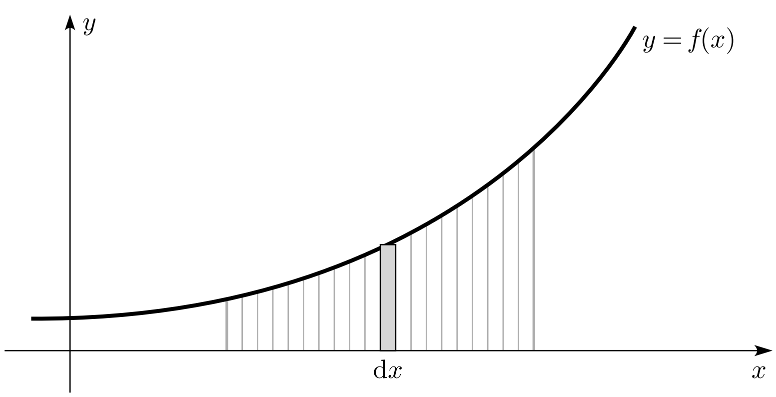


FIGURE 1.3. Using differentials to calculate the area under a curve. The total area is the sum of the small rectangles whose areas are given by the products $f(x)\,\mathrm{d}x$.

Although Leibniz began working with finite differences, his success with infinitesimal methods eventually converted him, despite ongoing doubts about their logical basis. When asked for justification, he tended to hedge: an infinitesimal was merely a quantity which may be taken "as small as one wishes" [**3**, Ch. V]. Elsewhere he wrote that it is safe to calculate with infinitesimals, since "the whole matter can be always referred back to assignable quantities" [**7**, p. 6]. Leibniz did not explain how one may alternate between "assignable" and "inassignable" quantities, a serious gloss. But it serves to emphasize the confusion and ambivalence with which Leibniz regarded infinitesimals [**3**, Ch. V].

As a final example of infinitesimals in history, consider Leonhard Euler (1707–1783), the world's most prolific mathematician. He unabashedly used the infinitely large and the infinitely small to prove many striking results, including the beautiful relation known as Euler's Equation:

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

where $i = \sqrt{-1}$. From a modern perspective, his derivations are bizarre. For instance, he claims that if $N$ is infinitely large, then the quotient $\frac{N-1}{N} = 1$. This formula may seem awkward, yet Euler used it to obtain correct results [**7**, pp. 8–9].

### 1.6. Banished

As the 19$^{\text{th}}$ century dawned, there was a strong tension between the logical inconsistencies of infinitesimals and the fact that they often yielded the right answer. Objectors essentially reiterated Zeno's complaints, while proponents offered metaphysical speculations. As the century progressed, a nascent trend toward formalism accelerated. Analysts began to prove all theorems rigorously, with each step requiring justification. Infinitesimals could not pass muster.

The first casualty was Leibniz's view of the derivative as the quotient of differentials. Bernhard Bolzano (1781–1848) realized that the derivative is a *single* quantity, rather than a ratio. He defined the derivative of a continuous function $f(x)$ at a point $c$ as the number $f'(c)$ which the quotient

$$\frac{f(c+h)-f(c)}{h}$$

approaches with arbitrary precision as $h$ becomes small. Limits are evident in Bolzano's work, although he did not define them explicitly.

Augustin-Louis Cauchy (1789–1857) took the next step by developing an arithmetic formulation of the limit concept which did not appeal to geometry. Interestingly, he used this notion to define an infinitesimal as any sequence of numbers which has zero as its limit. His theory lacked precision, which prevented it from gaining acceptance.

Cauchy also defined the integral in terms of limits; he imagined it as the ultimate sum of the rectangles beneath a curve as the rectangles become smaller and smaller [**3**, Ch. VII]. Bernhard Riemann (1826–1866) polished this definition to its current form, which avoids all infinitesimal considerations [**16**, Ch. 5], [**12**, Ch. 6].

In 1872, the limit finally received a complete, formal treatment from Karl Weierstrass (1815–1897). The idea is that a function $f(x)$ will take on values arbitrarily close to its limit at the point $c$ whenever its argument $x$ is sufficiently close to $c$.[4] This definition rendered infinitesimals unnecessary [**3**, 287].

The killing blow also fell in 1872. Richard Dedekind (1831–1916) and Georg Cantor (1845-1918) both published constructions of the real numbers. Before their work, it was not clear that the real numbers actually existed. Dedekind and Cantor were the first to exhibit sets which

[4]More formally, $L = f(c)$ is the limit of $f(x)$ as $x$ aproaches $c$ if and only if the following statement holds. For any $\varepsilon > 0$, there must exist a $\delta > 0$ for which $|c - x| < \delta$ implies that $|L - f(x)| < \varepsilon$.

satisfied all the properties desired of the reals.[5] These models left no space for infinitesimals, which were quickly forgotten by mathematicians [**3**, Ch. VII].

### 1.7. Regained

In comparision with mathematicians, engineers and physicists are typically less concerned with rigor and more concerned with results. Since their studies revolve around dynamical systems and continuous phenomena, they continued to regard infinitesimals as useful heuristic aids in their calculations. A little care ensured correct answers, so they had few qualms about infinitely small quantities. Meanwhile, the formalists, led by David Hilbert (1862-1943), reigned over mathematics. No theorem was valid without a rigorous, deductive proof. Infinitesimals were scorned since they lacked sound definition.

In autumn 1960, a revolutionary, new idea was put forward by Abraham Robinson (1918–1974). He realized that recent advances in symbolic logic could lead to a new model of mathematical analysis. Using these concepts, Robinson introduced an extension of the real numbers, which he called the hyperreals. The hyperreals, denoted ${}^*\mathbb{R}$, contain all the real numbers and obey the familiar laws of arithmetic. But ${}^*\mathbb{R}$ also contains infinitely small and infinitely large numbers.

With the hyperreals, it became possible to prove the basic theorems of calculus in an intuitive and direct manner, just as Leibniz had done in the $17^{\text{th}}$ century. A great advantage of Robinson's system is that many properties of $\mathbb{R}$ still hold for ${}^*\mathbb{R}$ and that classical methods of proof apply with little revision [**6**, pp. 281–287]. Robinson's landmark book,

[5]Never mind the fact that their constructions were ultimately based on the natural numbers, which did not receive a satisfactory definition until Frege's 1884 book *Grundlagen der Arithmetik* [**14**].

*Non-standard Analysis* was published in 1966. Finally, the mysterious infinitesimals were placed on a firm foundation [**7**, pp. 10–11].

In the 1970s, a second model of infinitesimal analysis appeared, based on considerations in category theory, another branch of mathematical logic. This method develops the *nil-square infinitesimal*, a quantity $\varepsilon$ which is not necessarily equal to zero, yet has the property that $\varepsilon^2 = 0$. Like hyperreals, nil-square infinitesimals may be used to develop calculus in a natural way. But this system of analysis possesses serious drawbacks. It is no longer possible to assert that either $x = y$ or $x \neq y$. Points are "fuzzy"; sometimes $x$ and $y$ are indistinguishable even though they are *not* identical. This is Peirce's continuum: a series of overlapping infinitesimal segments [**2**, Introduction]. Although intuitionists believe that this type of model is the proper way to view a continuum, many standard mathematical tools can no longer be used.[6] For this reason, the category-theoretical approach to infinitesimals is unlikely to gain wide acceptance.

### 1.8. The Future

The hyperreals satisfy a rule called the *transfer principle*:

> Any appropriately formulated statement is true of ${}^*\mathbb{R}$
> if and only if it is true of $\mathbb{R}$.

As a result, any proof using nonstandard methods may be recast in terms of standard methods. Critics argue, therefore, that Nonstandard Analysis (NSA) is a trifle. Proponents, on the other hand, claim that infinitesimals and infinitely large numbers facilitate proofs and permit a more intuitive development of theorems [**7**, p. 11].

[6] The specific casualties are the Law of Excluded Middle and the Axiom of Choice. This fact prevents proof by contradiction and destroys many important results, including Tychonoff's Theorem and the Hahn-Banach Extension Theorem.

New mathematical objects have been constructed with NSA, and it has been very effective in attacking certain types of problems. A primary advantage is that it provides a more natural view of standard mathematics. For example, the space of distributions, $\mathscr{D}'(\mathbb{R})$, may be viewed as a set of nonstandard functions.[7] A second benefit is that NSA allows mathematicians to apply discrete methods to continuous problems. Brownian motion, for instance, is essentially a random walk with infinitesimal steps. Finally, NSA shrinks the infinite to a manageable size. Infinite combinatorial problems may be solved with techniques from finite combinatorics [**10**, Preface].

So, infinitesimals are back, and they can no longer be dismissed as logically unsound. At this point, it is still difficult to project their future. Nonstandard Analysis, the dominant area of research using infinitesimal methods, is not yet a part of mainstream mathematics. But its intuitive appeal has gained it some formidable allies. Kurt Gödel (1906–1978), one of the most important mathematicians of the 20th century, made this prediction: "There are good reasons to believe that nonstandard analysis, in some version or other, will be the analysis of the future" [**7**, p. v].

[7] Incredibly, $\mathscr{D}'(\mathbb{R})$ may even be viewed as a set of infinitely differentiable nonstandard functions.

CHAPTER 2

# Rigorous Infinitesimals

There are now several formal theories of infinitesimals, the most common of which is Robinson's Nonstandard Analysis (NSA). I believe that NSA provides the most satisfying view of infinitesimals. Furthermore, its toolbox is easy to use. Advanced applications require some practice, but the fundamentals quickly become arithmetic.

## 2.1. Developing Nonstandard Analysis

Different authors present NSA in radically different ways. Although the three major versions are essentially equivalent, they have distinct advantages and disadvantages.

**2.1.1. A Nonstandard Extension of $\mathbb{R}$.** Robinson originally constructed a proper nonstandard extension of the real numbers, which he called the set of hyperreals, ${}^*\mathbb{R}$ [**6**, 281–287]. One approach to NSA begins by defining the nonstandard extension ${}^*\mathbb{X}$ of a general set $\mathbb{X}$. This extension consists of a non-unique mapping $*$ from the subsets of $\mathbb{X}$ to the subsets of ${}^*\mathbb{X}$ which preserves many set-theoretic properties (see Appendix A). Define the *power set* of $\mathbb{X}$ to be the collection of all its subsets, i.e. $\mathscr{P}(\mathbb{X}) = \{A : A \subseteq \mathbb{X}\}$. Then, $* : \mathscr{P}(\mathbb{X}) \to \mathscr{P}({}^*\mathbb{X})$. It can be shown that any nonempty set has a proper nonstandard extension, i.e. $\mathbb{X} \subsetneqq {}^*\mathbb{X}$. The extension of $\mathbb{R}$ to ${}^*\mathbb{R}$ is just one example. Since $\mathbb{R}$ is already complete, it follows that ${}^*\mathbb{R}$ must contain infinitely small and infinitely large numbers. Infinitesimals are born [**8**].

I find this definition very unsatisfying, since it yields no information about what a hyperreal *is*. Before doing anything, it is also necessary to prove a spate of technical lemmata. The primary advantage of this method is that the extension can be applied to any set-theoretic object to obtain a corresponding nonstandard object.[1] A minor benefit is that this system is not tied to a specific nonstandard construction, e.g. $^*\mathbb{R}$. It specifies instead the properties which the nonstandard object should preserve. An unfortunate corollary is that the presentation is extremely abstract [**8**].

**2.1.2. Nelson's Axioms.** Nonstandard extensions are involved (at best). Ed Nelson has made NSA friendlier by axiomatizing it. The rules are given *a priori* (see Appendix B), so there is no need for complicated constructions. Nelson's approach is called Internal Set Theory (IST). It has been shown that IST is consistent with standard set theory,[2] which is to say that it does not create any (new) mathematical contradictions [**11**].

Several details make IST awkward to use. To eliminate $^*\mathbb{R}$ from the picture, IST adds heretofore unknown elements to the reals. In fact, every infinite set of real numbers contains these nonstandard members. But IST provides no intuition about the nature of these new elements. How big are they? How many are there? How do they relate to the standard elements? Alain Robert answers, "These nonstandard integers have a certain *charm* that prevents us from really grasping

[1] This version of NSA strictly follows the Zermelo-Fraenkel axiomatic in regarding every mathematical object as a set. For example, an ordered pair $(a, b)$ is written as $\{a, \{a, b\}\}$, and a function $f$ is identified with its graph, $f = \{(x, f(x)) : x \in \operatorname{Dom} f\}$. In my opinion, it is unnecessarily complicated to expand every object to its primitive form.

[2] Standard set theory presumes the Zermelo-Fraenkel axioms and the Axiom of Choice.

them!" [**11**]. I see no charm.[3] Another major complaint is that IST intermingles the properties of $\mathbb{R}$ and ${}^*\mathbb{R}$, which serves to limit comprehension of both. It seems more transparent to regard the reals and the hyperreals as distinct systems.

## 2.2. Direct Ultrapower Construction of ${}^*\mathbb{R}$

In my opinion, a direct construction of the hyperreals provides the most lucid approach to NSA. Although it is not as general as a nonstandard extension, it repays the loss with rich intuition about the hyperreals. Arithmetic develops quickly, and it is based largely on simple algebra and analysis.

Since the construction of the hyperreals from the reals is analogous to Cantor's construction of the real numbers from the rationals, we begin with Cantor. I follow Goldblatt throughout this portion of the development [**7**].

**2.2.1. Cantor's Construction of $\mathbb{R}$.** Until the end of the 1800s, the rationals were the only "real" numbers in the sense that $\mathbb{R}$ was purely hypothetical. Mathematicians recognized that $\mathbb{R}$ should be an ordered field with the least-upper-bound property, but no one had demonstrated the existence of such an object. In 1872, both Richard Dedekind and Georg Cantor published solutions to this problem [**3**, Ch. VII]. Here is Cantor's approach.

Since the rationals are well-defined, they are the logical starting point. The basic idea is to identify each real number $r$ with those sequences of rationals which want to converge to $r$.

[3]In Nelson's defense, it must be said that the reason the nonstandard numbers are so slippery is that all sets under IST are internal sets (see Section 2.3.2), which are fundamental to NSA. Only the standard elements of an internal set are arbitrary, and these dictate the nonstandard elements.

Definition 2.1 (Sequence). A sequence is a function defined on the set of positive integers. It is denoted by

$$\mathbf{a} = \{a_j\}_{j=1}^{\infty} = \{a_j\}.$$

We will indicate the entire sequence by a boldface letter or by a single term enclosed in braces, with or without limits. The terms are written with a subscript index, and they are usually denoted by the same letter as the sequence.

Definition 2.2 (Cauchy Sequence). A sequence $\{r_j\}_{j=1}^{\infty} = \{r_j\}$ is *Cauchy* if it converges within itself. That is, $\lim_{j,k\to\infty} |r_j - r_k| = 0$.

Consider the set of Cauchy sequences of rational numbers, and denote them by $\mathbb{S}$. Let $\mathbf{r} = \{r_j\}$ and $\mathbf{s} = \{s_j\}$ be elements of $\mathbb{S}$. Define addition and multiplication termwise:

$$\begin{aligned} \mathbf{r} \oplus \mathbf{s} &= \{r_j + s_j\}, \text{ and} \\ \mathbf{r} \odot \mathbf{s} &= \{r_j \cdot s_j\}. \end{aligned}$$

It is easy to check that these operations preserve the Cauchy property. Furthermore, $\oplus$ and $\odot$ are commutative and associative, and $\oplus$ distributes over $\odot$. Hence, $(\mathbb{S}, \oplus, \odot)$ is a commutative ring which has zero $\mathbf{0} = \{0, 0, 0, \ldots\}$ and unity $\mathbf{1} = \{1, 1, 1, \ldots\}$.

