TO CALCULATE IN CALCULUS

A. SCHREMMER

That *calculations* make mathematics "more accessible" has long been recognized, at least since Descartes and certainly by Newton and Leibniz. Yet, limits have now become, even though they cannot be *calculated*—looking for a Skolem function $\delta = S(\epsilon)$ hardly qualifies, the rock on which the *calculus* rests ... and beginners founder.[1]

Prefatory Remarks

When thinking about how to circumvent limits, one runs invariably into $f(x) = x^2 \sin \frac{1}{x}$ where the problem is that the derivative is not continuous at 0 and the first approach that comes to mind is to restrict *calculations* to a given class of functions as, for instance, Class $P(t, n)$ functions in Levi [8] or real analytic functions in Bassein [1]. But how can we plausibly introduce a class of allowable functions *up-front*?

On the other hand, the heart of the matter may be the way we think about x_0 and $x_0 + h$ as, for all practical purposes, only $x_0 + h$ is "real": try to cut a board of length x_0 and what you will get is a board of length $x_0 + h$. In [2], Dieudonné writes that "*a real number* [x_0] *is known only when a method to approximate it* [to $x_0 + h$] *has been given (with an approximation which the mathematician wants to be arbitrarily small, whereas the user of mathematics is content with much less)*" while, in [4], Gowers never even mentions real numbers while spending a whole chapter on *infinite decimals*. This is also the distinction Leibniz made possible and Robinson legal. But, if infinitesimals do allow for *calculations*, they seem difficult to introduce "intuitively" other than the way Leibniz did it.[2] See for instance Keisler [5], Robert [10], and even Freed [3], none of which could be transparent enough to gain broad acceptance.

Nevertheless, instead of looking at the differential calculus as being "the mathematics of change", the view taken here will be, to quote Dieudonné again, that "[*in analysis*] *we often have to study the "behaviour" of a function f in a neighborhood of x_0*", i.e. $f(x_0 + h)$.[3] We start from the fact that an engineer who wanted to compute $\frac{17}{7} + \sqrt{10} + \pi$ would use formal series $\sum_0^\infty a_n (\frac{1}{10})^n$ *truncated* to some approximation and write something like

---

Date: March 27, 2015.

[1]Even if introducing limits "intuitively" gets around Skolem functions, that still does not make them *calculatable*.

[2]An interesting parallel is with the Dirac and Heaviside "functions" with which physicists never had any problem and for whom the fact that Schwartz made them acceptable to mathematicians made no difference.

[3]So, for $f(x) = x^2 \sin \frac{1}{x}$, one could say that it is defining $f(0)$ that causes all the trouble.

$\frac{17}{7} + \sqrt{10} + \pi = 2.42857 + [...] + 3.16227 + [...] + 3.14159 + [...] = 8.73243 + [...]$ with [...] standing for the small difference between a real number and its decimal approximation.[4]

Similarly, given a function, Lagrange [6], who explicitly wanted to free the differential calculus from "*any consideration of infinitesimals, vanishing quantities, limits and fluxions and reduce it to the algebraic study of finite quantities*", used formal series $\sum_0^\infty A_n(x_0)h^n$ *truncated* to local best *polynomial* approximations of $f(x_0 + h)$ from which all the usual notions can be defined and, once limits have been confined within $o[h^n]$, all the usual theorems can be proven by *calculations*. That, unlike $\sum_0^\infty a_n(\frac{1}{10})^n$, $\sum_0^\infty A_n(x_0)h^n$ does not necessarily converge is completely irrelevant: Lagrange was using **asymptotic expansions** and, to quote Dieudonné again, "*one should be careful to realize that the notion of asymptotic expansion* has nothing to do with the notion of a series." (Emphasis in the original.)

More precisely, using asymptotic expansions "*is to* compare f *to* [gauge] *functions whose behaviour near* 0 [*or* ∞] *is considered* known," that is functions that i. approach 0 or ∞ as x approaches 0 or ∞ and ii. are closed for multiplication so that iii. they are totally ordered by the relation "f is equal to, or negligible compared to, g". See the appendix. Thus, "*calculating means approximating*" and "*one must learn to distinguish what is "large" from what is "small", what is "preponderant" and what is "negligible".*"

The implementation suggested here should be acceptable *to* aspiring engineers, the majority of the students in Calculus. We describe how much of the behavior of an elementary function can be *calculated* and then discuss in that context some of the usual theorems.

