Calculus Using Infinitesimals

Newton developed calculus using infinitesimal changes in the values of variables with relationships between them. Leibniz "formally proved it" using limits of functions. People like engineers and physicists who want to do quick and dirty and intuitive calculus tend to do the infinitesimal method. School teaches the limit method. The goal here is to introduce the infinitesimal method.

Limits: $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$.

Infinitesimals: Let x and y be variables. This topic is about *rates of change*. So we will change x by increasing it by some amount represented by δx. It should be noticed that values of x and δx live in completely different domains. The variable x takes on some *real value* like $3.23432\ldots$. In contrast δx will be a variable that takes on an *infinitesimal value* that is smaller than small. This value comes from some quantum mechanics fairy land. Hence $x + \delta x$ is not significantly bigger than x. On the other hand, if the value of y depends on the value of x with some relationship like $y = x^2$, then this small change δx causes some infinitesimal change δy in y. What we care about is the relationship between them. Set some value for x, y, and δx, then we can see that $\delta y = 2x\delta x$. What is interesting is that though the infinitesimal values δx and δy are too small to be of significant in our real world, the ratio $\frac{\delta y}{\delta x}$ often is a real value in this world like 5. Dividing through the last equation by δx gives $\frac{\delta y}{\delta x} = 2x$.

Derivative = Slope: The slope of the curve y vs x at the point $\langle x, y\rangle$ is $slope = \frac{rise}{run} = \frac{change\ in\ y}{change\ in\ x} = \frac{\delta y}{\delta x}$. We call this the *derivative of y with respect to x*.

Functions: A function f is a mapping $f(x) = x^2$ from the variable x to the value $f(x)$. Similarly, $g(z) = e^z$. Then the composition $g(f(x))$ gives a function from x to $g(f(x)) = e^{x^2}$.

Variables: It is also useful to think of f and g as variables that take on the value $f(x)$ and $g(f(x))$, namely $f = x^2$ and $g = e^f = e^{x^2}$.

Proof of Sum Rule Using Infinitesimals: Let $s(x) = f(x) + g(x)$ be the sum of the two amounts $f(x)$ and $g(x)$. The *sum rule* states that $s'(x) = f'(x) + g'(x)$, i.e. the derivative of the sum is the sum of the derivatives. We will now prove this using infinitesimals. Switching to the variable view gives $s = f + g$. If x increases by δx, then f increases by δf and g increases by δg. This first change increases the sum $s = f + g$ by δf and the second by δg. The total change to the sum is $\delta s = \delta f + \delta g$. Dividing through by δx gives $\frac{\delta(f+g)}{\delta x} = \frac{\delta s}{\delta x} = \frac{\delta f}{\delta x} + \frac{\delta g}{\delta x}$. This completes the proof. As an example $\frac{\delta(e^x+x)}{\delta x} = \frac{\delta(e^x)}{\delta x} + \frac{\delta x}{\delta x} = e^x + 1$.

Proof of Product Rule Using Infinitesimals: Let $a(x) = f(x) \times g(x)$ be the area of the square that is $f(x)$ wide and $g(x)$ high. The *product rule* states that $a'(x) = f(x) \times g'(x) + f'(x) \times g(x)$. We will now prove this using infinitesimals. Switching to the variable view gives $a = f \times g$. If x increases by δx, then f increases by δf and g increases by δg. Draw the square with this infinitesimal increase on the right and top side. See how the area increases by $\delta a = f\delta g + g\delta f + \delta f\delta g$. Besides looking at the area, another way of seeing this is to make the changes one at a time. If f was a constant, but g increases by δg, then $a = f \times g$ would increase by $\delta a = f \times \delta g$. If g was a constant, but f increases by δf, then $a = f \times g$ would increase by $\delta a = \delta f \times g$. Making both changes would increase a by $\delta a = f\delta g + g\delta f$ (and ok maybe an extra $\delta f\delta g$).