Next, we will say that $\mathbf{r}, \mathbf{s} \in \mathbb{S}$ are equivalent to each other if and only if they share the same limit. More precisely,

$$\mathbf{r} \equiv \mathbf{s} \qquad \text{if and only if} \qquad \lim_{j\to\infty} |r_j - s_j| = 0.$$

It is straightforward to check that $\equiv$ is an equivalence relation by using the triangle inequality, and we denote its equivalence classes by $[\cdot]$. Moreover, $\equiv$ is a congruence on the ring $\mathbb{S}$, which means $r \equiv r'$ and $s \equiv s'$ imply that $r \oplus s \equiv r' \oplus s'$ and $r \odot s \equiv r' \odot s'$.

Now, let $\mathbb{R}$ be the quotient ring given by $\mathbb{S}$ modulo the equivalence.

$$\mathbb{R} = \{[\mathbf{r}] : \mathbf{r} \in \mathbb{S}\}.$$

Define arithmetic operations in the obvious way, viz.

$$[\mathbf{r}] + [\mathbf{s}] = [\mathbf{r} \oplus \mathbf{s}] = [\{r_j + s_j\}]\,, \text{ and}$$
$$[\mathbf{r}] \cdot [\mathbf{s}] = [\mathbf{r} \odot \mathbf{s}] = [\{r_j \cdot s_j\}]\,.$$

The fact that $\equiv$ is a congruence on $\mathbb{S}$ shows that these operations are independent of particular equivalence class members; they are *well-defined.*

Finally, define an ordering: $[\mathbf{r}] < [\mathbf{s}]$ if and only if there exists a rational $\varepsilon > 0$ and an integer $J \in \mathbb{N}$ such that $r_j + \varepsilon < s_j$ for each $j > J$.[4] We must check the well-definition of this relation. Let $[\mathbf{r}] < [\mathbf{s}]$, which dictates constants $\varepsilon$ and $J$. Choose $r' \equiv r$ and $s' \equiv s$. There exists an $N > J$ such that $j > N$ implies $|r_j - r'_j| < \frac{1}{4}\varepsilon$ and $|s'_j - s_j| < \frac{1}{4}\varepsilon$. Then,

$$|r_j - r'_j| + |s'_j - s_j| < \tfrac{1}{2}\varepsilon,$$

which shows that

$$|(r_j - s_j) + (s'_j - r'_j)| < \tfrac{1}{2}\varepsilon, \text{ or}$$
$$-\tfrac{1}{2}\varepsilon < (r_j - s_j) + (s'_j - r'_j) < \tfrac{1}{2}\varepsilon, \text{ which gives}$$
$$(s_j - r_j) - \tfrac{1}{2}\varepsilon < (s'_j - r'_j)$$

for any $j > N$. Since $[\mathbf{r}] < [\mathbf{s}]$ and $N > J$, $\varepsilon < (s_j - r_j)$ for all $j > N$. Then,

$$0 < \varepsilon - \tfrac{1}{2}\varepsilon < (s'_j - r'_j), \text{ or}$$
$$r'_j + \tfrac{1}{2}\varepsilon < s'_j$$

for each $j > N$, which demonstrates that $[\mathbf{r}'] < [\mathbf{s}']$ by our definition.

It can be shown that $(\mathbb{R}, +, \cdot, <)$ is a complete, ordered field. Since all complete, ordered fields are isomorphic, we may as well identify this object as the set of real numbers. Notice that the rational numbers $\mathbb{Q}$

[4]The sequences $\mathbf{r}$ and $\mathbf{s}$ do not necessarily converge to rational numbers, which means that we cannot do arithmetic with their limits. In the current context, the more obvious definition "$[\mathbf{r}] < [\mathbf{s}]$ iff $\lim_{j\to\infty} r_j < \lim_{j\to\infty} s_j$" is meaningless.

are embedded in $\mathbb{R}$ via the mapping $q \mapsto [\{q, q, q, \ldots\}]$. At this point, the construction becomes incidental. We hide the details by labeling the equivalence classes with more meaningful symbols, such as 2 or $\sqrt{2}$ or $\pi$.

**2.2.2. Cauchy's Infinitesimals.** The question at hand is how to define infinitesimals in a consistent manner so that we may calculate with them. Cauchy's arithmetic definition of an infinitesimal provides a good starting point.

Cauchy suggested that any sequence which converges to zero may be regarded as infinitesimal.[5] In analogy, we may also regard divergent sequences as infinitely large numbers. This concept suggests that rates of convergence and divergence may be used to measure the magnitude of a sequence.

Unfortunately, when we try to implement this notion, problems appear quickly. We might say that

$$\{2, 4, 6, 8, \ldots\} \quad \text{is greater than} \quad \{1, 2, 3, 4, \ldots\}$$

since it diverges faster. But how does

$$\{1, 2, 3, 4, \ldots\} \quad \text{compare with} \quad \{2, 3, 4, 5, \ldots\}?$$

They diverge at exactly the same rate, yet the second seems like it should be a little greater. What about sequences like

$$\{-1, 2, -3, 4, -5, 6, \ldots\}?$$

How do we even determine its rate of divergence?

Clearly, a more stringent criterion is necessary. To say that two sequences are equivalent, we will require that they be "almost identical."

[5]Given such an infinitesimal, $\varepsilon = \{\varepsilon_j\}$, Cauchy also defined $\eta = \{\eta_j\}$ to be an infinitesimal of order $n$ with respect to $\varepsilon$ if $\eta_j \in O\left({\varepsilon_j}^n\right)$ and ${\varepsilon_j}^n \in O\left(\eta_j\right)$ as $j \to \infty$ [**3**, Ch. VII].

**2.2.3. The Ring of Real-Valued Sequences.** We must formalize these ideas. As in Cantor's construction, we will be working with sequences. This time, the elements will be real numbers with no convergence condition specified.

Let $\mathbf{r} = \{r_j\}$ and $\mathbf{s} = \{s_j\}$ be elements of $\mathbb{R}^{\mathbb{N}}$, the set of real-valued sequences. First, define

$$\mathbf{r} \oplus \mathbf{s} = \{r_j + s_j\}, \text{ and}$$
$$\mathbf{r} \odot \mathbf{s} = \{r_j \cdot s_j\}.$$

$(\mathbb{R}^{\mathbb{N}}, \oplus, \odot)$ is another commutative ring[6] with zero $\mathbf{0} = \{0, 0, 0, \ldots\}$ and unity $\mathbf{1} = \{1, 1, 1, \ldots\}$.

**2.2.4. When Are Two Sequences Equivalent?** The next step is to develop an equivalence relation on $\mathbb{R}^{\mathbb{N}}$. We would like $\mathbf{r} \equiv \mathbf{s}$ when $\mathbf{r}$ and $\mathbf{s}$ are "almost identical"—if their *agreement set*

$$E_{\mathbf{rs}} = \{j \in \mathbb{N} : r_j = s_j\}$$

is "large." A nice idea, but there seems to be an undefined term. What is a large set? What properties should it have?

- Equivalence relations are reflexive, which means that any sequence must be equivalent to itself. Hence $E_{\mathbf{rr}} = \{1, 2, 3, \ldots\} = \mathbb{N}$ must be a large set.
- Equivalence is also transitive, which means that $E_{\mathbf{rs}}$ and $E_{\mathbf{st}}$ large must imply $E_{\mathbf{rt}}$ large. In general, the only nontrivial statement we can make about the agreement sets is that $E_{\mathbf{rs}} \cap E_{\mathbf{st}} \subseteq E_{\mathbf{rt}}$. Thus, the intersection of large sets ought to be large.

[6] Note that $\mathbb{R}^{\mathbb{N}}$ is not a field, since it contains nonzero elements which have a $\odot$-product of $\mathbf{0}$, such as $\{1, 0, 1, 0, 1, \ldots\}$ and $\{0, 1, 0, 1, 0, \ldots\}$ .

- The empty set, $\emptyset$, should *not* be large, or else every subset of $\mathbb{N}$ would be large by the foregoing. In that case all sequences would be equivalent, which is less than useful.
- A set of integers $A$ is called *cofinite* if $\mathbb{N} \setminus A$ is a finite set. Declaring any cofinite set to be large would satisfy the first three properties. But consider the sequences

  $$\mathbf{o} = \{1, 0, 1, 0, 1, \ldots\} \quad \text{and} \quad \mathbf{e} = \{0, 1, 0, 1, 0, \ldots\}.$$

  They agree nowhere, so they determine two distinct equivalence classes. We would like the hyperreals to be totally ordered, so one of $\mathbf{e}$ and $\mathbf{o}$ must exceed the other. Let us say that $\mathbf{r} < \mathbf{s}$ if and only if $L_{\mathbf{rs}} = \{j \in \mathbb{N} : r_j < s_j\}$ is a large set. Neither $L_{\mathbf{oe}} = \{j : j \text{ is even}\}$ nor $L_{\mathbf{eo}} = \{j : j \text{ is odd}\}$ is cofinite, so $\mathbf{e} \not< \mathbf{o}$ and $\mathbf{e} \not> \mathbf{o}$. To obtain a total ordering using this potential definition, we need another stipulation: for any $A \subseteq \mathbb{N}$, exactly one of $A$ and $\mathbb{N} \setminus A$ must be large.

These requirements may seem rather stringent. But they are satisfied naturally by any *nonprincipal ultrafilter* $\mathscr{F}$ on $\mathbb{N}$. (See Appendix C for more details about filters.) The existence of such an object is not trivial. Its complexity probably kept Cauchy and others from developing the hyperreals long ago. We are more interested in the applications of ${}^*\mathbb{R}$ than the minutiae of its construction. Therefore, we will not delve into the gory, logical details. Here, suffice it to say that there exists a nonprincipal ultrafilter on $\mathbb{N}$.

Definition 2.3 (Large Set). A set $A \subseteq \mathbb{N}$ is *large* with respect to the nonprincipal ultrafilter $\mathscr{F} \in \mathscr{P}(\mathbb{N})$ if and only if $A \in \mathscr{F}$.

Notation ($\{\!\!\{\mathbf{r} \mathrel{\mathscr{R}} \mathbf{s}\}\!\!\}$). In the foregoing, $E_{\mathbf{rs}}$ denoted the set of places at which $\mathbf{r} = \{r_j\}$ and $\mathbf{s} = \{s_j\}$ are equal. We need a more general notation for the set of terms at which two sequences satisfy some relation. Write

$$
\begin{aligned}
\{\!\{\mathbf{r} = \mathbf{s}\}\!\} &= \{j \in \mathbb{N} : r_j = s_j\}, \\
\{\!\{\mathbf{r} < \mathbf{s}\}\!\} &= \{j \in \mathbb{N} : r_j = s_j\}, \text{ or in general} \\
\{\!\{\mathbf{r} \mathscr{R} \mathbf{s}\}\!\} &= \{j \in \mathbb{N} : r_j \mathscr{R} s_j\}.
\end{aligned}
$$

Sometimes, it will be convenient to use a similar notation for the set of places at which a sequence satisfies some predicate $P$:

$$\{\!\{P(\mathbf{r})\}\!\} = \{j \in \mathbb{N} : P(r_j)\}.$$

Now, we are prepared to define an equivalence relation on $\mathbb{R}^{\mathbb{N}}$. Let

$$\{r_j\} \equiv \{s_j\} \quad \text{iff} \quad \{\!\{\mathbf{r} = \mathbf{s}\}\!\} \in \mathscr{F}.$$

The properties of large sets guarantee that $\equiv$ is reflexive, symmetric and transitive. Write the equivalence classes as $[\cdot]$. And notice that $\equiv$ is a congruence on the ring $\mathbb{R}^{\mathbb{N}}$.

Definition 2.4 (The Almost-All Criterion). When $\mathbf{r} \equiv \mathbf{s}$, we also say that they agree on a large set or agree at *almost all* $n$. In general, if $P$ is a predicate and $\mathbf{r}$ is a sequence, we say that $P$ holds *almost everywhere* on $\mathbf{r}$ if $\{\!\{P(\mathbf{r})\}\!\}$ is a large set.

**2.2.5. The Field of Hyperreals.** Next, we develop arithmetic operations for the quotient ring ${}^*\mathbb{R}$ which equals $\mathbb{R}^{\mathbb{N}}$ modulo the equivalence:

$${}^*\mathbb{R} = \{[\mathbf{r}] : \mathbf{r} \in \mathbb{R}^{\mathbb{N}}\}.$$

Addition and multiplication are defined by

$$
\begin{aligned}
&[\mathbf{r}] + [\mathbf{s}] = [\mathbf{r} \oplus \mathbf{s}] = [\{r_j + s_j\}]\,, \text{ and} \\
&[\mathbf{r}] \cdot [\mathbf{s}] = [\mathbf{r} \odot \mathbf{s}] = [\{r_j \cdot s_j\}]\,.
\end{aligned}
$$

Well-definition follows from the fact that $\equiv$ is a congruence. Finally, define the ordering by

$$[\mathbf{r}] < [\mathbf{s}] \quad \text{iff} \quad \{\!\{\mathbf{r} < \mathbf{s}\}\!\} \in \mathscr{F} \quad \text{iff} \quad \{j \in \mathbb{N} : r_j < s_j\} \in \mathscr{F}.$$

This ordering is likewise well-defined.

With these definitions, it can be shown that $({}^*\mathbb{R}, +, \cdot, <)$ is an ordered field. (See Goldblatt for a proof sketch [**7**, Ch. 3.6].)

This presentation is called an ultrapower construction of the hyperreals.[7] Since our development depends quite explicitly on the choice of a nonprincipal ultrafilter $\mathscr{F}$, we might ask whether the field of hyperreals is unique.[8] For our purposes, the issue is tangential. It does not affect any calculations or proofs, so we will ignore it.

**2.2.6. $\mathbb{R}$ Is Embedded in ${}^*\mathbb{R}$.** Identify any real number $r \in \mathbb{R}$ with the constant sequence $\mathbf{r} = \{r, r, r, \ldots\}$. Now, define a map $* : \mathbb{R} \to {}^*\mathbb{R}$ by

$$ {}^*r = [\mathbf{r}] = [\{r, r, r, \ldots\}]\,. $$

It is easy to see that for $r, s \in \mathbb{R}$,

$$ \begin{array}{ccl} {}^*(r+s) & = & {}^*r + {}^*s, \\ {}^*(r \cdot s) & = & {}^*r \cdot {}^*s, \\ {}^*r = {}^*s & \text{iff} & r = s, \quad \text{and} \\ {}^*r < {}^*s & \text{iff} & r < s. \end{array} $$

In addition, ${}^*0 = [\mathbf{0}] = [\{0, 0, 0, \ldots\}]$ is the zero of ${}^*\mathbb{R}$, and ${}^*1 = [\mathbf{1}] = [\{1, 1, 1, \ldots\}]$ is the unit.

THEOREM 2.5. *The map $* : \mathbb{R} \to {}^*\mathbb{R}$ is an order-preserving field isomorphism.* □

---

[7]The term *ultrapower* means that ${}^*\mathbb{R}$ is the quotient of a direct power ($\mathbb{R}^{\mathbb{N}}$) modulo a congruence ($\equiv$) given by an ultrafilter ($\mathscr{F}$).

[8]Unfortunately, the answer depends on which set-theoretic axioms we assume. The continuum hypothesis (CH) implies that we will obtain the same field (to the point of isomorphism) for any choice of $\mathscr{F}$. Denying CH leaves the situation undetermined [**7**, 33]. Both CH and not-CH are consistent with standard set theory, but Schechter's reference, *Handbook of Analysis and Its Foundations*, gives no indication that either axiom has any effect on standard mathematics [**15**].

Therefore, the reals are embedded quite naturally in the hyperreals. As a result, we may identify $r$ with ${}^*r$ as convenient.