Polynomial Functions

We use the positive power functions ax^{+n} as gauges and the fact that, near 0, $h^m = o[h^n]$ for $m > n$ and, near ∞, $x^m = o[x^n]$ for $m < n$. With the binomial theorem as **addition formula**, we then get $f(x_0 + h) = \sum_0^\infty A_i(x_0)h^i$ and just separate a **principal part** $\sum_0^n A_i(x_0)h^i$, simple enough to give us the local information we seek, from a **remainder** $o[h^n]$ too small to be significant in that regard. Usually, $n \leq 2$.

Best Approximations. If we try to approximate $f(x_0 + h) = \sum_0^\infty A_i(x_0)h^i$ with a *constant* function, the simplest kind of non-zero function, the error for $k(x) = K$ is $f(x_0 + h) - k(x_0 + h) = [A_0(x_0) - K] + o[1]$ and therefore of the same order of magnitude as the approximate output. But if we take $K = A_0(x_0)$, the error is then $A_1h + o[h]$, that is smaller than the approximate output by an order of magnitude. Thus, the **Best Constant Approximation** of f near x_0 is $\mathrm{BCA}f(x_0 + h) = A_0(x_0)$. Similarly, the **Best Affine Approximation** is $\mathrm{BAA}f(x_0 + h) = A_0(x_0) + A_1(x_0)h$ and the **Best Quadratic Approximation** is $\mathrm{BQA}f(x_0 + h) = A_0(x_0) + A_1(x_0)h + A_2(x_0)h^2$.

Geometry. We look at the way $f(x_0 + h)$ differs qualitatively on each side of x_0 from simpler, already known functions. Since the zero function has no height, we define **Height-sign** f near x_0 as the way, up or down, $f(x_0 + h)$ differs on each side of x_0 from the zero function. Similarly, since constant functions have no slope, **Slope-sign** f **near** x_0 is the

[4]As engineers are wont to put it, "The *real* real numbers are the decimal numbers".

way $f(x_0+h)$ differs from BCA$f(x_0+h)$ and since affine functions have no concavity, **Concavity-sign** f **near** x_0 is the way $f(x_0+h)$ differs from BAA$f(x_0+h)$.

When x_0 is **regular**, whether f is **positive** or **negative** (resp. **increasing** or **decreasing**, **concave up** or **concave down**) at x_0, i.e. whether Heigh-sign f near $x_0 = \langle +,+\rangle$ or $\langle -,-\rangle$ (resp. Slope-sign f near $x_0 = \langle \diagup,\diagup\rangle$ or $\langle \diagdown,\diagdown\rangle$, Concavity-sign f near $x_0 = \langle \cup,\cup\rangle$ or $\langle \cap,\cap\rangle$), is thus determined by the sign of the coefficient of h^0 (resp. h^1, h^2) in $f(x_0+h)$. When x_0 is **critical**, i.e. when the relevant coefficient is 0, the information comes from the sign of the next non-zero coefficient. x_0 is a **height sign-change** (resp. **slope sign-change**, **concavity sign-change**) input iff Heigh-sign f near $x_0 = \langle +,-\rangle$ or $\langle -,+\rangle$ (resp. Slope-sign f near $x_0 = \langle \diagup,\diagdown\rangle$ or $\langle \diagdown,\diagup\rangle$, Concavity-sign f near $x_0 = \langle \cup,\cap\rangle$ or $\langle \cap,\cup\rangle$).

Example 1. Let $f(x) = (x-2)^3 + x - 1$. We get from $f(+2+h) = +1 + o[1]$ that Height-sign f near $+2 = (+,+)$, from $f(+2+h) = +1 + h + o[h]$ that Slope-sign f near $+2 = \langle \diagup,\diagup\rangle$ but we need $f(+2+h) = +1 + h + h^3 + o[h^3]$ to get that Concavity-sign f near $+2 = \langle \cap,\cup\rangle$ and therefore that $+2$ is a concavity sign-change input.

We can also characterize **notable inputs**. For instance, x_0 is a **local extremum** iff the first *non-constant* term of $f(x_0+h)$ is *even* and x_0 is an **inflection** iff the first *non-linear* term of $f(x_0+h)$ is *odd*. Thus, for polynomial functions, Slope-sign f near a local extremum $= \langle \diagup,\diagdown\rangle$ or $\langle \diagdown,\diagup\rangle$ and Concavity-sign f near an inflection $= \langle \cup,\cap\rangle$ or $\langle \cap,\cup\rangle$. A **0-height** (resp. **0-slope**, **0-concavity**) input is an input whose nearby inputs have *small* height (resp. slope, concavity).[5] So, to locate 0-height (resp. 0-slope, 0-concavity) inputs, we try to solve $A_0(x_0) = 0$ (resp. $A_1(x_0) = 0$, $A_2(x_0) = 0$).

