Real and Nonstandard Analyses: An Overview

Preston Vo

Introduction

A **set** is a collection of elements. These elements can be anything: numbers, shapes, colors, and so on. For instance, consider the set A of the primary colors:

$$A = \{\text{red, blue, yellow}\}$$

A **function** is a mapping between the elements of two sets where each element from one set is assigned to exactly one element from the other. In the realm of single variable calculus, functions predominantly deal with sets of real numbers. For example, the function

$$f(x) = x^2$$

takes a set of inputs (x) and produces a corresponding set of their squares f(x) in the form of **ordered pairs**:

x	f(x)
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9

Plotting the points generated by $f(x)$ on the XY coordinate plane produces the following graph:

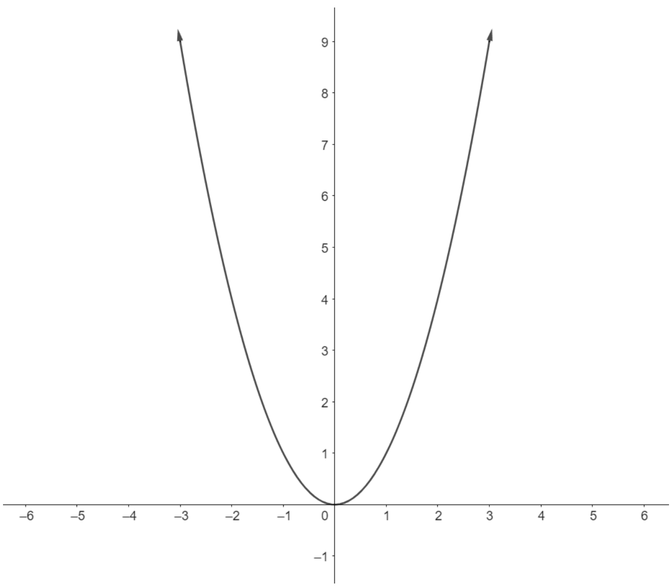

Figure 1: f(x) produces a parabolic curve

Limits analyze how the outputs of such functions behave as their inputs (x) approach either a particular point or ∞. For example, the parabolic function above approaches 4 as x approaches 2 and is denoted like so:

$$\lim_{x\to 2} x^2 = 4$$

As x approaches ∞, x^2 also approaches ∞:

$$\lim_{x\to \infty} x^2 = \infty$$

Limits are a result of **analysis**, an area of mathematics that deals with continuous change and continuous functions, broadly speaking. I intend to not only explore and compare two different branches of analysis that ultimately lead to the same results encountered in calculus, but to do so in a meticulously pedagogical manner. One branch relies on a rigorous definition of the limit, while the other abandons the idea of limits entirely in favor of a concept utilized in the upbringing of calculus over the past several centuries.

The Standard Approach

Introduction

Real analysis is the study of functions, sequences, and sets involving real numbers, and it is used as the traditional means of formalizing the mechanisms presented in calculus courses. This overview of how real analysis builds toward certain calculus principles will follow section A.2 of Appendix A in

> Simmons, G. (1996). *Calculus With Analytic Geometry* (2nd ed.), Mcgraw-Hill Education.

Being mindful of the fact that the absolute value of the difference between any two values, $|a - b|$, represents the distance between them will greatly aid in understanding the notation that follows. The limit is defined like so:

> Let a function $f(x)$ be defined on some interval containing the number c such that there are x's in the domain of $f(x)$ where
>
> $$0 < |x - c| < \delta$$
>
> for every positive number δ. The statement
>
> $$\lim_{x\to c} f(x) = L$$
>
> is then defined like so: For every positive number ε, there exists a positive number δ such that
>
> $$|f(x) - L| < \varepsilon$$
>
> for every x in the domain of $f(x)$ where
>
> $$0 < |x - c| < \delta$$

This definition states that $f(x)$ approaches some value L as x approaches a some number c if it can be shown that, for any set of outputs that lie within some distance ε from L, there exists a corresponding set of inputs (x's) that lie within some distance δ from c which *guarantees* that $f(x)$ falls within said range (ε) of L. In this sense, we can bring the range of outputs as close as we want to L (letting epsilon go to 0) while being absolutely sure that $f(x)$ lies within it. This is known as the **epsilon-delta definition** of the limit.

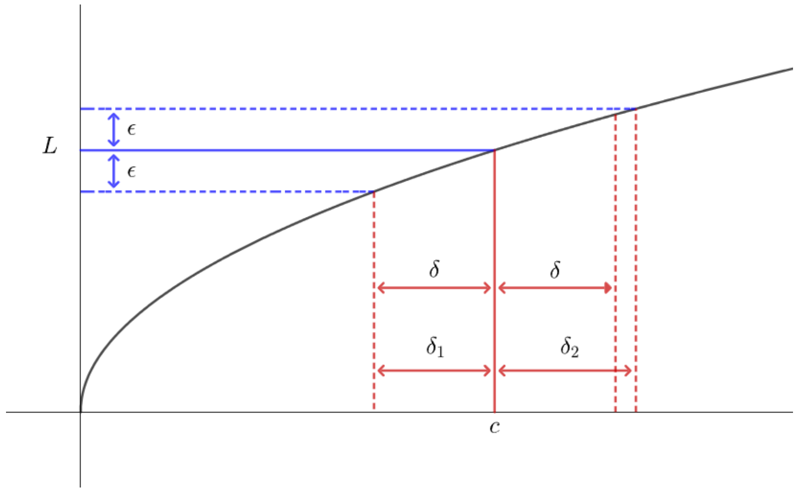

Figure 2: Epsilon-delta definition of the limit visualized

The interval $(L-\varepsilon, L+\varepsilon)$ describes the range of outputs that lie within a distance ε from L. One detail to note in the visualization above (Figure 2) is the presence of δ_1 and δ_2. This is because the x corresponding to $L-\varepsilon$ does not necessarily lie the same distance away from c as the x corresponding to $L+\varepsilon$, since the rate at which $f(x)$ changes may vary as x sweeps from $c-\delta_1$ to $c+\delta_2$. To illustrate this, notice how the curvature in the graph is steeper on the left-hand side of c compared to that on the right-hand side. This means that sweeping through some range of outputs on the left requires a smaller increment of x as opposed to sweeping through that same range of outputs on the right, since the output of the function increases at a faster rate on the left. Therefore, $\delta_1 < \delta_2$ for this particular graph.