**2.2.7. $\mathbb{R}$ Is a Proper Subset of ${}^*\mathbb{R}$.** Let $\varepsilon = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\} = \{\frac{1}{j}\}$. It is clear that $\varepsilon > \mathbf{0}$:

$$\{\!\!\{\mathbf{0} < \varepsilon\}\!\!\} = \{j \in \mathbb{N} : 0 < \tfrac{1}{j}\} = \mathbb{N} \in \mathscr{F}.$$

Yet, for any real number $r$, the set

$$\{\!\!\{\varepsilon < \mathbf{r}\}\!\!\} = \{j \in \mathbb{N} : \tfrac{1}{j} < r\}$$

is cofinite. Every cofinite set is large (see Appendix C), so $\{\!\!\{\varepsilon < \mathbf{r}\}\!\!\} \in \mathscr{F}$ which implies that $[\varepsilon] < {}^*r$. Therefore, $[\varepsilon]$ is a positive infinitesimal!

Analogously, let $\omega = \{1, 2, 3, \ldots\}$. For any $r \in \mathbb{R}$, the set

$$\{\!\!\{\mathbf{r} < \omega\}\!\!\} = \{j \in \mathbb{N} : r < j\}$$

is cofinite, because the reals are Archimedean. We have proved that ${}^*r < [\omega]$. Therefore, $[\omega]$ is infinitely large!

Remark 2.6. It is undesirable to discuss "infinitely large" and "infinitely small" numbers. These phrases are misleading because they suggest a connection between nonstandard numbers and the infinities which appear in other contexts. Hyperreals, however, have nothing to do with infinite cardinals, infinite sums, or sequences which diverge to infinity. Therefore, the terms *hyperfinite* and *unlimited* are preferable to "infinitely large." Likewise, *infinitesimal* is preferable to "infinitely small."

These facts demonstrate that $\mathbb{R} \subsetneq {}^*\mathbb{R}$. Here is an even more direct proof of this result. For any $r \in \mathbb{R}$, $\{\!\!\{\mathbf{r} = \omega\}\!\!\}$ equals $\emptyset$ or $\{r\}$. Thus $\{\!\!\{\mathbf{r} = \omega\}\!\!\} \notin \mathscr{F}$, which shows that ${}^*r \neq [\omega]$. Thus, $[\omega] \in {}^*\mathbb{R} \setminus \mathbb{R}$.

Definition 2.7 (Nonstandard Number). Any element of ${}^*\mathbb{R} \setminus \mathbb{R}$ is called a *nonstandard number*. For every $r \in \mathbb{R}$, ${}^*r$ is *standard*. In fact, *all* standard elements of ${}^*\mathbb{R}$ take this form.

This discussion also shows that any sequence $\varepsilon$ converging to zero generates an infinitesimal $[\varepsilon]$, which vindicates Cauchy's definition. Similarly, any sequence $\omega$ which diverges to infinity can be identified with an unlimited number $[\omega]$. Moreover, $[\varepsilon] \cdot [\omega] = [\mathbf{1}]$. So $[\varepsilon]$ and $[\omega]$ are multiplicative inverses.

Mission accomplished.

**2.2.8. The $*$ Map.** We would like to be able to extend functions from $\mathbb{R}$ to ${}^*\mathbb{R}$. As a first step, it is necessary to enlarge the function's domain.

Let $A \subseteq \mathbb{R}$. Define the extension or enlargement ${}^*A$ of $A$ as follows. For each $\mathbf{r} \in \mathbb{R}^{\mathbb{N}}$,

$$[\mathbf{r}] \in {}^*A \quad \text{iff} \quad \{\!\{\mathbf{r} \in A\}\!\} = \{j \in \mathbb{N} : r_j \in A\} \in \mathscr{F}.$$

That is, ${}^*A$ contains the equivalence classes of sequences whose terms are almost all in $A$. One consequence is that ${}^*a \in {}^*A$ for each $a \in A$.

Now, we prove a crucial theorem about set extensions.

Theorem 2.8. *Let $A \subseteq \mathbb{R}$. ${}^*A$ has nonstandard members if and only if $A$ is infinite. Otherwise, ${}^*A = A$.*

Proof. If $A$ is infinite, then there is a sequence $\mathbf{r}$, where $r_j \in A$ for each $j$, whose terms are all distinct. The set $\{\!\{\mathbf{r} \in A\}\!\} = \mathbb{N} \in \mathscr{F}$, so $[\mathbf{r}] \in {}^*A$. For any real $s \in A$, let $\mathbf{s} = \{s, s, s \ldots\}$. The agreement set $\{\!\{\mathbf{r} = \mathbf{s}\}\!\}$ is either $\emptyset$ or a singleton, neither of which is large. So ${}^*s = [\mathbf{s}] \neq [\mathbf{r}]$. Thus, $[\mathbf{r}]$ is a nonstandard element of ${}^*A$.

On the other hand, assume that $A$ is finite. Choose $[\mathbf{r}] \in {}^*A$. By definition, $\mathbf{r}$ has a large set of terms in $A$. For each $x \in A$, let

$R_x = \{\!\!\{\mathbf{r} = \mathbf{x}\}\!\!\} = \{j \in \mathbb{N} : r_j = x\}$. Now, $\{R_x\}_{x \in A}$ is a finite collection of pairwise disjoint sets, and their union is an element of $\mathscr{F}$, i.e. a large set. The properties of ultrafilters (see Appendix C) dictate that $R_x \in \mathscr{F}$ for exactly one $x \in A$, say $x_0$. Therefore, $\{\!\!\{\mathbf{r} = \mathbf{x_0}\}\!\!\} \in \mathscr{F}$, where $\mathbf{x_0} = \{x_0, x_0, x_0, \ldots\}$. And so $[\mathbf{r}] = {}^*x_0$.

As every element of $A$ has a corresponding element in ${}^*A$, we conclude that ${}^*A = A$ whenever $A$ is finite. □

The definition and theorem have several immediate consequences. ${}^*A$ will have infinitesimal elements at the accumulation points of $A$. In addition, the extension of an unbounded set will have infinitely large elements.

It should be noted that the $*$ map developed here is a special case of a nonstandard extension, described in Appendix A. Therefore, it preserves unions, intersections, set differences and Cartesian products.

Now, we are prepared to define the extension of a function, $f : \mathbb{R} \to \mathbb{R}$. For any sequence $\mathbf{r} \in \mathbb{R}^{\mathbb{N}}$, define $\mathbf{f}(\mathbf{r}) = \{f(r_j)\}$. Then let

$$^*f([\mathbf{r}]) = [\mathbf{f}(\mathbf{r})] \,.$$

In general,

$$\{\!\!\{\mathbf{r} = \mathbf{r}'\}\!\!\} \subseteq \{\!\!\{\mathbf{f}(\mathbf{r}) = \mathbf{f}(\mathbf{r}')\}\!\!\},$$

which means

$$\mathbf{r} \equiv \mathbf{r}' \quad \text{implies} \quad \mathbf{f}(\mathbf{r}) \equiv \mathbf{f}(\mathbf{r}')).$$

Thus, ${}^*f$ is well-defined. Now, ${}^*f : {}^*\mathbb{R} \to {}^*\mathbb{R}$.

We can also extend the partial function $f : A \to \mathbb{R}$ to the partial function ${}^*f : {}^*A \to {}^*\mathbb{R}$. This construction is identical to the last, except that we avoid elements outside $\operatorname{Dom} f$. For any $[\mathbf{r}] \in {}^*A$, let

$$s_j = \begin{cases} f(r_j) & \text{if } r_j \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Since $[\mathbf{r}] \in {}^*A$, $r_j \in A$ for almost all $j$, which means that $s_j = f(r_j)$ almost everywhere. Finally, we put

$$^*f([\mathbf{r}]) = [\mathbf{s}]\,.$$

Demonstrating well-definition of the extension of a partial function is similar to the proof for functions whose domain is $\mathbb{R}$.

It is easy to show that $^*(f(r)) = {}^*f({}^*r)$, so $^*f$ is an extension of $f$. Therefore, the $*$ is not really necessary, and it is sometimes omitted.

Definition 2.9 (Hypersequence). Note that this discussion also applies to sequences, since a sequence is a function $\mathbf{a} : \mathbb{N} \to \mathbb{R}$. The extension of a sequence is called a *hypersequence*, and it maps $^*\mathbb{N} \to {}^*\mathbb{R}$. The same symbol $\mathbf{a}$ is used to denote the hypersequence. Terms with hyperfinite indices are called *extended* terms.

Definition 2.10 (Standard Object). Any set of hyperreals, function on the hyperreals, or sequence of hyperreals which can be defined via this $*$ mapping is called *standard*.

## 2.3. Principles of NSA

Before we can exploit the power of NSA, we need a way to translate results from the reals to the hyperreals and vice-versa. I continue to follow Goldblatt's presentation [**7**].

**2.3.1. The Transfer Principle.** The Transfer Principle is the most important tool in Nonstandard Analysis. First, it allows us to recast classical theorems for the hyperreals. Second, it permits the use of hyperreals to prove results about the reals. Roughly, transfer states that

> any appropriately formulated statement is true of $^*\mathbb{R}$
> if and only if it is true of $\mathbb{R}$ [**7**, 11].

We must define what it means for a statement to be "appropriately formulated" and how the statement about $^*\mathbb{R}$ differs from the statement about $\mathbb{R}$.

Any mathematical statement can be written in logical notation using the following symbols:

> **Logical Connectives:** $\wedge$ (and), $\vee$ (or), $\neg$ (not), $\rightarrow$ (implies), and $\leftrightarrow$ (if and only if).
>
> **Quantifiers:** $\forall$ (for all) and $\exists$ (there exists).
>
> **Parentheses:** $()$, $[]$.
>
> **Constants:** Fixed elements of some fixed set or universe $\mathscr{U}$, which are usually denoted by letter symbols.
>
> **Variables:** A countable collection of letter symbols.

Definition 2.11 (Sentence). A *sentence* is a mathematical statement written in logical notation and which contains no *free* variables. In other words, every variable must be quantified to specify its *bound*, the set over which it ranges. For example, the statement $(x > 2)$ contains a free occurence of the variable $x$. On the other hand, the statement $(\forall y \in \mathbb{N})(y > 2)$ contains only the variable $y$, bound to $\mathbb{N}$, which means that it is a sentence. A sentence in which all terms are defined may be assigned a definite truth value.

Next, we explain how to take the $*$-transform of a sentence $\varphi$. This is a further generalization of the $*$ map which was discussed in Section 2.2.8.

- Replace each constant $\tau$ by $^*\tau$.
- Replace each relation (or function) $\mathscr{R}$ by $^*\mathscr{R}$.
- Replace the bound $A$ of each quantifier by its enlargement $^*A$.

Variables do not need to be renamed. Set operations like $\cup, \cap, \setminus, \times$, etc. are preserved under the $*$ map, so they do not need renaming. As we saw before, we may identify $r$ with ${}^*r$ for any real number, so these constants do not require a $*$. It is also common to omit the $*$ from standard relations like $=, \neq, <, \in$, etc. and from standard functions like $\sin, \cos, \log, \exp$, etc. The classical definition will dictate the $*$-transform. As before, $A \subsetneqq {}^*A$ whenever $A$ is infinite. Therefore, all sets must be replaced by their enlargements.

Be careful, however, when using sets as variables. The bound of a variable is the set over which it ranges, hence $(\forall A \subseteq \mathbb{R})$ must be written as $(\forall A \in \mathscr{P}(\mathbb{R}))$. Furthermore, the transform of $\mathscr{P}(\mathbb{R})$ is ${}^*\mathscr{P}(\mathbb{R})$ and *neither* $\mathscr{P}({}^*\mathbb{R})$ *nor* ${}^*\mathscr{P}({}^*\mathbb{R})$. This phenomenon results from the fact that $\mathscr{P}$ is not a function; it is a special notation for a specific set.

It will be helpful to provide some examples of sentences and their $*$-transforms.

$$(\forall x \in \mathbb{R})(\sin^2 x + \cos^2 x = 1) \quad \text{becomes}$$
$$(\forall x \in {}^*\mathbb{R})(\sin^2 x + \cos^2 x = 1).$$

$$(\forall x \in \mathbb{R})(x \in [a,b] \leftrightarrow a \leq x \leq b) \quad \text{becomes}$$
$$(\forall x \in {}^*\mathbb{R})(x \in {}^*[a,b] \leftrightarrow a \leq x \leq b).$$

$$(\exists y \in [a,b])(\pi < f(y)) \quad \text{becomes}$$
$$(\exists y \in {}^*[a,b](\pi < {}^*f(y)).$$

Now, we can restate the transfer principle more formally. If $\varphi$ is a sentence and ${}^*\varphi$ is its $*$-transform,

$${}^*\varphi \text{ is true} \quad \text{iff} \quad \varphi \text{ is true}.$$

The transfer principle is a special case of Łoś's Theorem, which is beyond the scope of this thesis.

As a result of transfer, many facts about real numbers are also true about the hyperreals. Trigonometric functions and logarithms, for instance, continue to behave the same way for hyperreal arguments.

Transfer also permits the use of infinitesimals and unlimited numbers in lieu of limit arguments (see Section 3.1).

One more caution about the transfer principle: although every sentence concerning $\mathbb{R}$ has a $*$-transform, there are many sentences concerning ${}^*\mathbb{R}$ which are *not* $*$-transforms.

The rules for applying the $*$-transform may seem arcane, but they quickly become second nature. The proofs in the next chapter will foster familiarity.

**2.3.2. Internal Sets.** For any sequence of subsets of $\mathbb{R}$, $\mathbf{A} = \{A_j\}$, define a subset $[\mathbf{A}] \subseteq {}^*\mathbb{R}$ by the following rule. For each $[\mathbf{r}] \in {}^*\mathbb{R}$,

$$[\mathbf{r}] \in [\mathbf{A}] \quad \text{iff} \quad \{\!\!\{\mathbf{r} \in A\}\!\!\} = \{j \in \mathbb{N} : r_j \in A_j\} \in \mathscr{F}.$$

Subsets of ${}^*\mathbb{R}$ formed in this manner are called *internal*.

As examples, the enlargement ${}^*A$ of $A \subseteq \mathbb{R}$ is internal, since it is constructed from the constant sequence $\{A, A, A, \ldots\}$. Any finite set of hyperreals is internal, and the hyperreal interval, $[a, b] = \{x \in {}^*\mathbb{R} : a \leq x \leq b\}$, is internal for any $a, b \in {}^*\mathbb{R}$.

Internal sets may also be identified as the elements of ${}^*\mathscr{P}(\mathbb{R})$. Thus the transfer principle gives internal sets a special status. For example, the sentence

$$(\forall A \in \mathscr{P}(\mathbb{N}))[(A \neq \emptyset) \to (\exists n \in \mathbb{N})(n = \min A)] \quad \text{becomes}$$
$$(\forall A \in {}^*\mathscr{P}(\mathbb{N}))[(A \neq \emptyset) \to (\exists n \in {}^*\mathbb{N})(n = \min A)].$$

Therefore, every nonempty internal subset of ${}^*\mathbb{N}$ has a least member.

Internal sets have many other fascinating properties, which are fundamental to NSA. It is also possible to define *internal functions* as the equivalence classes of sequences of real-valued functions. These, too, are crucial to NSA. Unfortunately, an explication of these facts would take us too far afield.

## 2.4. Working with Hyperreals

Having discussed some of the basic principles of NSA, we can begin to investigate the structure of the hyperreals. Then, we will be able to ignore the details of the ultrapower construction and use hyperreals for arithmetic. I am still following Goldblatt [**7**].