A **local graph** is the graph of a **Best Graphic Approximation** i.e. an approximation that has height, slope and concavity if any. Smoothly interpolating the local graph near ∞ gives an **essential global graph** which predicts the *existence* of **essential** notable inputs, i.e. those "visible from infinity", but not that of **fluctuations** i.e. pairs of opposite extrema separated by an inflection. In other words, polynomial functions are "essentially" positive power functions with a number of fluctuations thrown in.

Example 2. Let $Q(x) = ax^2 + bx + c$. Near ∞, $Q(x) = ax^2 + o[x^2]$ and interpolating the local graph near ∞ predicts the existence of a 0-slope extremum. From $Q(x_0+h) = [ax_0^2 + bx_0 + c] + [2ax_0 + b]h + [a]h^2$, we get that near $\frac{-b}{2a}$, the solution of $2ax_0 + b = 0$, what slope there is comes from ah^2. So, $\frac{-b}{2a}$ is the predicted extremum and the existence of 0-height inputs depends on whether Height-sign Q near $\frac{-b}{2a}$ is the opposite or the same as Height-sign Q near ∞. (Figure 1.)

Example 3. Let $C(x) = ax^3 + bx^2 + cx + d$. Near ∞, $C(x) = ax^3 + o[x^3]$ and interpolating the local graph near ∞ predicts the existence of an inflection. From $C(x_0+h) = \left[ax_0^3 + bx_0^2 + cx_0 + d\right] + \left[3ax_0^2 + 2bx + c\right]h + [3ax_0 + b]h^2 + ah^3$, we get that near $\frac{-b}{3a}$, the solution of $[3ax_0 + b] = 0$, what concavity there is comes from ah^3. So, $\frac{-b}{3a}$ is the predicted

[5]While nature abhors, say, 0° Kelvin or 0 Pascal, it has no issue with h° Kelvin or h Pascals, however small h is. And when physicists try to solve $f(x) = 0$, it is in the hope of locating those inputs *near which* $f(x)$ is small—not to say infinitesimal.

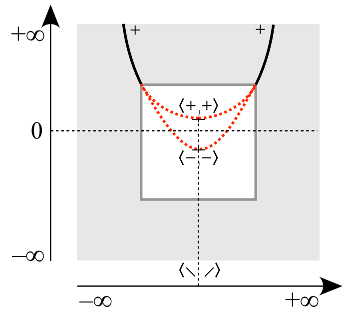

FIGURE 1. $Q(x) = ax^2 + bx + c$

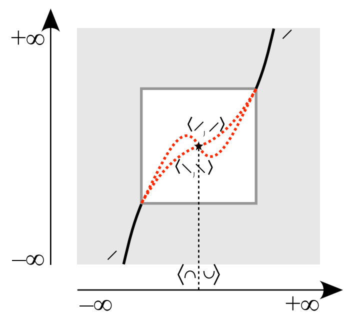

FIGURE 2. $C(x) = ax^3 + bx^2 + cx + d$

inflection and the existence of a fluctuation depends on whether Slope-sign C near $\frac{-b}{3a}$ is the opposite or the same as Slope-sign C near ∞. (Figure 2.)

LOCALLY APPROXIMATELY POLYNOMIAL FUNCTIONS

Constant (resp. affine, quadratic) functions have *globally* constant *height* (resp. *slope*, *concavity*) but what of functions that are only *locally approximately* constant (resp. *locally approximately* affine, *locally approximately* quadratic)? For instance, for $q(x) = ax^2+bx+c$, $q(x_0 + h) = A_0(x_0) + A_1(x_0)h + ah^2$ which shows that quadratic functions have constant concavity. But what of functions such that $f(x_0+h) = A_0(x_0)+A_1(x_0)h+A_2(x_0)h^2+o[h^2]$?

This is precisely where the difficulties begin. To avoid ambiguities, let us say that f is **n-Peano-differentiable** at x_0 iff f has a **Best Polynomial Approximation** of degree n. Then, for example, we obviously have

Theorem. *If f is Peano-continuous (resp. Peano-differentiable, Peano-2-differentiable)* at x_0*, then* near x_0*, $f(x)$ is positive/negative (resp. increasing/decreasing, concave up/concave down) if A_0 (resp. A_1, A_2) is positive/negative.*

But what happens when A_0 (resp. A_1, A_2) is 0?