This complication can be readily resolved by letting δ equal the smaller of δ_1 and δ_2:

$$\delta = \min(\delta_1, \delta_2).$$

min() is shorthand for taking the smallest value amongst the set of numbers present between the parentheses. For instance, if $x = \min(1, 2, 3)$, then $x = 1$. Allowing δ to be defined in this way works because of the following reasoning: Assume that $\delta_1 < \delta_2$ and that $|x-c| < \delta_2$ implies $|f(x) - L| < \varepsilon$ for some input x. If $|x-c| < \delta_1$, then $|x-c| < \delta_2$ since $\delta_1 < \delta_2$. It is therefore assured that $|f(x) - L| < \varepsilon$.

Employing the Definition

The epsilon-delta definition of the limit can now be used to prove various properties of functions. As a basic example, consider the following theorem:

Theorem 1
If f(x) = x, then $\lim_{x\to a} f(x) = a$, or

$$\lim_{x\to a} x = a$$

In order to prove this limit, we need to demonstrate that for any set of outputs that lie within ε of the limit (a), there exists a corresponding set of inputs that lie within δ of $x = a$ that assures that the output of $f(x) = x$ lies within the distance ε from a.

To start, choose some $\varepsilon > 0$, and let $\delta = \varepsilon$. For any x satisfying the inequalities $0 < |x-a| < \delta$, we know that $|f(x) - a| < \varepsilon$. This is because $f(x) = x$, and $\delta = \varepsilon$. The theorem is therefore proven.

We can also prove some essential limit laws: their sums, differences, products, and quotients.

Theorem 2 - Limit Laws
If $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$, then

(i) $\lim_{x \to a} [f(x) + g(x)] = L + M$
(ii) $\lim_{x \to a} [f(x - g(x))] = L - M$
(iii) $\lim_{x \to a} f(x)g(x) = LM$
(iv) $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{L}{M}$

To prove (i), we let $\varepsilon > 0$ be given and allow $\delta_1, \delta_2 > 0$ where

$$0 < |x - a| < \delta_1 \Rightarrow |f(x) - L| < \tfrac{1}{2}\varepsilon$$

and

$$0 < |x - a| < \delta_2 \Rightarrow |f(x) - M| < \tfrac{1}{2}\varepsilon.$$

For those unfamiliar with the $\Rightarrow$ symbol, it means that the statement following it is implied (or logically follows) from the statement preceding the symbol.

The $\frac{1}{2}$'s in front of the ε's may cause some confusion, but recall that when $\lim_{x \to c} f(x) = L$, the epsilon-delta definition tells us that there exists a set of x's lying within some distance δ from c such that the distance between $f(x)$ and L is always less than ε. Knowing this, it then follows that if there exists $\delta > 0$ such that $|f(x) - L| < \frac{1}{2}\varepsilon$ for some $\varepsilon > 0$, then $|f(x) - L| < \varepsilon$, because $\frac{1}{2}\varepsilon$ is smaller than ε.

Continuing the proof, we let $\delta = \min(\delta_1, \delta_2)$. If $0 < |x - a| < \delta$, then

$$\begin{aligned} |[f(x) + g(x)] - (L + M)| &= |[f(x) - L] + [g(x) - M]| && (1) \\ &\leq |f(x) - L| + |g(x) - M| && (2) \\ &< \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon, && (3) \end{aligned}$$

proving (i). Analyzing the three-step sequence above in further detail: (1) takes advantage of the associative property of addition and moves terms around. (2) is a subtle application of the distances variation of the **triangle inequality** which states that, for real numbers x and y,

$$|x + y| \leq |x| + |y|.$$

It essentially says that the distance between 0 and the sum of two numbers can be no more than the combined distances of x to 0 and y to 0. For example, if $x = 3$ and $y = -1$, then

$$|x + y| = |3 - 1| = 2 < 4 = |3| + |-1| = |x| + |y|.$$

While a formal proof is omitted, it may help to think of this inequality in terms of walking in two opposite directions. In particular, let positive numbers represent walking forward one way, while negative numbers represent walking backwards the other way. Therefore, if 0 is the position where one starts, $|x + y|$ represents the distance one stands from 0 after some combination of walking forwards and backwards, while $|x| + |y|$ represents the *total* distance walked. The distance from 0 can only ever be as large as the total distance walked (by walking only forwards or backwards), so $|x + y| \leq |x| + |y|$.

Returning to the limit proof, the two values involved in the triangle inequality are $f(x) - L$ and $g(x) - M$. (3) substitutes both $|f(x) - L|$ and $|g(x) - M|$ for $\frac{1}{2}\varepsilon$. Since $|f(x) - L| < \frac{1}{2}\varepsilon$ and $|g(x) - M| < \frac{1}{2}\varepsilon$, it follows that $|f(x) - L| + |g(x) - M| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon$. The theorem is ultimately proven, because the difference between the function, $f(x) + g(x)$, and the desired limit, $L + M$, was shown to be less than any $\varepsilon > 0$ given an appropriate δ.

(ii) $\lim_{x\to a}[f(x-g(x))] = L - M$

The proof of (ii) is similar to that of (i). We once again allow $\delta_1, \delta_2 > 0$ where

$$0 < |x-a| < \delta_1 \Rightarrow |f(x) - L| < \tfrac{1}{2}\varepsilon$$

and

$$0 < |x-a| < \delta_2 \Rightarrow |f(x) - M| < \tfrac{1}{2}\varepsilon.$$

Let $\delta = \min(\delta_1, \delta_2)$. If $0 < |x-c| < \delta$, then

$$\begin{aligned}
||[f(x)-g(x)| - (L-M)]| &= |[f(x)-L] + [M-g(x)]| \\
&\le |f(x)-L| + |M-g(x)| \\
&= |f(x)-L| + |g(x)-M| \\
&< \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon,
\end{aligned}$$

proving (ii).