**2.4.1. Types of Hyperreals.** ${}^*\mathbb{R}$ contains the *hyperreal* numbers. Similarly, ${}^*\mathbb{Q}$ contains *hyperrationals*, ${}^*\mathbb{Z}$ contains *hyperintegers* and ${}^*\mathbb{N}$ contains *hypernaturals*. The sentence

$$(\forall x \in \mathbb{R})[(x \in \mathbb{Q}) \leftrightarrow (\exists y, z \in \mathbb{Z})(z \neq 0 \wedge x = y/z)]$$

transfers to

$$(\forall x \in {}^*\mathbb{R})[(x \in {}^*\mathbb{Q}) \leftrightarrow (\exists y, z \in {}^*\mathbb{Z})(z \neq 0 \wedge x = y/z)],$$

which demonstrates that ${}^*\mathbb{Q}$ contains quotients of hyperintegers.

Another important set of hyperreals is the set of unlimited natural numbers, ${}^*\mathbb{N}_\infty = {}^*\mathbb{N} \setminus \mathbb{N}$. One of its key properties is that it has no least member.[9]

Hyperreal numbers come in several basic sizes. Terminology varies, but Goldblatt lists the most common definitions. The hyperreal $b \in {}^*\mathbb{R}$ is

- *limited* if $r < b < s$ for some $r, s \in \mathbb{R}$;
- *positive unlimited* if $b > r$ for every $r \in \mathbb{R}$;
- *negative unlimited* if $b < r$ for every $r \in \mathbb{R}$;
- *unlimited* or *hyperfinite* if it is positive or negative unlimited;
- *positive infinitesimal* if $0 < b < r$ for every positive $r \in \mathbb{R}$;
- *negative infinitesimal* if $r < b < 0$ for every negative $r \in \mathbb{R}$;
- *infinitesimal* if it is positive or negative infinitesimal or zero;[10]
- *appreciable* if $b$ is limited but not infinitesimal.

[9] Consequently, ${}^*\mathbb{N}_\infty$ is not internal.
[10] Zero is the only infinitesimal in $\mathbb{R}$.

Goldblatt also lists rules for arithmetic with hyperreals, although they are fairly intuitive. These laws follow from transfer of appropriate sentences about $\mathbb{R}$. Let $\varepsilon, \delta$ be infinitesimal, $b, c$ appreciable, and $N, M$ unlimited.

**Sums:** $\varepsilon + \delta$ is infinitesimal;
$b + \varepsilon$ is appreciable;
$b + c$ is limited (possibly infinitesimal);
$N + \varepsilon$ and $N + b$ are unlimited.

**Products:** $\varepsilon \cdot \delta$ and $\varepsilon \cdot b$ are infinitesimal;
$b \cdot c$ is appreciable;
$b \cdot N$ and $N \cdot M$ are unlimited.

**Reciprocals:** $\frac{1}{\varepsilon}$ is unlimited if $\varepsilon \neq 0$;
$\frac{1}{b}$ is appreciable;
$\frac{1}{N}$ is infinitesimal.

**Roots:** For $n \in \mathbb{N}$,
if $\varepsilon > 0$, $\sqrt[n]{\varepsilon}$ is infinitesimal;
if $b > 0$, $\sqrt[n]{b}$ is appreciable;
if $N > 0$, $\sqrt[n]{N}$ is unlimited.

**Indeterminate Forms:** $\frac{\varepsilon}{\delta}, \frac{N}{M}, \varepsilon \cdot N, N + M$.

Other rules follow easily from transfer coupled with common sense. On an algebraic note, these rules show that the set of limited numbers $\mathbb{L}$ and the set of infinitesimals $\mathbb{I}$ both form subrings of ${}^*\mathbb{R}$. $\mathbb{I}$ forms an ideal in $\mathbb{L}$, and it can be shown that the quotient $\mathbb{L}/\mathbb{I} = \mathbb{R}$.

**2.4.2. Halos and Galaxies.** The rich structure of the hyperreals suggests several useful new types of relations. The most important cases are when two hyperreals are infinitely near to each other and when they are a limited distance apart.

Definition 2.12 (Infinitely Near). Two hyperreals $b$ and $c$ are *infinitely near* when $b - c$ is infinitesimal. We denote this relationship by $b \simeq c$. This defines an equivalence relation on ${}^*\mathbb{R}$ whose equivalence classes are written

$$\mathrm{hal}(b) = \{c \in {}^*\mathbb{R} : b \simeq c\}.$$

Definition 2.13 (Limited Distance Apart). Two hyperreals $b$ and $c$ are at a *limited distance* when $b - c$ is appreciable. We denote this relationship by $b \sim c$. This also defines an equivalence relation on ${}^*\mathbb{R}$ whose equivalence classes are written

$$\mathrm{gal}(b) = \{c \in {}^*\mathbb{R} : b \sim c\}.$$

It is clear then that $b$ is infinitesimal if and only if $b \simeq 0$. Likewise, $b$ is limited if and only if $b \sim 0$. Equivalently, $\mathbb{I} = \mathrm{hal}(0)$ and $\mathbb{L} = \mathrm{gal}(0)$. This notation derives from the words "halo" and "galaxy," which illustrate the concepts well.

At this point, we can get some idea of how big the set of hyperreals is. Choose a positive unlimited number $N$. It is easy to see that $\mathrm{gal}(N)$ is disjoint from $\mathrm{gal}(2N)$. In fact, $\mathrm{gal}(N)$ does not intersect $\mathrm{gal}(nN)$ for any integer $n$. Furthermore, $\mathrm{gal}(N)$ is disjoint from $\mathrm{gal}(N/2)$, $\mathrm{gal}(N/3)$, etc. Moreover, none of these sets intersect $\mathrm{gal}(N^2)$ or the galaxy of any hypernatural power of $N$. The elements of $\mathrm{gal}(e^N)$ dwarf these numbers. Yet the elements of $\mathrm{gal}(N^N)$ are still greater.

Since the reciprocal of every unlimited number is an infinitesimal, we see that there are an infinite number of shells of infinitesimals surrounding zero, each of which has the same cardinality as a galaxy. Every real number has a halo of infinitesimals around it, and every galaxy contains a copy of the real line along with the infinitesimal halos of each element. Fleas on top of fleas.[11]

[11] More precisely, $|{}^*\mathbb{R}| = |\mathscr{P}(\mathbb{R})| = 2^c$, where $c$ is the cardinality of the real line. Therefore, the hyperreals have the same power as the set of functions on $\mathbb{R}$.

**2.4.3. Shadows.** Finally, we will discuss the shadow map which takes a limited hyperreal to its nearest real number.

THEOREM 2.14 (Unique Shadow). *Every limited hyperreal $b$ is infinitely close to exactly one real number, which is called its* shadow *and written* $\mathrm{sh}\,(b)$.

PROOF. Let $A = \{r \in \mathbb{R} : r < b\}$.

First, we find a candidate shadow. Since $b$ is limited, $A$ is nonempty and bounded above. $\mathbb{R}$ is complete, so $A$ has a least upper bound $c \in \mathbb{R}$.

Next, we show that $b \simeq c$. For any positive, real $\varepsilon$, the quantity $c+\varepsilon \notin A$, since $c$ is the least upper bound of $A$. Similarly, $c-\varepsilon < b$, or else $c - \varepsilon$ would be a smaller upper bound of $A$. So $c - \varepsilon < b \leq c + \varepsilon$, and $|b - c| \leq \varepsilon$. Since $\varepsilon$ is arbitrarily small, we must have $b \simeq c$.

Finally, uniqueness. If $b \simeq c' \in \mathbb{R}$, then $c \simeq c'$ by transitivity. The quantities $c$ and $c'$ are both real, so $c = c'$. $\square$

The shadow map preserves all the standard rules of arithmetic.

THEOREM 2.15. *If $b, c$ are limited and $n \in \mathbb{N}$, we have*

(1) $\mathrm{sh}\,(b \pm c) = \mathrm{sh}\,(b) \pm \mathrm{sh}\,(c)$;
(2) $\mathrm{sh}\,(b \cdot c) = \mathrm{sh}\,(b) \cdot \mathrm{sh}\,(c)$;
(3) $\mathrm{sh}\,(b/c) = \mathrm{sh}\,(b)\,/\,\mathrm{sh}\,(c)$, *provided that* $\mathrm{sh}\,(c) \neq 0$;
(4) $\mathrm{sh}\,(b^n) = (\mathrm{sh}\,(b))^n$;
(5) $\mathrm{sh}\,(|b|) = |\,\mathrm{sh}\,(b)\,|$;
(6) $\mathrm{sh}\left(\sqrt[n]{b}\right) = \sqrt[n]{\mathrm{sh}\,(b)}$ *if* $b \geq 0$; *and*
(7) *if* $b \leq c$ *then* $\mathrm{sh}\,(b) \leq \mathrm{sh}\,(c)$.

PROOF. I will prove 1 and 7; the other proofs are similar.

Let $\varepsilon = b - \mathrm{sh}\,(b)$ and $\delta = c - \mathrm{sh}\,(c)$. The shadows are infinitely near $b$ and $c$, so $\varepsilon$ and $\delta$ are infinitesimal. Then,

$$b + c = \mathrm{sh}\,(b) + \mathrm{sh}\,(c) + \varepsilon + \delta \simeq \mathrm{sh}\,(b) + \mathrm{sh}\,(c)\,.$$

Hence, $\mathrm{sh}\,(b+c) = \mathrm{sh}\,(b) + \mathrm{sh}\,(c)$. The proof for differences is identical.

Assume that $b \leq c$. If $b \simeq c$, then $\mathrm{sh}\,(b) \simeq c$. Thus, $\mathrm{sh}\,(b) = \mathrm{sh}\,(c)$. Otherwise, $b \not\simeq c$, so we have $c = b + \varepsilon$ for some positive, appreciable $\varepsilon$. Then, $\mathrm{sh}\,(c) = \mathrm{sh}\,(b) + \mathrm{sh}\,(\varepsilon)$, or $\mathrm{sh}\,(c) - \mathrm{sh}\,(b) = \mathrm{sh}\,(\varepsilon) > 0$. We conclude that $\mathrm{sh}\,(b) \leq \mathrm{sh}\,(c)$. □

REMARK 2.16. The shadow map does not preserve strict inequalities. If $b < c$ and $b \simeq c$, then $\mathrm{sh}\,(b) = \mathrm{sh}\,(c)$.

CHAPTER 3

# Straightforward Analysis

Finally, we will use the machinery of Nonstandard Analysis to develop some of the basic theorems of real analysis in an intuitive manner. In this chapter, I have drawn on Goldblatt [**7**], Rudin [**12**], Cutland [**5**] and Robert [**11**].

REMARK 3.1. Many of the proofs depend on whether a variable is real or hyperreal. Read carefully!

## 3.1. Sequences and Their Limits

The limit concept is the foundation of all classical analysis. NSA replaces limits with reasoning about infinite nearness, which reduces many complicated arguments to simple hyperreal arithmetic. First, we review the classical definition of a limit.

DEFINITION 3.2 (Limit of a Sequence). Let $\mathbf{a} = \{a_j\}_{j=1}^{\infty}$ be a real-valued sequence. Say that, for every real $\varepsilon > 0$, there exists $J(\varepsilon) \in \mathbb{N}$ such that $j > J$ implies $|a_j - L| < \varepsilon$. Then $L$ is the *limit* of the sequence $\mathbf{a}$. We also say that $\mathbf{a}$ converges to $L$ and write $a_j \to L$.

This definition is an awkward rephrasing of a simple concept. A sequence has a limit only if its terms get very close to that limit and stay there. NSA allows us to apply this idea more directly.

THEOREM 3.3. *Let* $\mathbf{a}$ *be a real-valued sequence. The following are equivalent:*

(1) $\mathbf{a}$ *converges to* $L$

(2) *$a_j \simeq L$ for every unlimited $j$.*

Proof. Assume that $a_j \to L$, and fix an unlimited $N$. For any positive, real $\varepsilon$, there exists $J(\varepsilon) \in \mathbb{N}$ such that

$$(\forall j \in \mathbb{N})(j > J \to |a_j - L| < \varepsilon).$$

By transfer,

$$(\forall j \in {}^*\mathbb{N})(j > J \to |a_j - L| < \varepsilon).$$

Since $N$ is unlimited, it exceeds $J$. Therefore, $|a_N - L| < \varepsilon$ for any positive, real $\varepsilon$, which means $|a_N - L|$ is infinitesimal, or equivalently $a_N \simeq L$.

Conversely, assume $a_j \simeq L$ for every unlimited $j$, and fix a real $\varepsilon > 0$. For unlimited $N$, any $j > N$ is also unlimited. So we have

$$(\forall j \in {}^*\mathbb{N})(j > N \to a_j \simeq L),$$

which implies

$$(\forall j \in {}^*\mathbb{N})(j > N \to |a_j - L| < \varepsilon).$$

Equivalently,

$$(\exists N \in {}^*\mathbb{N})(\forall j \in {}^*\mathbb{N})(j > N \to |a_j - L| < \varepsilon).$$

By transfer, this statement is true only if

$$(\exists N \in \mathbb{N})(\forall j \in \mathbb{N})(j > N \to |a_j - L| < \varepsilon)$$

is true. Since $\varepsilon$ was arbitrary, $a_j \to L$. □

As a consequence of this theorem and the Unique Shadow theorem, a convergent sequence can have only one limit.

### 3.1.1. Bounded Sequences.

Definition 3.4 (Bounded Sequence). A real-valued sequence **a** is *bounded* if there exists an integer $n$ such that $a_j \in [-n, n]$ for every index $j \in \mathbb{N}$. Otherwise, **a** is *unbounded*.

THEOREM 3.5. *A sequence is bounded if and only if its extended terms are limited.*

PROOF. Let $\mathbf{a}$ be bounded. Then, there exists $n \in \mathbb{N}$ such that $a_j \in [-n, n]$ for every $j \in \mathbb{N}$. Therefore, when $N$ is unlimited, $a_N \in {}^*[-n, n] = \{x \in {}^*\mathbb{R} : -n \leq x \leq n\}$. Hence $a_N$ is limited.

Conversely, let $a_j$ be limited for every unlimited $j$. Fix a hyperfinite $N \in {}^*\mathbb{N}$. Clearly, $a_j \in [-N, N]$. So

$$(\exists N \in {}^*\mathbb{N})(\forall j \in {}^*\mathbb{N})(-N \leq a_j \leq N).$$

Then, there must exist $n \in \mathbb{N}$ such that $-n \leq a_j \leq n$ for any standard term $a_j$. Therefore, the sequence is bounded. □

DEFINITION 3.6 (Monotonic Sequence). The sequence $\mathbf{a}$ *increases monotonically* if $a_j \leq a_{j+1}$ for each $j$. If $a_j \geq a_{j+1}$ for each $j$, then $\mathbf{a}$ *decreases monotonically.*

THEOREM 3.7. *Bounded, monotonic sequences converge.*

PROOF. Let $\mathbf{a}$ be a bounded, monotonically increasing sequence. Fix an unlimited $N$. Since $\mathbf{a}$ is bounded, $a_N$ is limited. Put $L = \mathrm{sh}\,(a_N)$. Now, $\mathbf{a}$ is nondecreasing, so $j \leq k$ implies $a_j \leq a_k$. In particular, $a_j \leq a_N \simeq L$ for every limited $j$. Thus, $L$ is an upper bound of the standard part of $\mathbf{a} = \{a_j : j \in \mathbb{N}\}$.

In fact, $L$ is the least upper bound of this set. If $r$ is any real upper bound of the limited terms of $\mathbf{a}$, it is also an upper bound the extended terms. The relation $L \simeq a_N \leq r$ implies that $L \leq r$.

Therefore, $a_j \simeq L$ for every unlimited $j$, and $a_j \to L$.

The proof for monotonically decreasing sequences is similar. □

REMARK 3.8. This result can be used to show that $\lim_{j\to\infty} c^j = 0$ for any real $c \in [0, 1)$. First, notice that $\{c^j\}$ is nonincreasing and that it is bounded below by 0. Thus, it has a real limit $L$. For unlimited $N$,

$$L \simeq c^{N+1} = c \cdot c^N \simeq c \cdot L.$$

Both $c$ and $L$ are real, so $L = c \cdot L$. But $c \neq 1$, so $L = 0$.