EXAMPLE 4. The fact that the coefficient of h^2 in EXAMPLE 1 was 0 did not prevent us from getting the concavity because we then used the h^3 term. But, for $g(x) =$

$x^3 \sin \frac{1}{x}$ when $x \neq 0$, 0 otherwise, that $g(0+h) = 0 + 0h + 0h^2 + o[h^2]$ is a dead end because the concavity is in $o[h^2]$ and, as can happen with asymptotic expansions, here we cannot get an h^3 term. (But then, $f''(0)$ doesn't exist either.)

More generally, we may ask what the connection is between $A_k k!$ and the **Cauchy-derivative** $f^{(k)}(x_0)$:

- If $f^{(k)}(x_0)$ exists for $k = 0, \ldots, n$, then the Taylor polynomial $\sum_{k=0}^{k=n} f^{(k)}(x_0)h^k$ is the Best Polynomial Approximation of degree n.
- But, as the function g in Example 4 shows, the existence of a Best Polynomial Approximation of degree $k > 1$ near x_0 does not ensure the existence of any Cauchy-derivative of order > 1 at x_0.

Peano-continuity and Peano-differentiability at x_0 are local properties and one would like to study them on an *interval*. For instance, the following would seem to be rather desirable:

Theorem. A continuous function on a closed bounded interval is bounded.

But if the proof is not obvious with Cauchy-continuity, Peano-continuity does not make it any easier. However, the latter isolates the difficulty very clearly. Say f is Peano-continuous on an interval $[a, b]$, then $\forall x_0 \in [a, b], f(x_0 + h) = f(x_0) + o(1)$. Suppose h is in a neighborhood of 0 whose size depends on x_0, for example such that $o(1) < \frac{1}{10}$. If we knew that we could cover $[a, b]$ using, say, N of these intervals, then $f(x) - f(a)$ would be bounded by $\frac{N}{10}$ and the theorem proved. This raises the question as to whether, from an open covering of a closed bounded interval, we can extract a finite one.

Rational Functions

To deal with rational functions, we include negative power functions among the gauges because there now may be: i. **poles**, namely **∞-height inputs**, i.e. inputs whose nearby inputs have *large* height, and/or ii. an **asymptote** near ∞. We also need the fact that, near 0, $h^{-m} = o[h^{-n}]$ for $m < n$ and, near ∞, $x^{-m} = o[x^{-n}]$ for $m > n$.

Local Behavior. Let $f(x) = \frac{A(x)}{B(x)} = \frac{a_m x^m + \ldots + a_p x^p}{a_n x^n + \ldots + a_q x^q}$, $m \geq p$, $n \geq q$. Near ∞, from $f(x) = \frac{a_m x^m + \ldots + a_p x^p}{a_n x^n + \ldots + a_q x^q}$ the next step is obvious: to get a best *polynomial* approximation we must *divide in descending powers* until we get the local information we seek. Near x_0, we compute $f(x_0 + h) = \frac{A(x_0+h)}{B(x_0+h)} = \frac{A_0 + A_1 h + A_2 h^2 + o[h^2]}{B_0 + B_1 h + B_2 h^2 + o[h^2]}$ as before and the next step is again obvious: to get a best *polynomial* approximation we must *divide in ascending powers*. If x_0 is a pole, we get a negative power function which is a Best Graphic Approximation.

Essential global graph. We interpolate the local graphs near ∞ and near the ∞-height input(s)—if any. Odd poles can be seen as infinite fluctuations and even poles as infinite extrema. In other words, rational functions are "essentially" integral power functions with bounded/infinite fluctuations and/or bounded/infinite extrema thrown in.

Example 5. Let $f(x) = \frac{x^2+x+1}{(x-2)^2}$. i. Near ∞, we get by long division in *descending* powers that $f(x) = +1 + 5x^{-2} + o[x^{-2}]$, i.e. a horizontal asymptote. ii. Near $+2$, the only