(iii) $\lim_{x\to a} f(x)g(x) = LM$

Proving (iii) is more complicated. We being by adding and subtracting $f(x)M$ to help relate the quantity $f(x)g(x) - LM$ to the differences $f(x) - L$ and $g(x) - M$:

$$\begin{aligned}
|f(x)g(x) - LM| &= |[f(x)g(x) - f(x)M] + [f(x)M - LM]| \\
&\le |f(x)g(x) - f(x)M| + |f(x)M - LM| \\
&= |f(x)||g(x) - M| + |M||f(x) - L| \\
&\le |f(x)||g(x) - M| + (|M|+1)|f(x) - L|. \qquad (4)
\end{aligned}$$

Provided some $\varepsilon > 0$, it is certain that $\delta_1, \delta_2, \delta_3 > 0$ all exist where

$$0 < |x-a| < \delta_1 \Rightarrow |f(x) - L| < 1 \Rightarrow |f(x)| < |L| + 1; \qquad (5)$$

$$0 < |x-a| < \delta_2 \Rightarrow |g(x) - M| < \frac{1}{2}\varepsilon\left(\frac{1}{|L|+1}\right); \qquad (6)$$

$$0 < |x-a| < \delta_3 \Rightarrow |f(x) - L| < \frac{1}{2}\varepsilon\left(\frac{1}{|M|+1}\right). \qquad (7)$$

(5) comes from the fact that $\lim_{x\to a} f(x) = L$, so a $\delta_1 > 0$ exists for every $\varepsilon > 0$. In the case of (5), $\varepsilon = 1$. While (6) and (7) may look confusing, they once again stem from the definition of the limit. For instance, since $\lim_{x\to a} g(x) = M$, a $\delta_2 > 0$ exists for every $\varepsilon > 0$. ε is fundamentally a positive value, which makes $\dfrac{1}{2}\varepsilon\left(\dfrac{1}{|L|+1}\right)$ also positive, justifying (6).

Resuming the proof, we let $\delta = \min(\delta_1, \delta_2, \delta_3)$. Then

$$0 < |x-a| < \delta \Rightarrow |f(x)g(x) - LM| < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon,$$

proving (iii). This final step is justified as so: (4) showed that

$$|f(x)g(x) - LM| \le |f(x)||g(x) - M| + (|M|+1)|f(x) - L|.$$

Since $|f(x)| < |L| + 1$ from (5), and $|g(x) - M| < \dfrac{1}{2}\varepsilon\left(\dfrac{1}{|L|+1}\right)$ from (6),

$$|f(x)||g(x) - M| < (|L|+1)\left[\frac{1}{2}\varepsilon\left(\frac{1}{|L|+1}\right)\right] = \frac{1}{2}\varepsilon.$$

Since $|f(x) - L| < \frac{1}{2}\varepsilon\left(\frac{1}{|M|+1}\right)$,

$$(|M|+1)|f(x) - L| < (|M|+1)\left[\frac{1}{2}\varepsilon\left(\frac{1}{|M|+1}\right)\right] = \frac{1}{2}\varepsilon,$$

justifying the final step.

(iv) $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{L}{M}$

To prove (iv), we take advantage of the fact that

$$\lim_{x\to a}\left[\frac{f(x)}{g(x)}\right] = \lim_{x\to a}\left[f(x)\cdot\frac{1}{g(x)}\right]$$

due to (iii), so all that is required is to show that

$$\lim_{x\to a}\left[\frac{1}{g(x)}\right] = \left[\frac{1}{M}\right].$$

If $g(x) \neq 0$, then

$$\left|\frac{1}{g(x)} - \frac{1}{M}\right| = \frac{|g(x)-M|}{|M||g(x)|} \tag{1}$$

Let $\delta_1 > 0$ where

$$0 < |x-a| < \delta_1 \Rightarrow |g(x) - M| < \tfrac{1}{2}\ |M| \tag{2}$$

so that

$$|g(x)| > \tfrac{1}{2}|M|$$
$$\frac{1}{|g(x)|} < \frac{2}{|M|}$$

which means

$$\left|\frac{1}{g(x)} - \frac{1}{M}\right| < \frac{2}{|M|^2}|g(x)-M|. \tag{3}$$

From (2), we know that the distance between $g(x)$ and M is smaller than $\frac{1}{2}\ |M|$. This means that $\frac{1}{2}|M| < |g(x)| < \frac{3}{2}|M|$, so $|g(x)| > \frac{1}{2}|M|$. (3) is achieved by substituting $\dfrac{1}{|g(x)|}$ on the right hand side of (1) with $\dfrac{2}{|M|}$.

Let $\varepsilon > 0$ be provided and $\delta_2 > 0$ such that

$$0 < |x-a| < \delta_2 \Rightarrow |g(x) - M| < \frac{|M|^2}{2}\varepsilon.$$

If $\delta = \min(\delta_1, \delta_2)$, then

$$0 < |x-a| < \delta \Rightarrow \left|\frac{1}{g(x)} - \frac{1}{M}\right| < \frac{2}{|M|^2}\cdot\frac{|M|^2}{2}\varepsilon = \varepsilon,$$

the final step coming from substituting $|g(x) - M|$ in (3) with $\dfrac{|M|^2}{2}\varepsilon$. This concludes the proof of (iv) and ultimately Theorem 2 in its entirety. It is interesting seeing how the limit laws are, in essence, results of brief sequences of subtle algebraic manipulations.

The final application of the epsilon-delta limit that will be analyzed is the classic **squeeze theorem**.