Infinitesimals Squared: Recall that an infinitesimal like δf is so small that it doesn't make a significant change in f. Recall that $0.01 \times 0.02 = 0.0002$ giving us that multiplying two really small numbers gives a really really small number. A infinitesimals times an infinitesimal not only makes no significant change in our world of values like f, it also only makes no significant change in the world infinitesimals like δf. The infinitesimal δf only has meaning in our world when divided by another infinitesimal like δx, giving the real valued derivative $\frac{\delta f}{\delta x} = 2x$. The *infinitesimal infinitesimal* $\delta f\delta g$ only has meaning in our world when divided by another infinitesimal infinitesimal like δx^2, giving the real value $\frac{\delta f\delta g}{\delta x^2}$. We call a change in a like $f\delta g$ a *first order* change an a change like $\delta f\delta g$ a *second order* change. Here we know that we are never going to divide by infinitesimal infinitesimals so we can ignore the change $\delta f\delta g$.

We simplify the change in a to just the first order changes $\delta a = f\delta g + g\delta f$. Dividing through by δx gives $\frac{\delta(f\times g)}{\delta x} = \frac{\delta a}{\delta x} = f\frac{\delta g}{\delta x} + g\frac{\delta f}{\delta x}$. This completes the proof. As an example $\frac{\delta(x\times x)}{\delta x} = x\cdot\frac{\delta x}{\delta x} + x\cdot\frac{\delta x}{\delta x} = x\cdot 1 + x\cdot 1 = 2x$.

Proof of Chain Rule Using Infinitesimals: Consider three gears labeled x, f, and g with 400, 100, and 20 teeth. If x turns by δx rotations then f turns by $\delta f = 4\delta x$ rotations or $\frac{\delta f}{\delta x} = \frac{400}{100} = 4$. If f turns by δf rotations then g turns by $\delta g = 5\delta x$ rotations or $\frac{\delta g}{\delta f} = \frac{100}{20} = 5$. It follows that if x turns by δx rotations then g turns by $\delta g = 4\cdot 5\delta x = 20\delta x$ rotations. The *chain rule* gives

$$\frac{\delta g}{\delta x} = \frac{\delta g}{\delta f} \times \frac{\delta f}{\delta x} = 5\cdot 4 = 20.$$

and is proved by simply write down what we want and what we have and cancel infinitesimals as needed.

Now consider our previous example. We know that if $f(x) = x^2$, then $f'(x) = 2x$ and that if $g(z) = e^z$, then $g'(z) = e^z$. One uses the *chain rule* to take the derivative of the composition $g(f(x)) = e^{x^2}$, namely $(g(f(x)))' = g'(f(x))\cdot f'(x) = e^{x^2}\cdot 2x$. We will now prove this using infinitesimals. Switching to the variable view gives that $f = x^2$, $\frac{\delta f}{\delta x} = 2x$, $g = e^f$, $\frac{\delta g}{\delta x} = e^f$, and composing $g = e^{x^2}$. Chaining things together, if x changes by δx, then this causes the change δf in f, which in turn causes the change δg in g. The chain rule wants to understand how this change δg in g compares with this initial change δx in x. The proof is easy. If you know that one is traveling $60\frac{km}{hr}$ for $2hrs$, then one gets the number of km the person has traveled by multiplying so that the units cancel as needed, namely $distance = speed \times time = 60\frac{km}{hr} \times 2hrs = 120km$. Similarly, knowing how the change δg in g in how it compares to the change δf in f and knowing how the change δf in f in how it compares to the change δx in x, we do what we did with the gears, $\frac{\delta g}{\delta x} = \frac{\delta g}{\delta f} \times \frac{\delta f}{\delta x}$. In our case, $\frac{\delta g}{\delta x} = e^f \times 2x = e^{x^2} \times 2x$.