**3.1.2. Cauchy Sequences.** Next, we will develop the nonstandard characterization of a Cauchy sequence.

THEOREM 3.9. *A real-valued sequence is Cauchy if and only if all its extended terms are infinitely close to each other, i.e.* $a_j \simeq a_k$ *for all unlimited* $j, k$.

PROOF. Assume that the real-valued sequence $\mathbf{a}$ is Cauchy:

$$(\forall \varepsilon \in \mathbb{R}^+)(\exists J \in \mathbb{N})(j, k > J \to |a_j - a_k| < \varepsilon).$$

Fix an $\varepsilon > 0$, which dictates $J(\varepsilon)$. Then,

$$(\forall j \in \mathbb{N})(\forall k \in \mathbb{N})(j, k > J \to |a_j - a_k| < \varepsilon).$$

By transfer,

$$(\forall j \in {}^*\mathbb{N})(\forall k \in {}^*\mathbb{N})(j, k > J \to |a_j - a_k| < \varepsilon).$$

All unlimited $j, k$ exceed $J$, which means that $|a_j - a_k| < \varepsilon$ for any epsilon. Thus, $a_j \simeq a_k$ whenever $j$ and $k$ are unlimited.

Now, assume that $a_j \simeq a_k$ for all unlimited $j, k$, and choose a real $\varepsilon > 0$. For unlimited $N$, any $j$ and $k$ exceeding $N$ are also unlimited. Then,

$$(\exists N \in {}^*\mathbb{N})(\forall j, k \in {}^*N)(j, k > N \to |a_j - a_k| < \varepsilon).$$

By transfer,

$$(\exists N \in \mathbb{N})(\forall j, k \in N)(j, k > N \to |a_j - a_k| < \varepsilon).$$

Since $\varepsilon$ was arbitrary, $\mathbf{a}$ is Cauchy. $\square$

This theorem suggests that a Cauchy sequence should not diverge, since its extended terms would have to keep growing. In fact, we can show that every Cauchy sequence of real numbers converges, and conversely. This property of the real numbers is called *completeness*, and it is equivalent to the least-upper-bound property, which is used to prove the Unique Shadow theorem. Before proving this theorem, we require a classical lemma.

LEMMA 3.10. *Every Cauchy sequence is bounded.*

PROOF. Let $\mathbf{a}$ be Cauchy. Pick a real $\varepsilon > 0$. There exists $J(\varepsilon)$ beyond which $|a_j - a_k| < \varepsilon$. In particular, for each $j \geq J$, $a_j$ is within $\varepsilon$ of $a_J$. Now, the set $E = \{a_j : j \leq J\}$ is finite, so we can put $m = \min E$ and $M = \max E$. Of course, $a_J \in [m, M]$. Thus every term of the sequence must be contained in the open interval $(m - \varepsilon, M + \varepsilon)$. As a result, $\mathbf{a}$ is bounded. □

THEOREM 3.11. *A real-valued sequence converges if and only if it is Cauchy.*

PROOF. Let $a_N$ be an extended term of the Cauchy sequence $\mathbf{a}$. By the lemma, $\mathbf{a}$ is bounded, hence $a_N$ is limited. Put $L = \operatorname{sh}(a_N)$. Since $\mathbf{a}$ is Cauchy, $a_j \simeq a_N \simeq L$ for every unlimited $j$. By Theorem 3.3, $a_j \to L$.

Next, assume that the real-valued sequence $a_j \to L$. For every unlimited $j$ and $k$, we have $a_j \simeq L \simeq a_k$. Therefore, $a_j \simeq a_k$, and $\mathbf{a}$ is Cauchy. □

**3.1.3. Accumulation Points.** If a real sequence does not converge, there are several other possibilities. The sequence may have multiple accumulation points; it may diverge to infinity; or it may have no limit whatsoever.

Definition 3.12 (Accumulation Point). A real number $L$ is called an *accumulation point* or a *cluster point* of the set $E$ if there are an infinite number of elements of $E$ within every $\varepsilon$-neighborhood of $L$, $(L-\varepsilon, L+\varepsilon)$, where $\varepsilon$ is a real number.

Theorem 3.13. *A real number $L$ is an accumulation point of the sequence* $\mathbf{a}$ *if and only if the sequence has an extended term infinitely near $L$. That is, $a_j \simeq L$ for some unlimited $j$.*

Proof. Assume that $L$ is a cluster point of $\mathbf{a}$. The logical equivalent of this statement is

$$(\forall \varepsilon \in \mathbb{R}^+)(\forall J \in \mathbb{N})(\exists j \in \mathbb{N})(j > J \wedge |a_j - L| < \varepsilon).$$

Fix a positive infinitesimal $\varepsilon$ and an unlimited $J$. By transfer, there exists an (unlimited) $j > J$ for which $|a_j - L| < \varepsilon \simeq 0$. So $a_j \simeq L$.

Next, let $a_j \simeq L$ for some unlimited $j$. Take $\varepsilon \in \mathbb{R}^+$ and $J \in \mathbb{N}$. Then $j > J$ and $|a_j - L| < \varepsilon$. Thus,

$$(\exists j \in {}^*\mathbb{N})(j > J \wedge |a_j - L| < \varepsilon).$$

Transfer demonstrates that $L$ is a cluster point of $\mathbf{a}$. $\square$

In other words, if $a_N$ is a hyperfinite term of a sequence, its shadow is an accumulation point of the sequence. This result yields a direct proof of the Bolzano-Weierstrass theorem.

Theorem 3.14 (Bolzano-Weierstrass). *Every bounded, infinite set has an accumulation point.*

Proof. Let $E$ be a bounded, infinite set. Since $E$ is infinite, we can choose a sequence $\mathbf{a}$ from $E$. Since $\mathbf{a}$ is bounded, all of its extended terms are limited, which means that each has a shadow. Each distinct shadow is a cluster point of the sequence, so $\mathbf{a}$ must have at least one accumulation point, which is simultaneously an accumulation point of the set $E$. $\square$

**3.1.4. Divergent Sequences.** Unbounded sequences do not need to have *any* accumulation points. One example is the sequence which diverges.

Definition 3.15 (Divergent Sequence). Let $\mathbf{a}$ be a real-valued sequence. We say the sequence *diverges to infinity* if, for any $n \in \mathbb{N}$, there exists $J(n)$ such that $j > J$ implies $a_j > n$. If, for any $n$, there exists $J(n)$ such that $j > J$ implies $a_j < -n$, then $\mathbf{a}$ *diverges to minus infinity.*

Theorem 3.16. *A real-valued sequence diverges to infinity if and only if all of its extended terms are positive unlimited. Likewise, it diverges to minus infinity if and only if each of its extended terms is negative unlimited.*

Proof. Let $\mathbf{a}$ be a divergent sequence. Fix an unlimited number $N$. For any $n \in \mathbb{N}$, there exists a $J$ in $\mathbb{N}$ such that

$$(\forall j \in \mathbb{N})(j > J \rightarrow a_j > n).$$

Since $N > J$, $a_N > n$. The integer $n$ was arbitrary, so $a_N$ must be unlimited.

Now, assume that $a_j$ is positive unlimited for every unlimited $j$, and choose an unlimited $J$. We have

$$(\exists J \in {}^*\mathbb{N})(\forall j \in {}^*\mathbb{N})(j > J \rightarrow a_j > n).$$

Transfer shows that $\mathbf{a}$ diverges to infinity.

The second part is almost identical. $\square$

**3.1.5. Superior and Inferior Limits.** Finally, we will define superior and inferior limits. Let $\mathbf{a}$ be a bounded sequence. Put $E =$

$\{\operatorname{sh}(a_j) : j \in {}^*\mathbb{N}_\infty\}$. We put

$$\limsup_{j\to\infty} a_j = \overline{\lim_{j\to\infty}}\, a_j = \sup E, \text{ and}$$
$$\liminf_{j\to\infty} a_j = \underline{\lim_{j\to\infty}}\, a_j = \inf E.$$

In other words, $\limsup_{j\to\infty} a_j$ is the supremum of the sequence's accumulation points, and $\liminf_{j\to\infty} a_j$ is the infimum of the accumulation points.

For unbounded sequences, there is a complication, since the set $E$ cannot be defined as before. When $\mathbf{a}$ is unbounded, put $E = \{\operatorname{sh}(a_j) : j \in {}^*\mathbb{N}_\infty \text{ and } a_j \in \mathbb{L}\}$. If $\mathbf{a}$ has no upper bound, then $\limsup_{j\to\infty} a_j = +\infty$. Similarly, if $\mathbf{a}$ has no lower bound, then $\liminf_{j\to\infty} a_j = -\infty$. Otherwise,

$$\limsup_{j\to\infty} a_j = \sup E, \text{ and}$$
$$\liminf_{j\to\infty} a_j = \inf E.$$

Some sequences, such as $\{(-2)^j\}$ neither converge nor diverge. Yet every sequence has superior and inferior limits, in this case $+\infty$ and $-\infty$.

Remark 3.17. Many results about real-valued sequences may be extended to complex-valued sequences by using transfer.

### 3.2. Series

Let $\mathbf{a} = \{a_j\}_{j=1}^{\infty}$ be a sequence. A *series* is a sequence $\mathbf{S}$ of partial sums,

$$S_n = \sum_{j=1}^{n} a_j = a_1 + a_2 + \cdots + a_n.$$

For $n \geq m$, it is common to denote $a_m + a_{m+1} + \cdots + a_n$ by

$$\sum_{j=m}^{n} a_j = \sum_{j=1}^{n} a_j - \sum_{j=1}^{m-1} a_j = S_n - S_{m-1}.$$

It is also common to drop the index from the sum if there is no chance of confusion.

If the sequence $\mathbf{S}$ converges to $L$, then we say that the series converges to $L$ and write

$$\sum_{1}^{\infty} a_j = L.$$

Extending $\mathbf{S}$ to a hypersequence yields a *hyperseries*. In this context, the summation of an unlimited number of terms of $\mathbf{a}$ becomes meaningful. The extended terms of $\mathbf{S}$ may be thought of as *hyperfinite sums*.

A series is just a special type of sequence, hence all the results for sequences apply. Notably,

THEOREM 3.18. *$\sum_1^\infty a_j = L$ if and only if $\sum_1^N a_j \simeq L$ for all unlimited $N$.* $\square$

THEOREM 3.19. *$\sum_1^\infty a_j$ converges if any only if $\sum_M^N a_j \simeq 0$ for all unlimited $M, N$ with $N \geq M$. In particular, the series $\sum_1^\infty a_j$ converges only if* $\lim_{j\to\infty} a_j = 0$. $\square$

It is crucial to remember that the converse of this last statement is *not* true. The fact that $\lim_{j\to\infty} a_j = 0$ does not imply the convergence of $\sum_1^\infty a_j$. For example, the series

$$\sum_{1}^{\infty} \frac{1}{j}$$

diverges. To see this, group the terms as follows:

$$\begin{aligned}\sum_{1}^{\infty} \frac{1}{j} &= 1 + \tfrac{1}{2} + (\tfrac{1}{3} + \tfrac{1}{4}) + (\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}) + \cdots \\ &> 1 + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots \\ &= +\infty.\end{aligned}$$

**3.2.1. The Geometric Series.** Now, we examine a fundamental type of series.

Definition 3.20 (Geometric Series). A sum of the form

$$\sum_{m}^{n} r^j = r^m + r^{m+1} + \cdots + r^n$$

is called a *geometric series.*

Theorem 3.21. *In general,*

$$\sum_{m}^{n} r^j = r^m \frac{1 - r^{n-m+1}}{1 - r}.$$

*Furthermore, if* $|r| < 1$*, the geometric series converges, and*

$$\sum_{1}^{\infty} r^j = \frac{r}{1 - r}.$$

Proof. Let $m, n$ be positive integers with $n \geq m$. Put

$$S = \sum_{m}^{n} r^j.$$

Then

$$rS = \sum_{m}^{n} r^{j+1} = \sum_{m+1}^{n+1} r^j.$$

Hence,

$$S - rS = r^m - r^{n+1}.$$

Simplifying, we obtain

$$S = r^m \frac{1 - r^{n-m+1}}{1 - r}.$$

Put $m = 1$. In this case,

$$\sum_{1}^{n} r^j = r \frac{1 - r^n}{1 - r}.$$

If we take $|r| < 1$, $r^N \simeq 0$ for every unlimited $N$. Thus

$$\sum_{1}^{N} r^j \simeq \frac{r}{1 - r} \in \mathbb{R}.$$

We conclude that

$$\sum_{1}^{\infty} r^j = \frac{r}{1 - r}.$$

$\square$

**3.2.2. Convergence Tests.** There are many tests to determine whether a given series converges. One of the most commonly used is the comparison test.

THEOREM 3.22 (Nonstandard Comparison Test). *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *and* $\mathbf{d}$ *be sequences of nonnegative real terms.*

*If* $\sum_1^\infty b_j$ *converges and* $a_j \le b_j$ *for all unlimited* $j$*, then* $\sum_1^\infty a_j$ *converges.*

*If, on the other hand,* $\sum_1^\infty d_j$ *diverges and* $c_j \ge d_j$ *for all unlimited* $j$*, then* $\sum_1^\infty c_j$ *diverges.*

PROOF. For limited $m, n$ with $n \ge m$,

$$0 \le \sum_m^n a_j \le \sum_m^n b_j$$

if $0 \le a_j \le b_j$ for all $m \le j \le n$. Therefore, the same relationship holds for unlimited $m, n$ when $0 \le a_j \le b_j$ for all unlimited $j$. Fix $M, N \in {}^*\mathbb{N}_\infty$ with $N \ge M$. Since $\sum_1^\infty b_j$ converges,

$$0 \le \sum_M^N a_j \le \sum_M^N b_j \simeq 0.$$

Hence $\sum_M^N a_j \simeq 0$, which implies that $\sum_1^\infty a_j$ converges.

Similar reasoning yields the second part of the theorem. □

Leibniz discovered a convergence test for alternating series. For historical interest, here is a nonstandard proof.

DEFINITION 3.23 (Alternating Series). If $a_j \le 0$ implies $a_{j+1} \ge 0$ and $a_j \ge 0$ implies $a_{j+1} \le 0$ then the series $\sum a_j$ is called an *alternating series.*

THEOREM 3.24 (Alternating Series Test). *Let* $\mathbf{a}$ *be a sequence of positive terms which decrease monotonically, with* $\lim_{j\to\infty} a_j = 0$*.*

$$\sum_1^\infty (-1)^{j+1} a_j = a_1 - a_2 + a_3 - a_4 + \cdots$$

*converges.*

Proof. First, we will show that $n \geq m$ implies

$$\left|\sum_m^n (-1)^{j+1} a_j\right| \leq |a_m|. \tag{3.1}$$

If $m$ is odd, the first term of $\sum_m^n (-1)^{j+1} a_j$ is positive. Now, we have two cases.

Let $n$ be odd. Then,

$$\sum_m^n (-1)^{j+1} a_j = (a_m - a_{m+1}) + (a_{m+2} - a_{m+3}) + \cdots + (a_n) \geq 0,$$

since each parenthesized group is positive due to the monotonicity of the sequence **a**. Similarly,

$$\sum_m^n (-1)^{j+1} a_j = a_m + (-a_{m+1} + a_{m+2}) + \cdots + (-a_{n-1} + a_n) \leq a_m,$$

since each group is negative. Therefore,

$$0 \leq \sum_m^n (-1)^{j+1} a_j \leq a_m$$

whenever $m$ and $n$ are both odd.