Theorem 3 - Squeeze Theorem

If there exists a $p > 0$ where

$$g(x) \leq f(x) \leq h(x)$$

for all x satisfying the inequalities $0 < |x - a| < p$, and if $\lim_{x \to a} g(x) = L$ and $\lim_{x \to a} h(x) = L$, then

$$\lim_{x \to a} f(x) = L$$

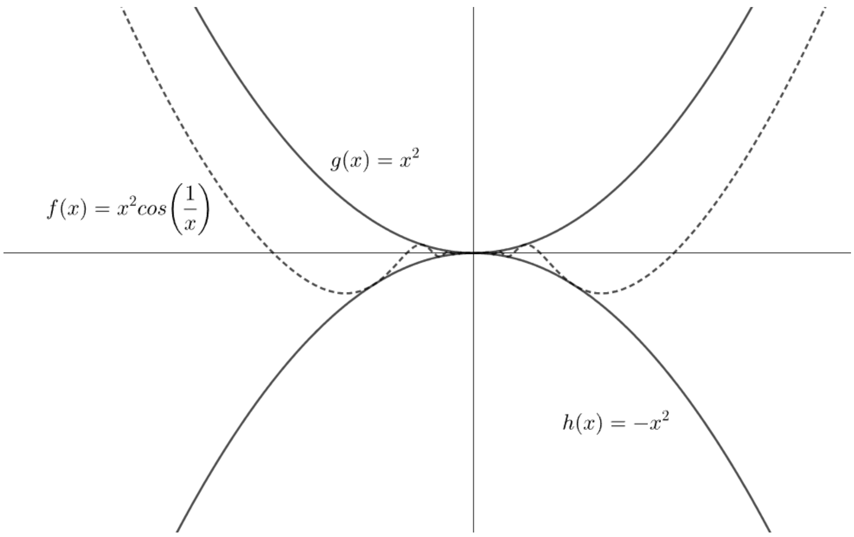

Figure 3: Squeeze theorem visualized

If a function $f(x)$ is bounded between two other functions $g(x)$ and $h(x)$, and $g(x)$ and $h(x)$ both approach the same limit L as $x \to a$, then $f(x)$ is "squeezed" toward L. For example, consider the functions $f(x) = x^2 \cos(\frac{1}{x^2}), g(x) = x^2$, and $h(x) = -x^2$ displayed in Figure 3. We can show that $x^2 \cos(\frac{1}{x^2})$ lies in between x^2 and $-x^2$ like so:

$$-1 \leq \cos\left(\frac{1}{x^2}\right) \leq 1$$

$$-x^2 \leq x^2 \cos\left(\frac{1}{x^2}\right) \leq x^2.$$

Since $\lim_{x \to 0} -x^2 = \lim_{x \to 0} x^2 = 0$, the squeeze theorem tells us that $\lim_{x \to 0} x^2 \cos(\frac{1}{x^2}) = 0$.

To prove the theorem, let $\varepsilon > 0$ be provided, and choose $\delta_1, \delta_2 > 0$ where

$$0 < |x - a| < \delta_1 \Rightarrow L - \varepsilon < g(x) < L + \varepsilon$$

and

$$0 < |x - a| < \delta_2 \Rightarrow L - \varepsilon < h(x) < L + \varepsilon$$

Let $\delta = \min(p, \delta_1, \delta_2)$. Then

$$0 < |x - a| < \delta \Rightarrow L - \varepsilon < g(x) \leq f(x) \leq h(x) < L + \varepsilon$$

$$L - \varepsilon < f(x) < L + \varepsilon$$

$$|f(x) - L| < \varepsilon$$

and thus the theorem is proven.

As a whole, we see how real analysis employs a delicate system of distances and algebraic manipulations in order to concisely define what exactly a limit is and how its definition can be used to prove related theorems. This system is then used as a base to build up toward various other ideas in calculus including differentiation and integration. Such details will not be covered, as
1) the goal here is to merely gain a baseline understanding of the nature of the argumentation present in real analysis and 2) they warrant a deeper level of study far beyond the scope of my efforts. We will, however, get to see how the next field of analysis approaches various ideas of calculus without the concept of limits.

The Nonstandard Approach

One drawback with the real analysis approach involving the epsilon-delta definition of the limit is its cumbersome notation. Proofs can quickly become cluttered with absolute values and inequalities that at times can be challenging to keep track of and connect to one another. By formalizing the idea of the infinitesimal to where basic algebraic techniques can be applied to it, a new system of analysis can be developed to study the same material that real analysis does while taking advantage of simpler and more concise notation.

Introduction

This introduction follows section 1.4 of

Keisler, H. (1976). *Foundations of Infinitesimal Calculus*, Prindle Weber & Schmidt.

Let us first consider two points on the parabola $f(x) = x^2$. One will be a fixed point at (x_0, y_0), while the other will lie some distance away at $(x_0 + \Delta x, y_0 + \Delta y)$ where Δx and Δy represent the horizontal and vertical distances between the two points.

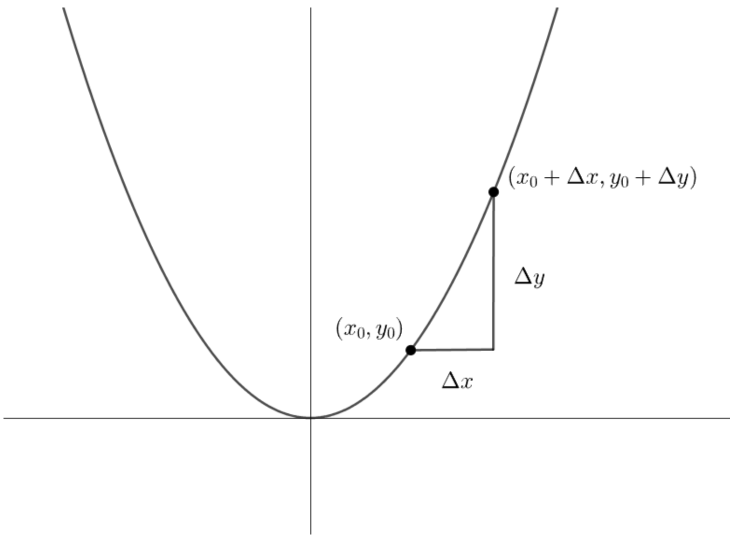

Figure 4: (x_0, y_0) and $(x_0 + \Delta x, y_0 + \Delta y)$ on $f(x) = x^2$

The average slope between any two points (x_1, y_1) and (x_2, y_2) is the ratio of the change in y to the change in x:

$$\frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Therefore, the average slope between the two points on the parabola above is

$$\frac{(y_0 + \Delta y) - y_0}{(x_0 + \Delta x) - x_0}. \tag{1}$$

The function is $f(x) = x^2$, so any y value is determined by plugging in its corresponding input x into it. The points then become (x_0, x_0^2) and $\left(x_0 + \Delta x, (x_0 + \Delta x)^2\right)$. Substituting in these new y coordinates into (1) gives

$$\begin{aligned}\frac{\Delta y}{\Delta x} &= \frac{(x_0 + \Delta x)^2 - x_0^2}{(x_0 + \Delta x) - x_0} \\ &= \frac{\left[x_0^2 + 2x_0 \Delta x + (\Delta x)^2\right] - x_0^2}{(x_0 + \Delta x) - x_0} \\ &= \frac{2x_0 \Delta x + (\Delta x)^2}{\Delta x} \\ &= 2x_0 + \Delta x. \qquad (2)\end{aligned}$$

(2) gives us the slope in terms of a fixed point's x-coordinate and its horizontal distance Δx from some variable point. If we want to find the slope of a line that lies **tangent** to the curve – meaning that it intersects the curve at exactly one point – at (x_0, x_0^2), then we can treat Δx as a very small number so that the tangent line's slope is very close to (2). We *cannot* set Δx equal to zero, because (2) follows from a quotient where Δx is present in the denominator of a quotient. Anyhow, treating it as a minisule value leads to the intuitive result that the slope of some line tangent to the point (x_0, x_0^2) on the parabola is essentially

$$2x_0.$$

This loose usage of an infinitesimal – treating a value as infinitely small to where it can be ignored – has in fact produced decently accurate results throughout the development of calculus over the past several centuries, being used by those such as Newton, Leibniz, Euler, and others in varying forms. However, it is by no means rigorous and does not establish results with absolute certainty. Why? This is because we are left with a problem: How do we know exactly when numbers are small enough to be treated as negligible in a calculation?

This now brings us to our first formal definition:

> An **infinitesimal** is a number ε where
>
> $$-a < \varepsilon < a$$
>
> for any positive real number a.

Amongst the real numbers, 0 would be the only value that qualifies as infinitesimal. What we now do is expand the real number system by introducing **hyperreal numbers**. These include the real numbers along with nonzero infinitesimals, which can be thought of as numbers that lie infinitely close to 0. The set of all real numbers is represented by $\mathbb{R}$, while the set of all hyperreal numbers is represented by $\mathbb{R}^*$. Various symbols including $\Delta x, \Delta y, \varepsilon$, and δ are used to symbolize infinitesimals. For instance, $x + \Delta x$ is interpretted as some quantity that lies *infinitely* close to x, and $\frac{1}{\varepsilon}$ represents an **infinite positive number**. Hyperreal numbers that are not infinitely positive or negative are **finite numbers**.

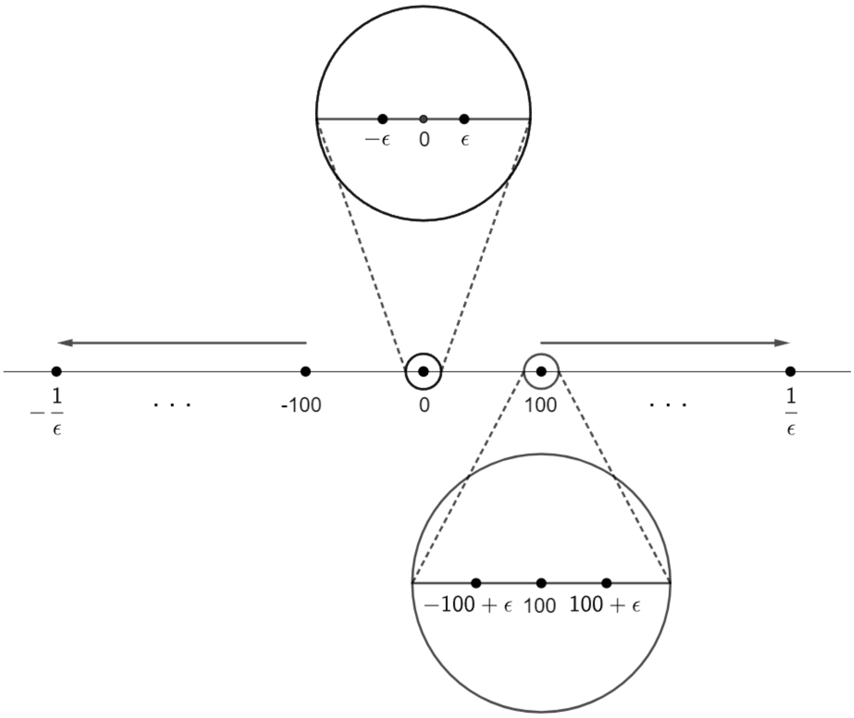

Figure 5: Hyperreal number line visualized

We can conceptualize the scale of hyperreal numbers by imagining ourselves "zooming in" infinitely close onto on a real number such as 0 or 100 on the hyperreal number line (Figure 5). This helps to clarify the nature of hyperreal numbers as merely an extension of the real numbers, so much so to where its properties and the arithmetic involved with them match exactly with that of the reals. This approach to analysis is known as **nonstandard analysis**.

Going back to the calculation of the instantaneous slope of $f(x) = x^2$, we now treat the value of Δx as a nonzero infinitesimal under this new lense of analysis. This means that the expression

$$2x_0 + \Delta x$$

lies infinitely close to $2x_0$, so the slope of any line tangent to (x_0, x_0^2) is concluded to be $2x_0$.