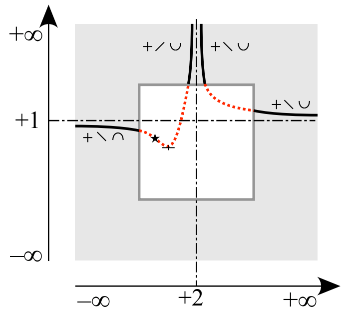

FIGURE 3. $f(x) = \frac{x^2+x+1}{(x-2)^2}$

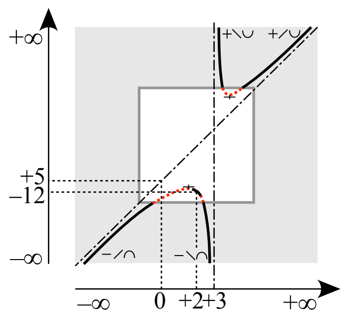

FIGURE 4. $f(x) = \frac{x^3-8}{x^2-5x+6}$

possible pole, we get by short division that $f(+2+h) = \frac{+7+o[1]}{h^2} = +7h^{-2} + o[h^{-2}]$, i.e. that $+2$ is indeed a pole. iii. A smooth interpolation of the local graphs near ∞ and $+2$ gives the essential global graph in Figure 3 which shows that f has *at least* one inflection and one minimum.

EXAMPLE 6. Let $f(x) = \frac{x^3-8}{x^2-5x+6}$. i. Near ∞, we get by long division in *descending* powers that $f(x) = x + 5 + 19x^{-1} + o[x^{-1}]$ i.e. an oblique asymptote. ii. Near $+2$, one of two possible poles, we get by long division in *ascending* powers that $f(+2+h) = \frac{+12h+6h^2+h^3}{-h+h^2} = -12 - 18h - 19h^2 + o[h^2]$ i.e. that $+2$ is regular. iii. Near $+3$, the other possible pole, we get by short division that $f(+3+h) = \frac{+19+27h+9h^2+h^3}{h+h^2} = \frac{+19+o[1]}{h+o[h]} = +19h^{-1} + o[h^{-1}]$ i.e. that $+3$ is indeed a pole. iv. A smooth interpolation of the local graphs near ∞ and $+3$—the local graph near $+2$ provides only confirmation—gives the essential global graph in Figure 4 which shows that f has *at least* one maximum and one minimum.

RADICAL FUNCTIONS

Rather that defining radical functions as functions which output irrational numbers, we look at them as solutions of functional equations so that the approximating issue is built-in.

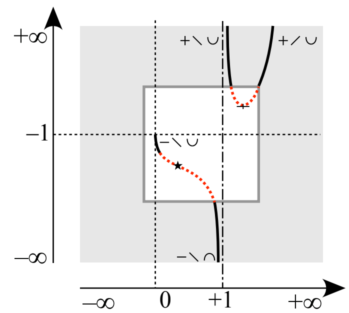

FIGURE 5. $f(x) = \frac{\sqrt{x}+e^x}{x^2-1}$

EXAMPLE 7. Let $\sqrt{\ }$ be a solution of $f^2(x) = x$. From $\sqrt{x_0+h} = A_0(x_0) + A_1(x_0)h + A_2(x_0)h^2 + o[h^2]$ and therefore $\left(\sqrt{x_0+h}\right)^2 = A_0(x_0)^2 + 2A_0(x_0)A_1(x_0)h + \left[A_1(x_0)^2 + 2A_0(x_0)A_2(x_0)\right] h^2 + o[h^2]$, we get for $x_0 > 0$ that $\sqrt{x_0+h} = \sqrt{x_0} + \frac{1}{2\sqrt{x_0}}h - \frac{1}{8x_0\sqrt{x_0}}h^2 + o[h^2]$.

TRANSCENDENTAL FUNCTIONS

Similarly, we look at the exponential, logarithmic and circular functions as solution of initial value problems as, for instance, in Lang [7][pp. 65-76]. However, for much of what is needed with beginners, the following is sufficient:

i. Let exp be the solution of $f' = f$ with $f(0) = 1$. From $\exp(x_0+h) = A_0(x_0) + A_1(x_0)h + A_2(x_0)h^2 + A_3(x_0)h^3 + o[h^3]$ and therefore $\exp'(x_0+h) = A_1(x_0) + 2A_2(x_0)h + 3A_3(x_0)h^2 + o[h^2]$, we get $\exp(x_0+h) = \exp(x_0)\left[1 + h + \frac{h^2}{2} + \frac{h^3}{3!}\right] + o[h^3]$ and, since $\exp(0) = 1$, $\exp(h) = 1 + h + \frac{h^2}{2} + \frac{h^3}{3!} + o[h^3]$. Hence the approximate addition formula $\exp(x_0+h) = \exp(x_0)\cdot\exp(h) + o[h^n]$. Near ∞, $x^n = o[e^x]$, $e^{-x} = o[x^{-n}]$

ii. Near 0^+, $h^{-n} = o[\log h]$ all n and near $+\infty$, $\log x = o[x^n]$ all n.