Let $n$ be even. Then,

$$\sum_m^n (-1)^{j+1} a_j = (a_m - a_{m+1}) + (a_{m+2} - a_{m+3}) + \cdots + (a_{n-1} - a_n) \geq 0,$$

since each group is positive, and

$$\sum_m^n (-1)^{j+1} a_j = a_m + (-a_{m+1} + a_{m+2}) + \cdots + (-a_n) \leq a_m,$$

as each group is negative. Hence,

$$0 \leq \sum_m^n (-1)^{j+1} a_j \leq a_m$$

whenever $m$ is odd and $n$ is even.

If $m$ is even, identical reasoning shows that

$$0 \leq -\sum_m^n (-1)^{j+1} a_j \leq a_m.$$

Therefore, relation 3.1 holds for any $m, n \in \mathbb{N}$ with $n \geq m$.

Now, if $m$ is unlimited and $n \geq m$,

$$0 \leq \left| \sum_{m}^{n} (-1)^{j+1} a_j \right| \leq |a_m| \simeq 0.$$

We conclude that the alternating series converges. □

There are also nonstandard versions of other convergence tests. The proofs are not especially enlightening, so I omit these results.

### 3.3. Continuity

Since infinitesimals were invoked to understand continuous phenomena, it seems as if they should have an intimate connection with the mathematical concept of continuity. Indeed, they do.

Definition 3.25 (Continuity at a Point). Fix a function $f$ and a point $c$ at which $f$ is defined. $f$ is *continuous* at $c$ if and only if, for every real $\varepsilon > 0$, there exists a real $\delta(\varepsilon) > 0$ for which

$$|c - x| < \delta \rightarrow |f(c) - f(x)| < \varepsilon.$$

In other words, the value of $f(x)$ will be arbitrarily close to $f(c)$ if $x$ is close enough to $c$. We also write

$$\lim_{x \to c} f(x) = f(c)$$

to indicate the same relationship.

Theorem 3.26. *$f$ is continuous at $c \in \mathbb{R}$ if and only if $x \simeq c$ implies $f(x) \simeq f(c)$. Equivalently,*[1]

$$f(\text{hal}(c)) \subseteq \text{hal}(f(c)).$$

[1]Notice how closely this condition resembles the standard topological definition of continuity: $f$ is continuous at $c$ if and only if the inverse image of every neighborhood of $f(c)$ is contained in some neighborhood of $c$.

Proof. Assume that $f$ is continuous at $c$. Choose a real $\varepsilon > 0$. There exists a real $\delta > 0$ for which

$$(\forall x \in \mathbb{R})(|c - x| < \delta \to |f(c) - f(x)| < \varepsilon).$$

If $x \simeq c$, then $|c - x| < \delta$. Thus, $|f(c) - f(x)| < \varepsilon$. But $\varepsilon$ is arbitrarily small, so we must have $f(x) \simeq f(c)$.

Conversely, assume that $x \simeq c$ implies $f(x) \simeq f(c)$. Fix a positive, real number $\varepsilon$. For any infinitesimal $\delta > 0$, $|c - x| < \delta$ implies that $x \simeq c$. Then, $|f(x) - f(c)| < \varepsilon$. So,

$$(\exists \delta \in {}^*\mathbb{R}^+)(|c - x| < \delta \to |f(c) - f(x)| < \varepsilon).$$

By transfer, $f$ is continuous at $c$. □

**3.3.1. Continuous Functions.** Continuous functions are another bedrock of analysis, since they behave quite pleasantly.

Definition 3.27 (Continuous Function). A function is *continuous* on its domain if and only if it is continuous at each point in its domain.

Theorem 3.28. *A function $f$ is continuous on a set $A$ if and only if $x \simeq c$ implies $f(x) \simeq f(c)$ for every* real *$c \in A$ and every* hyperreal *$x \in {}^*A$.*

Proof. This fact follows immediately from transfer of the definitions. □

Theorem 3.28 shows that we can check continuity algebraically, rather than concoct a limit argument. (See Example 3.31.)

**3.3.2. Uniform Continuity.** The emphasis in the statement of Theorem 3.28 is crucial. If $c$ is allowed to range over the hyperreals, the condition becomes stronger.

Definition 3.29 (Uniformly Continuous). A function is *uniformly continuous* on a set $A$ if and only if, for each real $\varepsilon > 0$, there exists a *single* real $\delta > 0$ such that

$$|x - y| < \delta \rightarrow |f(x) - f(y)| < \varepsilon$$

for every $x, y \in A$. It is clear that every uniformly continuous function is also continuous.

Theorem 3.30. *$f$ is uniformly continuous if and only if $x \simeq y$ implies $f(x) \simeq f(y)$ for every* hyperreal *$x$ and $y$.*

Proof. The proof is so similar to the proof of Theorem 3.26 that it would be tiresome to repeat. □

An example of the difference between continuity and uniform continuity may be helpful.

Example 3.31. Let $f(x) = x^2$. Fix a real $c$, and let $x = c + \varepsilon$ for some $\varepsilon \in \mathbb{I}$.

$$f(x) - f(c) = (c + \varepsilon)^2 - c^2 = 2c\varepsilon + \varepsilon^2 \simeq 0,$$

so $f(x) \simeq f(c)$. Thus $f$ is continuous on $\mathbb{R}$.

But something else happens if $c$ is unlimited. Put $x = c + \frac{1}{c} \simeq c$. Then,

$$f(x) - f(c) = (c + \tfrac{1}{c})^2 - c^2 = 2c \cdot \tfrac{1}{c} + (\tfrac{1}{c})^2 = 2 + (\tfrac{1}{c})^2 \simeq 2.$$

Therefore, $f(x) \not\simeq f(c)$, which means that $f$ is not uniformly continuous on $\mathbb{R}$.

Although continuity and uniform continuity are generally distinct, they coincide for some sets.

Theorem 3.32. *If $f$ is continuous on a closed interval $[a, b] \subseteq \mathbb{R}$, then $f$ is uniformly continuous on this interval.*

Proof. Pick hyperreals $x, y \in {}^*[a, b]$ for which $x \simeq y$. Now, $x$ is limited, so we may put $c = \mathrm{sh}\,(x) = \mathrm{sh}\,(y)$. Since $a \le x \le b$ and $c \simeq x$, we have $c \in [a, b]$. Therefore $f$ is continuous at $c$, which implies that $f(x) \simeq f(c)$ and $f(y) \simeq f(c)$. By transitivity, $f(x) \simeq f(y)$, which means that $f$ is uniformly continuous on the interval. $\square$

**3.3.3. More about Continuous Functions.** As we mentioned before, the special properties of continuous functions are fundamental to analysis. One of the most basic is the intermediate value theorem, which has a very attractive nonstandard proof.

Theorem 3.33 (Intermediate Value). *If $f$ is continuous on the interval $[a, b]$ and $d$ is a point strictly between $f(a)$ and $f(b)$, then there exists a point $c \in [a, b]$ for which $f(c) = d$.*

To prove the theorem, the interval $[a, b]$ is partitioned into segments of infinitesimal width. Then, we locate a segment whose endpoints have $f$-values on either side of $d$. The common shadow of these endpoints will be the desired point $c$.

Proof. Without loss of generality, assume that $f(a) < f(b)$, so $f(a) < d < f(b)$. Define

$$\Delta_n = \frac{b - a}{n}.$$

Now, let $\mathbf{P}$ be a sequence of partitions of $[a, b]$, in which $P_n$ contains $n$ segments of width $\Delta_n$:

$$P_n = \{x \in [a, b] : x = a + j\Delta_n \text{ for } j \in \mathbb{N} \text{ with } 0 \le j \le n\}.$$

Define a second sequence, $\mathbf{s}$, where $s_n$ is the last point in the partition $P_n$ whose $f$-value is strictly less than $d$:

$$s_n = \max\{x \in P_n : f(x) < d\}.$$

Thus, for any $n$, we must have

$$a \le s_n < b \quad \text{and} \quad f(s_n) < d \le f(s_n + \Delta_n).$$

Fix an unlimited $N$. By transfer, $a \le s_N < b$, which implies that $s_N$ is limited. Put $c = \mathrm{sh}\,(s_N)$. The continuity of $f$ shows that $f(c) \simeq f(s_N)$. Now, it is clear that $\Delta_N \simeq 0$, which means that $s_N \simeq s_N + \Delta_N$. Therefore, $f(s_N) \simeq f(s_N + \Delta_N)$. Transfer shows that $f(s_N) < d \le f(s_N + \Delta_N)$. Hence, we also have $d \simeq f(s_N)$. Both $f(c)$ and $d$ are real, so $f(c) = d$. □

The extreme value theorem is another key result. It shows that a continuous function must have a maximum and a minimum on any closed interval.

Definition 3.34 (Absolute Maximum). The quantity $f(c)$ is an *absolute maximum* of the function $f$ if $f(x) < f(c)$ for every $x \in \mathbb{R}$. The *absolute minimum* is defined similarly. The maximum and minimum of a function are called its *extrema*.

Theorem 3.35 (Extreme Value). *If the function $f$ is continuous on $[a, b]$, then $f$ attains an absolute maximum and minimum on the interval $[a, b]$.*

Proof. This proof is similar to the proof of the intermediate value theorem, so I will omit the details. We first construct a uniform, finite partition of $[a, b]$. Now, there exists a partition point at which the function's value is greater than or equal to its value at any other partition point. (The existence of this point relies on the fact that the interval is closed. If the interval were open, the function might approach—but never reach—an extreme value at one of the endpoints.) Transfer yields a uniform, hyperfinite partition which has points infinitely near every real number in the interval. Fix a real point $x \in [a, b]$. Then there exists a partition point $p \in \mathrm{hal}(x)$. Since the function is continuous, $f(x) \simeq f(p)$. But there still exists a partition point $P$ at which the function's value is at least as great as at any other partition point. Hence, $f(x) \simeq f(p) \leq f(P)$. Taking shadows, we see that $f(x) \leq \mathrm{sh}\,(f(P)) = f(\mathrm{sh}\,(P))$. Therefore, the function takes its maximum value at the real point $\mathrm{sh}\,(P)$. The proof for the minimum is the same. □

## 3.4. Differentiation

Differentiation involves finding the "instantaneous" rate of change of a continuous function. This phrasing emphasizes the intimate relation between infinitesimals and derivative. Leibniz used this connection to develop his calculus. As we shall see, the nonstandard version of differentiation closely resembles Leibniz's conception.

Definition 3.36 (Derivative). If the limit

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

exists, then the function $f$ is said to be *differentiable* at the point $c$ with *derivative* $f'(c)$.

Theorem 3.37. *If $f$ is defined at the point $c \in \mathbb{R}$, then $f'(c) = L$ if and only if $f(x + \varepsilon)$ is defined for each $\varepsilon \in \mathbb{I}$, and*

$$\frac{f(c+\varepsilon) - f(c)}{\varepsilon} \simeq L.$$

Proof. This theorem follows directly from the characterization of continuity given in Section 3.3. □

Corollary 3.38. *If $f$ is differentiable at $c$, then $f$ is continuous at $c$.*

Proof. Fix a nonzero infinitesimal, $\varepsilon$.

$$f'(c) \simeq \frac{f(c+\varepsilon) - f(c)}{\varepsilon}.$$

Since $f'(c)$ is limited,

$$0 \simeq \varepsilon f'(c) \simeq f(c+\varepsilon) - f(c).$$

Therefore, $x \simeq c$ implies that $f(x) \simeq f(c)$. We conclude that $f$ is continuous at $c$. □

The next corollary reduces the process of taking derivatives to simple algebra. It legitimates Leibniz's method of differentiation, which we discussed in the Introduction and in Section 1.5.

COROLLARY 3.39. *When $f$ is differentiable at $c$,*

$$f'(c) = \operatorname{sh}\left(\frac{f(c+\varepsilon) - f(c)}{\varepsilon}\right)$$

*for any nonzero infinitesimal $\varepsilon$.* □

**3.4.1. Rules for Differentiation.** NSA makes it easy to demonstrate the rules governing the derivative. These principles allow us to differentiate algebraic combinations of functions, such as sums and products.

THEOREM 3.40. *Let $f, g$ be functions which are differentiable at $c \in \mathbb{R}$. Then $f+g$ and $fg$ are also differentiable at $c$, as is $f/g$ when $g(c) \neq 0$. Their derivatives are*

(1) $(f+g)'(c) = f'(c) + g'(c)$,
(2) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$ *and*
(3) $(f/g)'(c) = [f'(c)g(c) + f(c)g'(c)]/[g(c)]^2$.

PROOF. We prove the first two; the third is similar.

Fix a nonzero infinitesimal $\varepsilon$. Since $f$ and $g$ are differentiable at $c$, $f(c+\varepsilon)$ and $g(c+\varepsilon)$ are both defined.

$$\begin{aligned}(f+g)'(c) &= \frac{(f+g)(c+\varepsilon)-(f+g)(c)}{\varepsilon}\\ &= \frac{f(c+\varepsilon)+g(c+\varepsilon)-f(c)-g(c)}{\varepsilon}\\ &= \frac{f(c+\varepsilon)-f(c)}{\varepsilon}+\frac{g(c+\varepsilon)-g(c)}{\varepsilon}\\ &\simeq f'(c)+g'(c).\end{aligned}$$

Similarly,

$$\begin{aligned}(fg)'(c) &= \frac{(fg)(c+\varepsilon)-(fg)(c)}{\varepsilon}\\ &= \frac{f(c+\varepsilon)g(c+\varepsilon)-f(c)g(c)}{\varepsilon}\\ &= \frac{f(c+\varepsilon)g(c+\varepsilon)-f(c)g(c+\varepsilon)+f(c)g(c+\varepsilon)-f(c)g(c)}{\varepsilon}\\ &= \frac{f(c+\varepsilon)-f(c)}{\varepsilon}\cdot g(c+\varepsilon)+f(c)\cdot\frac{g(c+\varepsilon)-g(c)}{\varepsilon}\\ &\simeq f'(c)g(c+\varepsilon)+f(c)g'(c)\\ &\simeq f'(c)g(c)+f(c)g'(c).\end{aligned}$$

□

The chain rule is probably the most important tool for computing derivatives. It is only slightly more difficult to demonstrate.

THEOREM 3.41 (Chain Rule). *Fix $c \in \mathbb{R}$. If $g$ is differentiable at $c$, and $f$ is differentiable at $g(c)$, then $(f \circ g)(c) = f(g(c))$ is differentiable, and*

$$(f\circ g)'(c) = (f'\circ g)(c)\cdot g'(c) = f'(g(c))\cdot g'(c).$$

PROOF. Fix a nonzero $\varepsilon \in \mathbb{I}$. We must show that

$$\frac{f(g(c+\varepsilon))-f(g(c))}{\varepsilon} \simeq f'(g(c))\cdot g'(c). \tag{3.2}$$

There are two cases.

If $g(c+\varepsilon) = g(c)$ then both sides of relation 3.2 are zero.

Otherwise, $g(c+\varepsilon) \neq g(c)$. Put $\delta = g(c+\varepsilon) - g(c) \simeq 0$. Then,

$$\begin{aligned}\frac{f(g(c+\varepsilon)) - f(g(c))}{\varepsilon} &= \frac{f(g(c)+\delta) - f(g(c))}{\delta} \cdot \frac{\delta}{\varepsilon} \\ &\simeq f'(g(c)) \cdot \frac{g(c+\varepsilon) - g(c)}{\varepsilon} \\ &\simeq f'(g(c)) \cdot g'(c).\end{aligned}$$

□

**3.4.2. Extrema.** Derivatives are also useful for detecting at which points a function takes extreme values.

DEFINITION 3.42 (Local Maximum). The quantity $f(c)$ is a *local maximum* of the function $f$ if there exists a real number $\varepsilon > 0$ such that $f(x) \leq f(c)$ for every $x \in (c-\varepsilon, c+\varepsilon)$. A *local minimum* is defined similarly. Local minima and maxima are called *local extrema* of $f$.