Let's use infinitesimals to try computing the instantaneous slope of the function

$$f(x) = \sqrt{x}.$$

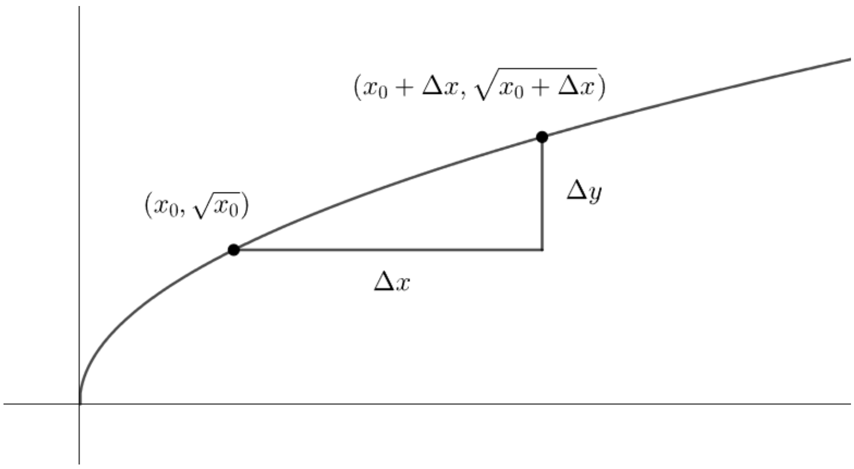

Figure 6: $f(x) = \sqrt{x}$

The average slope between a fixed point and a variable point on the square root function is

$$\begin{aligned}\frac{\Delta y}{\Delta x} &= \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \\ &= \frac{\sqrt{x_0 + \Delta x} - \sqrt{x_0}}{\Delta x}\end{aligned}$$

In order to calculate the instaneous slope, we need a way to cancel out Δx from the denominator. This can be readily accomplished by multiplying the quotient by the conjugate of the numerator in the form of 1 to induce a difference of squares:

$$\begin{aligned}\frac{\sqrt{x_0 + \Delta x} - \sqrt{x_0}}{\Delta x} \cdot \left(\frac{\sqrt{x_0 + \Delta x} + \sqrt{x_0}}{\sqrt{x_0 + \Delta x} + \sqrt{x_0}}\right) &= \frac{\left(\sqrt{x_0 + \Delta x}\right)^2 - \left(\sqrt{x_0}\right)^2}{\Delta x\left(\sqrt{x_0 + \Delta x} + \sqrt{x_0}\right)} \\ &= \frac{(x_0 + \Delta x) - x_0}{\Delta x\left(\sqrt{x_0 + \Delta x} + \sqrt{x_0}\right)} \\ &= \frac{\Delta x}{\Delta x\left(\sqrt{x_0 + \Delta x} + \sqrt{x_0}\right)} \\ &= \frac{1}{\sqrt{x_0 + \Delta x} + \sqrt{x_0}}\end{aligned}$$

Since the Δx in the denominator is infinitesimal, this means that the quantity $\sqrt{x_0 + \Delta x}$ is infinitely close to $\sqrt{x_0}$, so

$$\begin{aligned}\frac{1}{\sqrt{x_0 + \Delta x} + \sqrt{x_0}} &= \frac{1}{\sqrt{x_0} + \sqrt{x_0}} \\ &= \frac{1}{2\sqrt{x_0}}.\end{aligned}$$

Those who have experience with calculating the derivatives of functions using the limit definition of the derivative will be quick to notice that these calculations play out mostly the same way with the exception of the absence of the limit.

At this stage, the skeptical reader may question how exactly this system has solved the concern regarding rigor. In particular, how do the ideas presented so far formalize the concept of deeming a quantity in a given calculation as neglible? What has happened so far is that we have *defined* what it means for a number to be considered negligibly small, so if a given quantity does not meet the criteria specified, then we know that it cannot be ignored. Beyond that, however, we have done nothing more than explore some basic ideas of nonstandard analysis, and rigorous justifications require a working knowledge of abstract algebra. The details are obviously left out, as they lie far beyond my expertise. While the curious reader may feel discouraged or unfulfilled by this, keep in mind that the formalization of many concepts require complicated machinery, so it is commonplace for students of mathematics to take a result at face value initially and uncover its logical justification far later. As a direct analogue to the current situation, consider the analysis behind limits covered in the first half of this paper. It is likely that most undergraduates who have studied limits will never encounter this rigorous treatment of them in their lifetime – let alone give it any consideration. Ambitious students, however, will inevitably run across it in a real analysis course, and the formalities are unveiled then. We now lie in the very same boat with infinitesimals!

Basic Applications of Infinitesimals

This section follows

> Stroyan, Keith. (2012). *A Brief Introduction to Infinitesimal Calculus*, can be found at https://homepage.math.uiowa.edu/~stroyan/InfsmlCalculus/Lecture1/Lect1.pdf.

It is due time for us to see how infinitesimal logic can be used to provide basic intuitive proofs of a few ideas seen in calculus.

> **Theorem 4 - The Extreme Value Theorem**
> Consider a function $f(x)$ that is continuous on a closed interval $[a, b]$. There then exist two numbers x_{min} and x_{max} at which $f(x)$ achieves its minimum and maximum values respectively where
>
> $$f(x_{\text{min}}) \leq f(x) \leq f(x_{\text{max}})$$
>
> for all other x's.

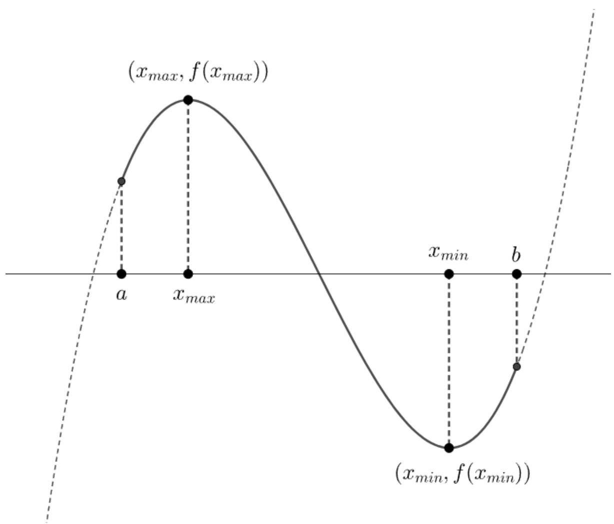

Figure 7: Extreme value theorem visualized

The extreme value theorem asserts that if a function smoothly travels over an interval of inputs, then it must take on a maximum and a minimum value at some inputs x_{max} and x_{min}.