$y = e^x$: You likely know that e^x is defined to be $\overbrace{e\times e\times e\times\ldots\times e}^{x}$ where $e = 2.7182818\ldots$. But where does that come from. Suppose that y represents the value of your bank account at time x. Suppose that due to interest, after each year the current value of y increases by the current value at the beginning of the year, namely $y(x+1) = y(x) + y(x) = 2y(x)$. The \$1 turns into \$2, \$4, \$8, … after the years go by to give the value $y = 2^x$ after x years. If after one year your bank account value y increase by y, then after δx years, it should increase proportionally by $y\delta x$, namely $y(x+\delta x) = y(x) + y(x)\delta x$. The change in y is $\delta y = y\delta x$. Dividing through gives the differential equation $y' = \frac{\delta y}{\delta x} = y$. When we get our interest every δx time intervals instead of every year, the interest compounds faster, getting interest on the interest faster which makes your account grow faster. Instead of growing to 2^x, it grows to e^x where e is a little bigger than 2, specifically is worked out to be $e = 2.7182818\ldots$. Look at the plot of $y = e^x$ and verify that the slope at $\langle x, y\rangle$ is y. As y grows, this slope grows. In fact, the tangent line through this point slopes so quickly that it goes down to the point $\langle x-1, 0\rangle$ and up to the point $\langle x+1, 2y\rangle$.

Formally, we want
$e^x = y' = \lim_{h\to 0}\frac{e^{x+h} - e^x}{h} = \lim_{h\to 0} e^x \times \frac{e^h - 1}{h}$.
Dividing through by e^x, gives that $\lim_{h\to 0}\frac{e^h - 1}{h} = 1$.
Solving for h gives $e^h - 1 = h$, $e^h = 1 + h$, and
$e = \lim_{h\to 0}(1+h)^{1/h}$.
This is the formal definition of e.

For example if $1/h = 10$, then $(1+h)^{1/h} = (1.1)^{10} = 2.59$. Similarly $1.05^{20} = 2.65$, $1.01^{100} = 2.704$.

This value of e is the result of the tension between two effects. The first effect is that when $h = 0$, then $(1+h)^n = 1^n = 1$ for every value n. The second effect is that when h goes to zero, $1/h$ goes to infinity and so $(1 + 0.000001)^{1/h}$ also goes to infinity.

$y = c^x$: We take the derivative of exponentials with a different base as follows. $(c^x)' = \left(\left(e^{\ln(c)}\right)^x\right)' = \left(e^{\ln(c)x}\right)' = e^{\ln(c)x}\cdot\ln(c) = \ln(c)\cdot c^x$.

Or $(x^x)' = \left(\left(e^{\ln(x)}\right)^x\right)' = \left(e^{\ln(x)x}\right)' = e^{\ln(x)x} \cdot (\frac{x}{x} + \ln(x)) = (\ln(x)+1) \cdot x^x$.

The Derivative of $y = \ln(x)$: Instead of remembering that Christ was 2014 years ago, homo erectus appeared 1 million years ago, life appeared 4.5 billion years ago, and the universe appeared 13.8 billion years old, it is easier to remember that these numbers have 3, 6, 9 and 10 zeros. We use the log function for this. $\log_{10}(1000) = 3$. Instead of a base of 10, computer science uses a base of 2 and science uses a base of e. We write $y = \log_e(x) = \ln(x)$, pronounced log and lawn. This is the inverse function of $x = e^y$. Though we have switched the role of x and y, we still know that if $x = e^y$, then $\frac{\delta x}{\delta y} = e^y$. You can do anything to one side if you do the same thing to the other side. Flipping both sides over gives $\frac{\delta y}{\delta x} = \frac{1}{e^y} = \frac{1}{x}$. Look at the plot of $y = \ln(x)$ and verify that the slope at $\langle x, y\rangle$ is $\frac{1}{x}$.