THEOREM 3.43. *The function $f$ has a local maximum at the point $c$ if and only if $x \simeq c$ implies that $f(x) \leq f(c)$. An analogous theorem is true of local minima.*

PROOF. Take $f(c)$ to be a local maximum. Then, there exists a real $\varepsilon > 0$ for which

$$(\forall x \in (c-\varepsilon, c+\varepsilon))(f(x) \leq f(c)).$$

If $x \simeq c$, then $x \in (c-\varepsilon, c+\varepsilon)$, and $f(x) \leq f(c)$.

Conversely, assume that $x \simeq c$ implies $f(x) \leq f(c)$. When $\varepsilon \in \mathbb{I}^+$, $c - \varepsilon < x < c + \varepsilon$ implies that $x \simeq c$. Therefore,

$$(\exists \varepsilon \in {}^*\mathbb{R}^+)(\forall x \in {}^*\mathbb{R})(c-\varepsilon < x < c+\varepsilon \rightarrow f(x) \leq f(c)).$$

By transfer, $f(c)$ is a local maximum. □

THEOREM 3.44 (Critical Point). *If $f$ takes a local maximum at $c$ and $f$ is differentiable at $c$, then $f'(c) = 0$. The same is true for local minima.*

Proof. Fix a positive infinitesimal, $\varepsilon$. Since $f$ is differentiable at $c$, $f(c+\varepsilon)$ and $f(c-\varepsilon)$ are defined. Now,

$$f'(c) \simeq \frac{f(c+\varepsilon)-f(c)}{\varepsilon} \le 0 \le \frac{f(c-\varepsilon)-f(c)}{-\varepsilon} \simeq f'(c).$$

$f'(c)$ is real, which forces $f'(c)=0$. □

The mean value theorem now follows from the critical point and extreme value theorems by standard reasoning.

Theorem 3.45 (Mean Value). *If $f$ is differentiable on $[a,b]$, there exists a point $x \in (a,b)$ at which*

$$f'(x) = \frac{f(b)-f(a)}{b-a}.$$

## 3.5. Riemann Integration

Since the time of Archimedes, mathematicians have calculated areas by summing thin rectangular strips. Riemann's integral retains this geometrical flavor. The nonstandard approach to integration elaborates on Riemann sums by giving the rectangles infinitesimal width. This view recalls Leibniz's process of summing ($\int$) rectangles with height $f(x)$ and width $\mathrm{d}x$.

**3.5.1. Preliminaries.** To develop the integral, we need an extensive amount of terminology. In the following, $[a,b]$ is a closed, real interval and $f:[a,b]\to\mathbb{R}$ is a bounded function, i.e. it takes finite values only.

Definition 3.46 (Partition). A *partition* of $[a,b]$ is a finite set of points, $P=\{x_0,x_1,\ldots,x_n\}$ with $a=x_0\le x_1\le\cdots\le x_{n-1}\le x_n=b$. Define for $1\le j\le n$

$$M_j=\sup f(x) \quad\text{and}\quad m_j=\inf f(x) \quad\text{where } x\in[x_{j-1},x_j].$$

We also set $\Delta x_j = x_j - x_{j-1}$.

Definition 3.47 (Refinement). Take two partitions, $P$ and $P'$, of the interval $[a, b]$. $P'$ is said to be a *refinement* of $P$ if and only if $P \subseteq P'$.

Definition 3.48 (Common Refinement). A partition $P'$ which refines the partition $P_1$ and which also refines the partition $P_2$ is called a *common refinement* of $P_1$ and $P_2$.

Definition 3.49 (Riemann Sum). With reference to a function $f$, an interval $[a, b]$ and a partition $P$, define the

- upper Riemann sum by $U_a^b(f, P) = U(f, P) = \sum_1^n M_j \Delta x_j$,
- lower Riemann sum by $L_a^b(f, P) = L(f, P) = \sum_1^n m_j \Delta x_j$ and
- ordinary Riemann sum by $S_a^b(f, P) = S(f, P) = \sum_1^n f(x_{j-1}) \Delta x_j$.

The endpoints $a$ and $b$ are omitted from the notation when there is no chance of error.

Several facts follow immediately from the definitions.

Proposition 3.50. *Let $M$ be the supremum of $f$ on $[a, b]$ and $m$ be the infimum of $f$ on $[a, b]$. For any partition $P$,*

$$m(b-a) \le L(f, P) \le S(f, P) \le U(f, P) \le M(b-a). \tag{3.3}$$

Proof. The first inequality holds since $m \le m_j$ for each $j$. The second holds since $m_j \le f(x_j)$ for each $j$. The other two inequalities follow by symmetric reasoning. □

Proposition 3.51. *Let $P$ be a partition of $[a, b]$ and $P'$ be a refinement of $P$. Then*

$$U(f, P') \le U(f, P) \quad \textit{and} \quad L(f, P') \ge L(f, P).$$

Proof. Suppose that $P'$ contains exactly one point more than $P$, and let this extra point $p$ fall within the interval $[x_j, x_{j+1}]$, where $x_j$

and $x_{j+1}$ are consecutive points in $P$. Put

$$z_1 = \sup_{[x_j,p]} f(x) \quad \text{and} \quad z_2 = \sup_{[p,x_{j+1}]} f(x).$$

Both $z_1 \le M_j$ and $z_2 \le M_j$, since $M_j$ was the supremum of the function over the entire subinterval $[x_j, x_{j+1}]$. Now, we calculate

$$\begin{aligned} U(f,P) - U(f,P') &= M_j(x_{j+1} - x_j) - z_1(p - x_j) - z_2(x_{j+1} - p) \\ &= (M_j - z_1)(p - x_j) + (M_j - z_2)(x_{j+1} - p) \\ &\ge 0. \end{aligned}$$

Thus, $U(f,P') \le U(f,P)$.

If $P'$ has additional points, the result follows by iteration. The proof of the corresponding inequality for lower Riemann sums is analogous.

□

PROPOSITION 3.52. *For any two partitions $P_1$ and $P_2$, $L(f,P_1) \le U(f,P_2)$.*

PROOF. Let $P'$ be a common refinement of $P_1$ and $P_2$.

$$L(f,P_1) \le L(f,P') \le S(f,P') \le U(f,P') \le U(f,P_2).$$

□

**3.5.2. Infinitesimal Partitions.** Now, given a real number $\Delta x > 0$, define $P_{\Delta x} = \{x_0, x_1, \ldots, x_N\}$ to be the partition of $[a,b]$ into $N = \lceil (b-a)/\Delta x \rceil$ equal subintervals of width $\Delta x$. (The last segment may be smaller). For the sake of simplicity, write $U(f,\Delta x)$ in place of the notation $U(f,P_{\Delta x})$. We can now regard $U(f,\Delta x)$, $L(f,\Delta x)$ and $S(f,\Delta x)$ as functions of the real variable $\Delta x$.

THEOREM 3.53. *If $f$ is continuous on $[a,b]$ and $\Delta x$ is infinitesimal,*

$$L(f,\Delta x) \simeq S(f,\Delta x) \simeq U(f,\Delta x).$$

Proof. First, define for each $\Delta x$ the quantity

$$\mu(\Delta x) = \max\{M_j - m_j : 1 \leq j \leq N\},$$

which represents the maximum oscillation in any subinterval of the partition $P_{\Delta x}$.

Now, fix an infinitesimal $\Delta x$. Since $f$ is continuous and $x_j \simeq x_{j-1}$ for each $j$, $M_j \simeq m_j$. Therefore, the maximum difference $\mu(\Delta x)$ must be infinitesimal.

Form the difference

$$\begin{aligned}
U(f, \Delta x) - L(f, \Delta x) &= \sum_1^N (M_j - m_j)\Delta x \\
&\leq \mu(\Delta x) \sum_1^n \Delta x \\
&\leq \mu(\Delta x) \cdot N \cdot \Delta x \\
&= \mu(\Delta x) \left\lceil \frac{b-a}{\Delta x} \right\rceil \Delta x \\
&\leq \mu(\Delta x) \left( \frac{b-a}{\Delta x} + 1 \right) \Delta x \\
&= \mu(\Delta x)(b-a) + \mu(\Delta x)\Delta x \\
&\simeq 0.
\end{aligned}$$

By transfer of relation 3.3, the ordinary Riemann sum $S(f, \Delta x)$ is sandwiched between the upper and lower sums, so it is infinitely near both. □

**3.5.3. The Riemann Integral.** Finally, we are prepared to define the integral in the sense of Riemann.

Definition 3.54 (Riemann Integrable). Let $\Delta x$ range over $\mathbb{R}$. If

$$L = \lim_{\Delta x \to 0} L(f, \Delta x) \quad \text{and} \quad U = \lim_{\Delta x \to 0} U(f, \Delta x)$$

both exist and $L = U$, then $f$ is *Riemann integrable* on $[a, b]$. We write

$$\int_a^b f(x)\,\mathrm{d}x$$

to denote the common value of the limits.

THEOREM 3.55. *If $f$ is continuous on $[a, b]$, then $f$ is Riemann integrable, and*

$$\int_a^b f(x)\,\mathrm{d}x = \mathrm{sh}\,(S(f, \Delta x)) = \mathrm{sh}\,(L(f, \Delta x)) = \mathrm{sh}\,(U(f, \Delta x))$$

*for every infinitesimal $\Delta x$.*

PROOF. For any two infinitesimals, $\Delta x, \Delta y > 0$,

$$L(f, \Delta x) \leq U(f, \Delta y) \simeq L(f, \Delta y) \leq U(f, \Delta x) \simeq L(f, \Delta x).$$

Therefore, $L(f, \Delta x) \simeq L(f, \Delta y)$ and $U(f, \Delta x) \simeq U(f, \Delta y)$ whenever $\Delta x \simeq \Delta y \simeq 0$. Therefore, $L(f, \Delta x)$ and $U(f, \Delta x)$ are continuous at $\Delta x = 0$. Theorem 3.53 shows that

$$\lim_{\Delta x \to 0} L(f, \Delta x) = \lim_{\Delta x \to 0} U(f, \Delta x).$$

The result follows immediately. □

**3.5.4. Properties of the Integral.** The standard properties of integrals follow easily from the definition of the integral as the shadow of a Riemann sum, the properties of sums and the properties of the shadow map.

THEOREM 3.56. *If $f$ and $g$ are integrable over $[a, b] \subseteq \mathbb{R}$, then*

- $\int_a^b cf(x)\,\mathrm{d}x = c\int_a^b f(x)\,\mathrm{d}x$;
- $\int_a^b [f(x) + g(x)]\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x$;
- $\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x$;
- $\int_a^b f(x)\,\mathrm{d}x \leq \int_a^b g(x)\,\mathrm{d}x$ *if* $f(x) \leq g(x)$ *for all* $x \in [a, b]$;
- $m(b - a) \leq \int_a^b f(x)\,\mathrm{d}x \leq M(b - a)$ *where* $m \leq f(x) \leq M$ *for all* $x \in [a, b]$.

**3.5.5. The Fundamental Theorem of Calculus.** Finally, we will prove the Fundamental Theorem of Calculus using nonstandard methods. This theorem bears its impressive name because it demonstrates the intimate link between the processes of differentiation and integration—they are inverse operations. Newton and Leibniz are credited with the discovery of calculus because they were the first to develop this theorem. Nonstandard Analysis furnishes a beautiful proof.

THEOREM 3.57. *If $f$ is continuous on $[a, b]$, the area function*

$$F(x) = \int_a^x f(t)\,\mathrm{d}t$$

*is differentiable on $[a, b]$ with derivative $f$.*

There is an intuitive reason that this theorem holds: the change in the area function over an infinitesimal interval $[x, x+\varepsilon]$ is approximately equal to the area of a rectangle with base $[x, x+\varepsilon]$ which fits under the curve (see Figure 3.1).

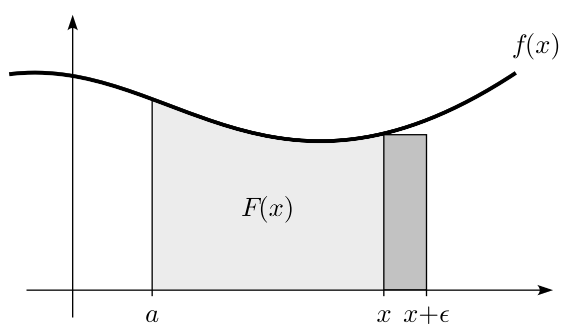


FIGURE 3.1. Differentiating the area function.

Algebraically,

$$F(x+\varepsilon) - F(x) \approx \varepsilon \cdot f(x).$$

Dividing this relation by $\varepsilon$ suggests the result. Of course, we must formalize this reasoning.

PROOF. If $\varepsilon$ is a positive real number less than $b - x$,

$$F(x+\varepsilon) - F(x) = \int_x^{x+\varepsilon} f(t)\,\mathrm{d}t.$$

By the extreme value theorem, the continuous function $f$ attains a maximum at some real point $M$ and a minimum at some real point $m$, so

$$[(x+\varepsilon) - x] \cdot f(m) \le \int_x^{x+\varepsilon} f(t)\,\mathrm{d}t \le [(x+\varepsilon) - x] \cdot f(M), \text{ or}$$

$$\varepsilon \cdot f(m) \le \int_x^{x+\varepsilon} f(t)\,\mathrm{d}t \le \varepsilon \cdot f(M).$$

Dividing by $\varepsilon$,

$$f(m) \le \frac{F(x+\varepsilon) - F(x)}{\varepsilon} \le f(M). \tag{3.4}$$

By transfer, if $\varepsilon \in \mathbb{I}^+$, there are hyperreal $m, M \in {}^*[x, x+\varepsilon]$ for which equation 3.4 holds.

But now, $x + \varepsilon \simeq x$, so $m \simeq x$ and $M \simeq x$. The continuity of $f$ shows that

$$\frac{F(x+\varepsilon) - F(x)}{\varepsilon} \simeq f(x). \tag{3.5}$$

A similar procedure shows that relation 3.5 holds for any negative infinitesimal $\varepsilon$.

Therefore, the area function $F$ is differentiable at $x$ for any $x \in [a, b]$ and its derivative $F'(x) = f(x)$. □

COROLLARY 3.58 (Fundamental Theorem of Calculus). *If a function $F$ has a continuous derivative $f$ on $[a, b]$, then*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

PROOF. Let $A(x) = \int_a^x f(x)\,\mathrm{d}x$. For $x \in [a, b]$,

$$(A(x) - F(x))' = A'(x) - F'(x) = f(x) - f(x) = 0,$$

which implies that $(A - F)$ is constant on $[a, b]$. Then

$$F(b) - F(a) = A(b) - A(a) = \int_a^b f(x)\,\mathrm{d}x.$$

$\square$

# Conclusion

In the last chapter, we saw how NSA offers intuitive direct proofs of many classical theorems. Nonstandard Analysis would be a curiosity if it only allowed us to reprove theorems of real analysis in a streamlined fashion. But its application in other areas of mathematics shows it to be a powerful tool. Here are two examples.

**Topology:** Topology studies the spatial structure of sets. The key concepts are proximity and adjacency, which are formalized by defining the *open neighborhood* of a point. Intuitively, an open set about $p$ contains all the points near $p$ [**7**, 113]. In metric spaces, topology can be arithmetized: the open neighborhoods of $p$ contain those points which are less than a certain distance from $p$. The distance between any two points is determined by a function which returns a positive, real value. With NSA, the *distance function* can be extended, so that it returns positive hyperreals. Then, we can say that two points are near each other if and only if they are at an infinitesimal distance. This definition simplies many fundamental ideas in the topology of metric spaces. Furthermore, the nonstandard extension of a topological space can facilitate the proof of general topological theorems, just as the hyperreals facilitate proofs about $\mathbb{R}$ [**9**].