Continuity entails that the outputs of a function are close to one another if the inputs are close. In other words, a small change in the input should only ever produce a proportionately small change in the output. Mathematically speaking, a function $f(x)$ is continuous on a closed interval $[a, b]$ only when

$$a \leq x_1 \approx x_2 \leq b \Longrightarrow f(x_1) \approx f(x_2)$$

where the $\approx$ symbol means that two quantities are *approximately* equal (very close to one another).

With this property fleshed out, we can now go about providing an intuitive argument for the extreme value theorem. We start by dividing up the interval $[a, b]$ into small increments like so:

$$a < a + \frac{b-a}{H} < a + \frac{2(b-a)}{H} < \cdots < a + \frac{k(b-a)}{H} < \cdots < b,$$

where $b - a$ is the length of the interval, H is the number of parts the interval is partitioned into, and k is a positive integer. Amongst the partition points, $f(x)$ achieves some maximum value at one (or possibly more than one) of the partition points $x_M = a + \dfrac{k(b-a)}{H}$, so

$$f(x_M) \geq f(x_1) \text{ for any } x_1 = a + \frac{j(b-a)}{H}, j \neq k.$$

We know that any input x within the interval $[a, b]$ lies within a distance of $\frac{b-a}{2H}$ from one of the partition points. To see why, consider the closed interval [2,12] divided into 5 equal subintervals ($H = 5$) so that the partition points are

$$2, \quad 2+\frac{12-2}{5}, \quad 2+\frac{2(12-2)}{5}, \quad 2+\frac{3(12-2)}{5}, \quad 2+\frac{4(12-2)}{5}, \quad 2+\frac{5(12-2)}{5}$$
$$= 2, \ 4, \ 6, \ 8, \ 10, \ 12.$$

Since the size of the interval is 10, and it is being partitioned into 5 subintervals, the distance between each partition point is

$$\frac{b-a}{H} = \frac{12-2}{5} = 2.$$

and

$$\frac{b-a}{2H} = 1.$$

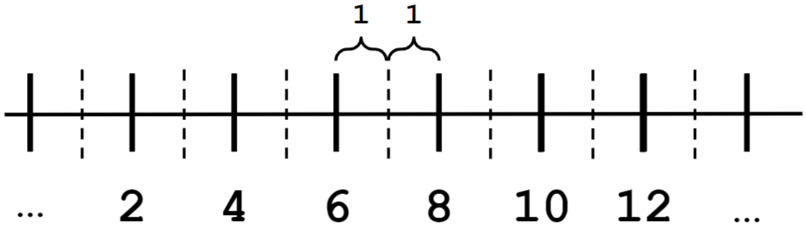

Figure 8: Partition points visualized

If we think of the quantity $\frac{b-a}{2H}$ as half of the distance between consecutive partition points and imagine a one-dimensional radius of $\frac{b-a}{2H}$ around each partition, then we can see that these radii cover all of the values between the partition points. Therefore, any value of x within the interval must lie within a distance of $\frac{b-a}{2H}$ from one of the partition points.

If we now let H grow exceedingly large to the point where $\frac{1}{H}$ becomes infinitesimal, then the set of partition points will eventually "fill up" the rest of the interval $[a, b]$ so that any value of x within the interval lies infinitely close to one of the partition points. In others words,

$$x \approx x_1 \text{ for any } x \text{ in } [a, b],$$

which means

$$f(x_M) \geq f(x_1) \approx f(x),$$

and this gives us the approximate maximum of $f(x)$. A similar argument can then be made for the approximation of the function's minimum, and this concludes the intuitive argument for the extreme value theorem, although this is by no means a formal proof.

We now employ nonstandard analysis to prove the incredible fundamental theorem of calculus, which relates a function's area under its curve to its antiderivative. Before we can do this, however, we must first establish the following result:

> Consider a small change in input δx for some function $f(x)$ that is continuous some interval. The corresponding small change in the output can be expressed like so:
>
> $$f(x+\delta x) - f(x) = f'(x) \cdot \delta x + \varepsilon \cdot \delta x,$$

where $f'(x)$ is the derivative of $f(x)$ and $\varepsilon \approx 0$.

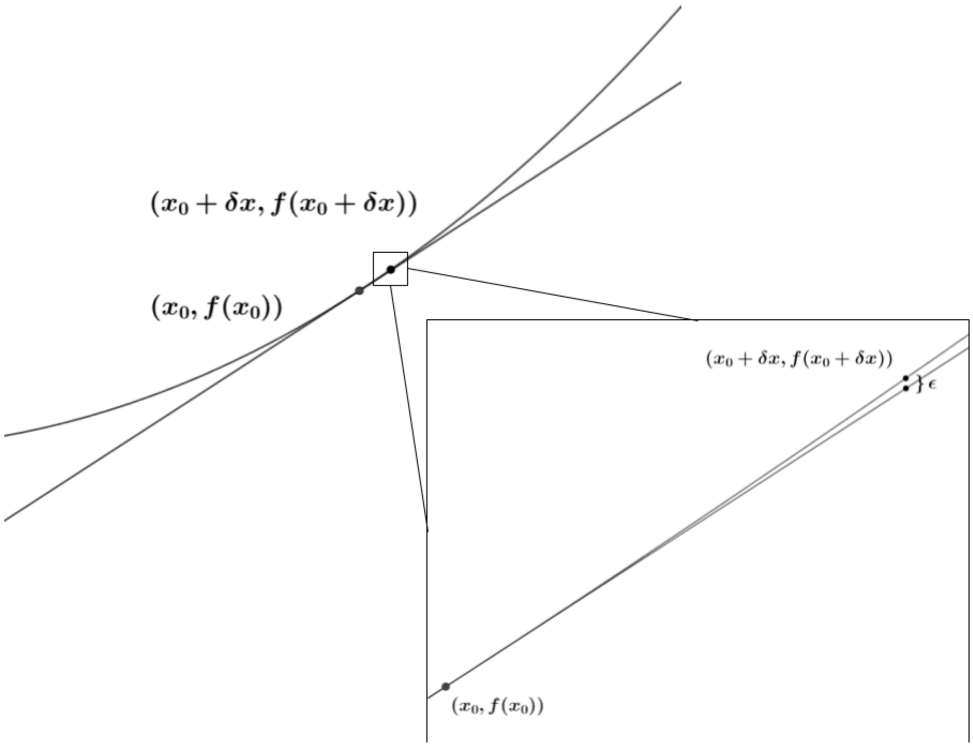

Figure 9: Tangent line visualization

A few small things to note:

1) Δ is typically used represent a small difference between two quantities, while δ is typically used to represent an infinitesimal value. The difference is merely semantical, as both represent the same core idea.