Approximating $y = f(x)$: Suppose you have strange curve defined by $y = f(x)$. Suppose you completely understand the function at some value $x = x_0$, namely you know the values $f(x_0)$, $f'(x_0)$, and maybe even $f''(x_0)$. We want to approximate the value $f(x_0 + \delta x)$ at $x_0 + \delta x$ for some small value δx. Let $F(x_0 + \delta x)$ be this approximation. The constant approximation is $F(x_0 + \delta x) = f(x_0)$ by the approximation that f does not change that quickly. The *first order* approximation has F be the tangent line through $\langle x_0, y_0\rangle$ with slope $f'(x_0)$. This equation is $F(x_0+\delta x) = f(x_0)+f'(x_0)\delta x$. The *second order* approximation has F be the quadratic function $F(x) = ax^2+bx+c$ such that $F(x_0) = f(x_0)$, $F'(x_0) = f'(x_0)$, and $F''(x_0) = f''(x_0)$. This turns out to be $F(x_0 + \delta x) = f(x_0) + f'(x_0)\delta x + \frac{1}{2}f''(x_0)\delta x^2$.

Other Rules:

- The derivative of a constant is zero because when x changes by δx, 5 changes by $\delta 5 = 0$.
- The polynomial rule is that for all constants c, we have that $(x^c)' = cx^{c-1}$.
  - With $c = 2$, $(x^2)' = 2x$.
  - With $c = 1$, $(x^1)' = x' = 1x^0 = 1$, which is good because $x' = \frac{\delta x}{\delta x} = 1$.
  - With $c = \frac{1}{2}$, $(\sqrt{x})' = (x^{1/2})' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}$.
  - With $c = 0$, $(1)' = (x^0)' = 0x^{-1} = 0$, which is good because $x^0 = 1$ which is a constant.
  - With $c = -1$, $(\frac{1}{x})' = (x^{-1})' = -1x^{-2} = -\frac{1}{x^2}$.
- $(sinx)' = cosx$ and $(cosx)' = -sinx$ as long as x is measure in radians instead of degrees so that an angle of x sweeps an arc of length x on the unit circle and the full circle has an angle of 2π.

Recursion: Trusting you friends. Build parse tree and differentiate recursively.

Two Ways: One thing I find fun is taking the derivative of something two different ways and seeing that they give the same answer. The first three could actually be used as proofs of the polynomial rule that $(x^c)' = cx^{c-1}$.

- $f^2 = f \times f$. Take the derivative of the first using the chain rule and the second using the product rule.
- $x^{c+d} = x^c \times x^d$. Take the derivative of the first as a polynomial and the second using the product rule.
- $x^{cd} = (x^c)^d$. Take the derivative of the first as a polynomial and the last using the chain rule.
- $x^c = \left(e^{\ln x}\right)^c = e^{c \ln x}$. Take the derivative of the first as a polynomial and the second using the chain rule.
- $cf(x) = c \times f(x)$. Take the derivative of the first using the fact that $(cf(x))' = cf'(x)$ and the second using the product rule.
- $\ln(cx) = \ln(x) + \ln(c)$. Take the derivative of the first using the chain rule and the second using the sum rule.

- One that has bothered me is the following. The polynomial rule gives that the derivative of x^c is a constant times x^{c-1} which means that taking the derivative shifts the power of a polynomial by one. On the other hand, $(\ln x)' = \frac{1}{x} = x^{-1}$, the later being a polynomial and the former not being. How do these mesh when $c = \epsilon$ is close to zero?

  The key is that all the constants matter. Note that because the derivative of a constant is zero that the dirivativers of $f(x) + a$ and $f(x) + b$ are the same.

  But one thing that is true is that if two cars start in the same place and aways have the same velocity (in the same direction) then they will always be in the same place. Hence if functions f and g have the same starting points, say $f(1) = g(1)$ and always have the same derivative, i.e. $f'(x) = g'(x)$ for all x, then $f(x) = g(x)$ are the same.

  Let $f(x) = \ln x$ and $g(x) = \lim_{\epsilon \to 0} \frac{x^\epsilon - 1}{\epsilon}$. You can easily confirm that $f(1) = g(1) = 0$ and that $f'(x) = g'(x) = \frac{1}{x}$. Hence, it must be the case that $f(x) = g(x)$. If you know how to do it using l'hopital's rule, confirm that $g(x) = \lim_{\epsilon \to 0} \frac{x^\epsilon - 1}{\epsilon} = \ln x = f(x)$.

- Think of others.