**Distributions:** Distributions are generalized functions which are extremely useful in electrical engineering and modern physics.

> The space of distributions is somewhat complicated to define from a traditional perspective, because it contains elements like the Dirac $\delta$ function. Conceptually, this "function" of the reals is zero everywhere except at the origin, where it is infinite—but only so infinite that the area beneath it equals 1. NSA allows us to view the $\delta$ function as a nonstandard function which has an unlimited value on an infinitesimal interval [**11**, 93–95]. It turns out that all distributions can be seen as internal functions. In fact, using suitable definitions, the distributions may even be realized as a subset of ${}^*C^\infty(\mathbb{R})$, the infinitely differentiable internal functions. But that is another theorem for another day.

Other areas of application include differential equations, probability, combinatorics and functional analysis [**10**], [**7**], [**11**].

---

Classical analysis is often confusing and technical. Fiddling with epsilons and deltas obscures the conceptual core of a proof. Infinitesimals and unlimited numbers, however, brightly illuminate many mathematical concepts. If logic had advanced as quickly as analysis, NSA might well be the dominant paradigm. And if Gödel is right, it may yet be.

APPENDIX A

# Nonstandard Extensions

The most general method of developing Nonstandard Analysis begins with the concept of a nonstandard extension. It can be shown that every nonempty set $\mathbb{X}$ has a proper nonstandard extension ${}^*\mathbb{X}$ which is a strict superset of $\mathbb{X}$. This is accomplished using an ultrapower construction, which is similar to that in Section 2.2.

Henson suggests that the properties of a proper nonstandard extension are best considered from a geometrical standpoint. Since functions and relations are identified with their graphs, this view is appropriate for all mathematical objects. The essential idea is that the geometric nature of an object does not change under a proper nonstandard extension, although it may be comprised of many more points. For example, the line segment $[0, 1]$ is still a line segment of unit length under the mapping, yet it contains nonstandard elements. Similarly, the unit square remains a unit square, with new, nonstandard elements. Et cetera. This explanation indicates why the nonstandard extension preserves certain set-theoretic properties like Cartesian products [**8**].

DEFINITION A.1 (Nonstandard Extension of a Set). Let $\mathbb{X}$ be any nonempty set. A *nonstandard extension of* $\mathbb{X}$ consists of a mapping that assigns a set ${}^*A$ to each $A \subseteq \mathbb{X}^m$ for all $m \geq 0$, such that ${}^*\mathbb{X}$ is nonempty and the following conditions are satisfied for all $m, n \geq 0$:

(1) The mapping preserves Boolean operations on subsets of $\mathbb{X}^m$. If $A, B \subseteq \mathbb{X}^m$ then

- ${}^*A \subseteq ({}^*\mathbb{X})^m$;

- ${}^*(A \cap B) = ({}^*A \cap {}^*B)$;
- ${}^*(A \cup B) = ({}^*A \cup {}^*B)$;
- ${}^*(A \setminus B) = ({}^*A) \setminus ({}^*B)$.

(2) The mapping preserves basic diagonals. If $\Delta = \{(x_1, \ldots, x_m) \in \mathbb{X}^m : x_i = x_j, 1 \leq i < j \leq m\}$ then ${}^*\Delta = \{(x_1, \ldots, x_m) \in ({}^*\mathbb{X})^m : x_i = x_j, 1 \leq i < j \leq m\}$.

(3) The mapping preserves Cartesian products. If $A \subseteq \mathbb{X}^m$ and $B \subseteq \mathbb{X}^n$, then ${}^*(A \times B) = {}^*A \times {}^*B$. (We regard $A \times B$ as a subset of $\mathbb{X}^{m+n}$.)

(4) The mapping preserves projections that omit the final coordinate. Let $\pi$ denote projection of $(n+1)$-tuples on the first $n$ coordinate. If $A \subseteq \mathbb{X}^{n+1}$ then ${}^*(\pi(A)) = \pi({}^*A)$.

APPENDIX B

# Axioms of Internal Set Theory

Nelson's Internal Set Theory (IST) adds a new predicate, *standard*, to classical set theory. Three primary axioms govern the use of this new predicate. Note that the term *classical* refers to any sentence which does use the term "standard" [**11**].

**Idealization:** For any classical, binary relation $\mathscr{R}$, the following are equivalent:

(1) For any standard and finite set $E$, there is an $x = x(E)$ such that $x \mathscr{R} y$ holds for each $y \in E$.

(2) There is an $x$ such that $x \mathscr{R} y$ holds for all standard $y$.

**Standardization:** Let $E$ be a standard set and $P$ be a predicate. Then there is a unique, standard subset $A = A(P) \subseteq E$ whose standard elements are precisely the standard elements $x \in E$ for which $P(x)$ is true.

**Transfer:** Let $F$ be a classical formula with a finite number of parameters. $F(x, c_1, c_2, \ldots, c_n)$ holds for all standard values of $x$ if and only if $F(x, c_1, c_2, \ldots, c_n)$ holds for all values of $x$, standard and nonstandard.

APPENDIX C

# About Filters

The direct power construction of the hyperreals depends crucially on the properties of filters and the existence of a nonprincipal ultrafilter on $\mathbb{N}$. Here are some key definitions, lemmata and theorems about filters, taken from Goldblatt [**7**, pp. 18–21]. $\mathbb{X}$ will denote a nonempty set.

Definition C.1 (Power Set). The *power set* of $\mathbb{X}$ is the set of all subsets of $\mathbb{X}$:

$$\mathscr{P}(\mathbb{X}) = \{A : A \subseteq \mathbb{X}\}.$$

Definition C.2 (Filter). A *filter* on $\mathbb{X}$ is a nonempty collection, $\mathscr{F} \subseteq \mathscr{P}(\mathbb{X})$, which satisfies the following axioms:

- If $A, B \in \mathscr{F}$, then $A \cap B \in \mathscr{F}$.
- If $A \in F$ and $A \subseteq B \subseteq \mathbb{X}$, then $B \in \mathscr{F}$.

$\emptyset \in \mathscr{F}$ if and only if $\mathscr{F} = \mathscr{P}(\mathbb{X})$. $\mathscr{F}$ is a *proper filter* if and only if $\emptyset \notin \mathscr{F}$. Any filter has $\mathbb{X} \in \mathscr{F}$, and $\{\mathbb{X}\}$ is the smallest filter on $\mathbb{X}$.

Definition C.3 (Ultrafilter). An *ultrafilter* is a filter which satisfies the additional axiom that

- For any $A \subseteq \mathbb{X}$, exactly one of $A$ and $\mathbb{X} \setminus A$ is an element of $\mathscr{F}$.

Definition C.4 (Principal Ultrafilter). For any $x \in \mathbb{X}$,

$$\mathscr{F}^x = \{A \subseteq \mathbb{X} : x \in A\}$$

is an ultrafilter, called the *principal ultrafilter generated by* $x$. If $\mathbb{X}$ is finite, then every ultrafilter is principal. A *nonprincipal ultrafilter* is an ultrafilter which is not generated in this fashion.

DEFINITION C.5 (Filter Generated by $\mathscr{H}$). Given a nonempty collection, $\mathscr{H} \subseteq \mathscr{P}(\mathbb{X})$, the *filter generated by* $\mathscr{H}$ is the collection

$$\mathscr{F}^{\mathscr{H}} = \{A \subseteq \mathbb{X} : A \subseteq B_1 \cap \cdots \cap B_k \text{ for some } k \text{ and some } B_j \in \mathscr{H}\}.$$

DEFINITION C.6 (Cofinite Filter). $\mathscr{F}^{\mathrm{co}} = \{A \subseteq \mathbb{X} : \mathbb{X} \setminus A \text{ is finite}\}$ is called the *cofinite* filter on $\mathbb{X}$. It is proper if and only if $\mathbb{X}$ is infinite. $\mathscr{F}^{\mathrm{co}}$ is not an ultrafilter.

PROPOSITION C.7. *An ultrafilter* $\mathscr{F}$ *satisfies*

- $A \cap B \in \mathscr{F}$ *iff* $A \in \mathscr{F}$ *and* $B \in \mathscr{F}$,
- $A \cup B \in \mathscr{F}$ *iff* $A \in \mathscr{F}$ *or* $B \in \mathscr{F}$, *and*
- $\mathbb{X} \setminus A \in \mathscr{F}$ *iff* $A \notin \mathscr{F}$.

PROPOSITION C.8. *If* $\mathscr{F}$ *is an ultrafilter and* $\{A_1, A_2, \ldots, A_k\}$ *is a finite collection of pairwise disjoint sets such that*

$$A_1 \cup A_2 \cup \cdots \cup A_k \in \mathscr{F},$$

*then precisely one of these* $A_j \in \mathscr{F}$.

PROPOSITION C.9. *If an ultrafilter contains a finite set, then it contains a singleton* $\{x\}$. *Then, this ultrafilter equals* $\mathscr{F}^x$, *which means that it is principal. As a result, a nonprincipal ultrafilter must contain all cofinite sets. This fact is crucial in the construction of the hyperreals.*

PROPOSITION C.10. $\mathscr{F}$ *is an ultrafilter on* $\mathbb{X}$ *if and only if it is a* maximal proper filter, *i.e. a proper filter which cannot be extended to a larger proper filter.*

Definition C.11 (Finite Intersection Property). We say that the collection $\mathscr{H} \subseteq \mathscr{P}(\mathbb{X})$ has the *finite intersection property* or *fip* if the intersection of each nonempty finite subcollection is nonempty. That is,

$$B_1 \cap \cdots \cap B_k \neq \emptyset$$

for any finite $k$ and subsets $B_j \in \mathscr{H}$. Note that a filter $\mathscr{F}^{\mathscr{H}}$ is proper if and only if $\mathscr{H}$ has the fip.

Proposition C.12. *If $\mathscr{H}$ has the fip and $A \subseteq \mathbb{X}$, then at least one of $\mathscr{H} \cup \{A\}$ and $\mathscr{H} \cup \{\mathbb{X} \setminus A\}$ has the fip.*

Finally, I give Goldblatt's proof that there exists a nonprincipal ultrafilter on any infinite set.

Proposition C.13 (Zorn's Lemma). *Let $(P, \leq)$ be a set endowed with a partial ordering, under which every linearly ordered subset (or "chain") has an upper bound in $P$. Then $P$ contains a $\leq$-maximal element.*

Zorn's lemma is equivalent to the Axiom of Choice.

Theorem C.14. *Any collection of subsets of $\mathbb{X}$ that has the finite intersection property can be extended to an ultrafilter on $\mathbb{X}$.*

Proof. If $\mathscr{H}$ has the fip, then $\mathscr{F}^{\mathscr{H}}$ is proper. Let $\mathscr{Z}$ be the collection of all proper filters on $\mathbb{X}$ that include $\mathscr{F}^{\mathscr{H}}$, partially ordered by set inclusion, $\subseteq$. Choose any totally ordered subset of $\mathscr{Z}$. The union of the members of this chain is in $\mathscr{Z}$. Hence every totally ordered subset of $\mathscr{Z}$ has an upper bound in $\mathscr{Z}$. By Zorn's Lemma, $\mathscr{Z}$ has a maximal element, which will be a maximal proper filter on $\mathbb{X}$ and therefore an ultrafilter. $\square$

Corollary C.15. *Any infinite set has a nonprincipal ultrafilter on it.*

Proof. If $\mathbb{X}$ is infinite, then the cofinite filter on $\mathbb{X}$, $\mathscr{F}^{\text{co}}$ is proper and has the fip. Therefore, it is contained in some ultrafilter $\mathscr{F}$. For any $x \in \mathbb{X}$, the set $\mathbb{X} \setminus \{x\} \in \mathscr{F}^{\text{co}} \subseteq \mathscr{F}$. Since $\{x\} \in \mathscr{F}^x$, we conclude that $\mathscr{F} \neq \mathscr{F}^x$. Thus $\mathscr{F}$ in nonprincipal. $\square$

In fact, an infinite set supports a vast number of nonprincipal ultrafilters. The set of nonprincipal ultrafilters on $\mathbb{N}$ has the same cardinality as $\mathscr{P}(\mathscr{P}(\mathbb{N}))$ [**7**, 33].

# Bibliography


[1] Bell, E.T. *Men of Mathematics: The Lives and Achievements of the Great Mathematicians from Zeno to Poincaré.* Simon and Schuster, New York, 1937.

[2] Bell, J.L. *A Primer of Infinitesimal Analysis.* Cambridge University Press, Cambridge, 1998

[3] Boyer, Carl B. *The History of the Calculus and its Conceptual Development.* Dover Publications, New York, 1949.

[4] "Luitzen Egbertus Jan Brouwer." *The MacTutor History of Mathematics Archive.* http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Brouwer.html. 11 April 1999.

[5] Cutland, Nigel J. "Nonstandard Real Analysis." *Nonstandard Analysis: Theory and Applications.* Eds. Lief O. Arkeryd *et al.* NATO ASI Series C 493. Kluwer Academic Publishers, Dordrecht, 1996.

[6] Dauben, Joseph Warren. *Abraham Robinson: The Creation of Nonstandard Analysis; A Personal and Mathematical Odyssey.* Princeton University Press, Princeton, 1995.

[7] Goldblatt, Robert. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.* Graduate Texts in Mathematics #188. Springer-Verlag, New York, 1998.

[8] Henson, C. Ward. "Foundations of Nonstandard Analysis." *Nonstandard Analysis: Theory and Applications.* Eds. Lief O. Arkeryd *et al.* NATO ASI Series C 493. Kluwer Academic Publishers, Dordrecht, 1996.

[9] Loeb, Peter A. "Nonstandard Analysis and Topology." *Nonstandard Analysis: Theory and Applications.* Eds. Lief O. Arkeryd *et al.* NATO ASI Series C 493. Kluwer Academic Publishers, Dordrecht, 1996.

[10] *Nonstandard Analysis: Theory and Applications.* Eds. Lief O. Arkeryd *et al.* NATO ASI Series C 493. Kluwer Academic Publishers, Dordrecht, 1996.

[11] Robert, Alain. *Nonstandard Analysis.* John Wiley & Sons, Chichester, 1988.

[12] Rudin, Walter. *Principles of Mathematical Analysis.* 3rd ed. International Series in Pure and Applied Mathematics. McGraw-Hill, New York, 1976.

[13] Russell, Bertrand. *Principles of Mathematics.* 2nd ed. W.W. Norton & Company, New York, 1938.

[14] Russell, Bertrand. "Definition of Number." *The World of Mathematics.* Vol. 1. Simon and Schuster, New York, 1956.

[15] Schechter, Eric. *Handbook of Analysis and Its Foundations.* Academic Press, San Diego, 1997.

[16] Varberg, Dale and Edwin J. Purcell. *Calculus with Analytic Geometry.* Prentice Hall, Englewood Cliffs, 1992.

This thesis is set in the Computer Modern family of typefaces, designed by Dr. Donald Knuth for the beautiful presentation of mathematics. It was composed on a PowerMacintosh 6500/250 using Knuth's typesetting software TeX.

## About the Author

Joel A. Tropp was born in Austin, Texas on July 18, 1977. He was deported to Durham, NC in 1988. He sojourned there until 1995, at which point he graduated from Charles E. Jordan high school. Mr. Tropp then matriculated in the Plan II honors program at the University of Texas at Austin, thereby going back where he came from. At the University, he participated in the Normandy Scholars, Junior Fellows and Dean's Scholars programs. He was an entertainment writer for the *Daily Texan*, and he edited the Plan II feature magazine, *The Undecided*, for three years. In 1998, he won a Barry M. Goldwater Scholarship, and he was a semi-finalist for the British Marshall. Mr. Tropp is a member of Phi Beta Kappa, and he is the 1999 Dean's Honored Graduate in Mathematics. After graduating, he will remain at the University as a Ph.D. student in the Computational Applied Math program, supported by the CAM graduate fellowship.