iii. Let sin and cos be the solutions of $f'' = -f$ with $\sin(0) = 0$, $\sin'(0) = 1$ and $\cos(0) = 1$, $\cos'(0) = 0$. From $\sin(x_0+h) = A_0(x_0) + A_1(x_0)h + A_2(x_0)h^2 + A_3(x_0)h^3 + o[h^3]$ and therefore $\sin''(x_0+h) = 2A_2(x_0) + 3!A_3(x_0)h + o[h]$ we get $\sin(x_0+h) = A_0(x_0) + A_1(x_0)h - \frac{1}{2}A_0(x_0)h^2 - \frac{1}{3!}A_1(x_0)h^3 + o[h^3]$. Since $\sin(0) = 0$ and $\sin'(0) = 1$, $\sin(h) = h - \frac{1}{3!}h^3 + o[h^3]$. Similarly, $\cos(h) = 1 - \frac{1}{2}h^2 + o[h^2]$. Hence the approximate addition formulas $\sin(x_0+h) = \sin x_0 \cos h + \cos x_0 \sin h + o[h^n]$ and $\cos(x_0+h) = \cos x_0 \cos h - \sin x_0 \sin h + o[h^n]$.

We can then investigate a rather wide range of functions.

EXAMPLE 8. Let $f(x) = \frac{\sqrt{x}+e^x}{x^2-1}$. i. Near $+\infty$, we have $f(x) = \frac{e^x+[\ldots]}{x^2+[\ldots]} = e^x + o[???]$. ii. Near 0^+, we have $f(h) = \frac{h^{\frac{1}{2}}+(1+h+\frac{h^2}{2}+[\ldots])}{-1+h^2} = -1 - h^{\frac{1}{2}} + o[h^{\frac{1}{2}}]$. iii. Near $+1$, the only possible pole, we have $f(+1+h) = \frac{\sqrt{1+h}+e^{1+h}}{(1+h)^2-1} = \frac{1+\frac{h}{2}+[\ldots]+e(1+h+[\ldots])}{1+2h+[\ldots]-1} = \frac{e+1}{2}h^{-1} + o[???]$. iv. A smooth interpolation of the local graphs near 0, $+1$ and $+\infty$ gives the essential global graph in Figure 5 which in turn shows that f has *at least* one minimum and one inflection.

Derivative Rules

To find the derivative of $[f \star g]$ where $\star$ is one of the algebraic operations, we need only get the coefficient of h in $[f \star g](x_0 + h)$. For instance, to find the derivative of $f(x) = \frac{N(x)}{D(x)}$, we compute $f(x_0 + h) = \frac{N(x_0+h)}{D(x_0+h)} = \frac{N_0(x_0)+N_1(x_0)h+o[h]}{D_0(x_0)+D_1(x_0)h+o[h]} = \frac{N_0(x_0)}{D_0(x_0)} + \frac{N_1(x_0)D_0(x_0)-N_0(x_0)D_1(x_0)}{D(x_0)^2}h + o[h]$ by long division in ascending powers of h.

However, rather than to invoke the rules, it is often easier and faster to compute directly.

Example 9. To find the equation of the tangent to $f(x) = \frac{x^2-1}{x-2}$ at $x_0 = 3$, we compute BAA$f(3 + h) = \frac{(3+h)^2-1}{3+h-2} = \frac{8+6h+h^2}{1+h} = 8 - 2h$ by division in ascending powers, from which $y = 8 - 2(x - 3)$.

As might be expected, the chain rule is a bit more subtle:

– Since f is differentiable, $f(x_0 + h) = f(x_0) + f'(x_0)h + o[h] = f(x_0) + k$ where $k = f'(x_0)h + o[h]$ so that $o[k] = o[f'(x_0)h + o[h]] = f'(x_0)o[h] + o[o[h]] = o[h] + o[o[h]] = o[h]$.