2) In Figure 9, $f'(x_0)$ does not represent the actual equation of the tangent line but merely serves as a label showing that the tangent line has a slope of $f'(x_0)$.

3) In Figure 9, ε is not intended to represent the vertical distance between the tangent line and $f(x_0)$. It's meaning will be clarified shortly.

The result can quickly be shown by starting with the limit definition of the derivative at $(x_0, f(x_0))$,

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

As a reminder, this says that the slope of the tangent line at $(x_0, f(x_0))$ is given by the limit of the slope of the secant line between $(x_0, f(x_0))$ and $(x_0 + \Delta x, f(x_0 + \Delta x))$ as $\Delta x \to 0$, where a secant line is simply a line that intersects a curve at least twice.

We can see in Figure 9 how the slope of the tangent line slightly differs from the slope of the secant line. Although the secant line is omitted from the figure, this is certain to be the case, because if they were equal, then the gap marked by ε would not exist[1]. Therefore, we can say that

$$\begin{aligned} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} &= f'(x_0) + \varepsilon \\ f(x + \Delta x) - f(x) &= f'(x_0) \cdot \Delta x + \varepsilon \cdot \Delta x, \end{aligned} \tag{1}$$

where ε represents the small difference in slope between the secant and tangent lines. It can also be thought of as the error inherent in the limit from the slope of the tangent line.

[1]It is possible for this gap to not be present, but $f(x)$ would have to be linear. This would result in ε being 0.

As $\Delta x \to 0$, the slope of the secant line approaches that of the tangent line, so $\varepsilon \to 0$ as well. From an infinitesimal perspective, both values are infinitesimal or approximately zero, and the overarching idea applies broadly to any x where $f(x)$ is continuous, so (1) can be rewritten more generally as

$$f(x + \delta x) - f(x) = f'(x) \cdot \delta x + \varepsilon \cdot \delta x, \tag{2}$$

where ε and δx are both infinitesimal. (2) tells us that a nonlinear change – the left side – is equal to the sum of i) a linear change, $f'(x) \cdot \delta x$, and ii) the product of the error and the change in x, and this product is even smaller than δx.

We are now ready to prove the first fundamental theorem of calculus.

Theorem 5 - The Fundamental Theorem of Calculus

For a function $f(x)$ that is continuous over some closed interval [a,b], if another function $F(x)$ exists where

$$dF(x) = F'(x)dx = f(x)dx$$

for all x in [a,b], then

$$\int_a^b f(x)dx = F(b) - F(a).$$

If $F(x)$ is an **antiderivative** of $f(x)$, or $\frac{d}{dx}F(x) = f(x)$, over the interval $[a, b]$, then the integral of $f(x)$ on said interval is simply the difference between $F(x)$ evaluated at $x = b$ and $x = a$.

As a reminder, a **Riemann sum** is an approximation of the area under a curve via a division into small rectangles. These rectangles are obtained by first partitioning the interval $[a, b]$ into n subintervals where each partition point x_k is given by

$$x_k = a + \frac{k(b-a)}{n},$$

where $0 \leq k \leq n$ and k is a positive integer.Each subinterval acts as a base for a rectangle, and the rectangles' heights depend on $f(x)$. For example, if the interval ()Riemann sums are essentially summations ($\sum$), so n acts as an **increment**. For the proof, we will be focused on a variation of this concept known as the **left Riemann sum**, where the height of each rectangle is determined by the left endpoints of the subintervals.

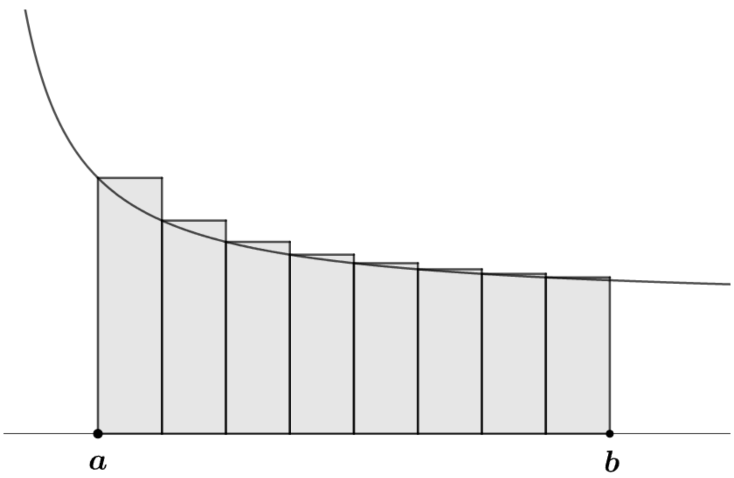

Figure 10: Riemann sum visualized

The integral is approximated by the left Riemann sum of $f(x)$ from a to b where the increment is an infinitesimal δx:

$$\int_a^b f(x)dx \approx \sum_{\substack{x=a \\ \text{step } \delta x}}^{b-\delta x} f(x) \cdot \delta x, \text{ where } \delta x \approx 0.$$

In this case, one can imagine the visual shown in Figure 10 but instead with the area divided up into an infinitely large number of rectangles. Notice how the summation only increments from a to $b - \delta x$, as $f(b - \delta x)$ corresponds to the height of the final rectangle toward the right of Figure 10.