– Since g is differentiable,

$$\begin{aligned} g(f(x_0) + k) &= g(f(x_0)) + g'(f(x_0))k + o[k] \\ &= g(f(x_0)) + g'(f(x_0))\{f'(x_0)h + o[h]\} + o[h] \\ &= g(f(x_0)) + g'(f(x_0))f'(x_0)h + g'(f(x_0))o[h] + o[h] \\ &= g(f(x_0)) + g'(f(x_0))f'(x_0)h + o[h] \end{aligned}$$

Limits

In the conventional approach, we define

$$\lim_{x \to x_0} f(x) = L \qquad \text{iff} \qquad \forall \epsilon \, \exists \delta \ulcorner 0 < |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon \urcorner$$

but once we decide, as in the "intuitive" approach, to avoid ϵ's and δ's, we are left with

$$\lim_{x \to x_0} f(x) = L \qquad \text{iff}$$

i.e. without even the appearance of a definition. By contrast, here we have

$$\lim_{x \to x_0} f(x) = \lim_{h \to 0} f(x_0 + h) = \lim_{h \to 0} [L + R_0(h) = L + \lim_{h \to 0} R_0(h)$$

so that

$$\lim_{x \to x_0} f(x) = L \qquad \text{iff} \qquad f(x_0 + h) = L + o[h^0]$$

In fact, we can just as easily find *sided* limits by looking at the **First Non-Constant Approximation** $f(x_0 + h) = L + A_n h^n + o[h^n]$.

Moreover, here we obviously need neither the Derivative Tests nor L'Hôpital's Rule.

Example 10. $\lim_{x\to 0} \frac{\sin x}{x}$, $\lim_{x\to 0} \frac{e^x-1}{x^3}$, $\lim_{x\to 0} \frac{1-\cos x}{x^2}$, $\lim_{x\to 0} \frac{e^x}{x^2}$, $\lim_{x\to 0} \frac{\tan x}{x^2}$, are all obvious as soon as we replace the functions by polynomial approximations.

Remarks On Some Theorems

Mean Value Theorem. In this context, the Mean Value Theorem is seen as a remainder theorem, that is as providing bounds on the error made when we approximate $f(x_0+h)$ by $f(x_0)$ by saying that, when f is differentiable, the remainder $R^{(0)}(h)$ in $f(x_0+h) = f(x_0)+R^{(0)}(h)$ is of the form $h\cdot f'(c)$ with c between x_0 and x_0+h. This is of course a special case of Taylor's formula with remainder, also called Extended Mean Value Theorem and due, significantly, to Lagrange:

$$f(x_0+h) = f(x_0)+f'(x_0)h+\cdots+\frac{f^{(n)}(x_0)}{n!}\cdot h^n + R^{(n)}(h)$$

in which $R^{(n)}(h) = \frac{f^{(n+1)}(c)}{(n+1)!}h^{n+1}$ with c between x_0 and x_0+h. It gives as an easy consequence that if $f'(x) = 0$ on (a,b) and if $f(x)$ is continuous on $[a,b]$ then $f(x)$ is constant, and that if $f'(x) > 0$ then $f(x)$ is increasing.

We should stress that, in this context, $\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{h^k}{k!}$ is *not* to be thought of as the n^{th} partial sum of a Taylor series. When writing

$$f(x) = \sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{h^k}{k!} + h^n R_n(x_0,h)$$

the remainder, $h^n R_n(x_0,h)$, for x_0 fixed, is a function of two variables, h and n, so that, in order to try to make it small, we can do either one of two things:

- For fixed n, we can make $|h|$ small (this is our viewpoint). For example, by integration by parts, we have:

  $$\int_0^\infty \frac{e^{-t}}{1+xt}dt = \sum_0^n (-1)^k k! x^k + (-x)^{n+1}\int_0^\infty \frac{e^{-t}\cdot t^{n+1}}{1+xt}dt$$

  If $x \geqq 0$, the last term is, in absolute value, less than or equal to $(n+1)!|x|^{n+1}$ and even though the absolute value of the remainder approaches ∞ as $n\to\infty$, for fixed n, it can be made as small as we wish by choosing x close enough to 0.
- For fixed x, we can try to make R_n small by letting n approach ∞ which leads to analytic functions theory. The theory is not local anymore as we are approximating f in a fixed neighborhood of x_0.

Inverse Function Theorem. *If $f'(x_0)\neq 0$ and if f' is continuous at x_0, then f has an inverse, defined in a neighborhood of $f(x_0)$ and which is continuously differentiable:* $(f^{-1}(f(x))'\big|_{x=x_0} = \frac{1}{f'(x_0)}$

In other words, letting $\xi = f^1(x)$, there exists a change of variable ξ, which is continuously differentiable so that $f(\xi(x)) = x$ and, locally, the graph of f can be rectified—but the rectification can be quite cumbersome, e.g. for $f(x) = x + x^3\sin\frac{1}{x}, x\neq 0, f(0) = 0$.

To show that f' is differentiable, we check that $f^{-1}(f(x_0)+k)$ is approximately affine. We have $f(x_0)+k = f(x_0+h) = f(x_0)+f'(x_0)h+o[h]$ with $h = \frac{k}{f'(x_0)}+\frac{o[h]}{f'(x_0)}$. Then,

$f^{-1}(f(x_0)+k) = x_0+h = f^{-1}(f(x_0)) + \frac{1}{f'(x_0)}k - \frac{1}{f'(x_0)} \cdot o[h]$ in which the remainder $-\frac{1}{f'(x_0)} \cdot o[h]$ is plausibly small.

For a *proof*, the remainder must be shown to be $o_k[k]$. From $k = h(f'(x_0))$, $h \to 0$ implies $k \to 0$ and, since $f'(x_0) \neq 0$, $k \to 0$ implies $h \to 0$ so that $o_h[1]$ iff $o_k[1]$. Then, $\frac{h}{f'(x_0)} \cdot o_k[1] = \frac{k \cdot o_k[1]}{f'(x_0)(f'(x_0)+o_k[1])} = k \cdot o_k[1]$

Fundamental Theorem

For most calculus students, stressing the "antiderivative aspect" is more importan than stressing the "measure theoretic aspect". So, following Picard [9], the Fundamental Theorem is proven by solving, using finite differences, the initial value problem: Given a function $f(x)$, find the value at x_1 of a function $F(x)$ such that $F(x_0) = y_0$ and $F'(x) = f(x)$.

If we assume the existence of an antiderivative $F(x)$, we have immediately:

$$\begin{aligned}
F(x_0+h) - F(x_0) &= F'(x_0)h + o_1[h] \\
&= f(x_0) \cdot h + o_1[h] \\
F(x_0+2h) - F(x_0+h) &= f(x_0+h) \cdot h + o_2[h] \\
F(x_0+3h) - F(x_0+2h) &= f(x_0+2h) \cdot h + o_3[h] \\
\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\
F(x_0+nh) - F(x_0+(n-1)h) &= f(x_0+(n-1)) \cdot h + o_n[h]
\end{aligned}$$

Adding and cancelling on the left, we get:

$$F(x)|_{x_0}^{x_1} = h \sum_{i=0}^{i=n-1} f(x_0+ih) + h \sum_{k=1}^{k=n} o_i[1]$$

where the sum $\sum_{k=1}^{k=n} o_k[1]$ is extremely complicated which is a good reason to let $n \to \infty$ as then, for $f(x)$ smooth enough, $\sum_{k=1}^{k=n} o_k[1]$ will approach 0 and we have

$$F(x)|_{x_0}^{x_1} = h \lim_{n\to\infty} \sum_{k=0}^{k=n-1} f(x_0+kh) = h \lim_{n\to\infty} \sum_{i=0}^{i=n-1} f(x_k),$$

where $h\sum_{k=0}^{k=n-1} f(x_i)$ is called a **Riemann sum**, and which we can *then* easily interpret geometrically as the approximation of $\int_{x_0}^{x_1} f(x)dx$, the area under the graph of f. We thus have the Fundamental Theorem:

$$F(x)|_{x_0}^{x_1} = \int_{x_0}^{x_1} f(x)dx$$

Conclusion

Of course Lagrange did not really succeed since the $o[h^n]$ rest on the notion of limit.[6] But, with polynomial algebra as sole prerequisite[7] and whether the rules for *calculating* with $o[h^n]$ be proven or merely stated, the use of asymptotic expansions empowers beginners and, not least, makes it possible to develop in a particularly transparent way a "coherent view" of the calculus.[8]

Appendix: Landau's Little o's

i. Given a function g, $o[g]$ near 0 is the set of all functions f that are **negligible** compared to, i.e. strictly dominated by, g near 0. By an abuse of notation, instead of $f \in o[g]$, we write $f = o[g]$ to mean that $f(h) < g(h)$—or that $\lim_{h\to 0} \frac{f(h)}{g(h)} = 0$. Similarly near ∞.

ii. We have $o[g] \pm o[g] = o[g]$ and $A \cdot o[g] = o[g]$ so that, for instance, $A_3h^3 + A_4h^4 + A_5h^5 = o[h^2]$. We also have $o[g] + o[o[g]] = o[g]$ and $|o[g]|^\lambda = o[g^\lambda]$.

References

[1] S. Bassein, *An infinite series approach to calculus*, Publish or Perish, 1993.
[2] J. Dieudonné, *Infinitesimal calculus*, Houghton-Mifflin, 1971.
[3] W. Freed, *Infinitesimal calculus*, Concordia College, Edmonton, Alberta, 1986.
[4] T. Gowers, *Mathematics, a very short introduction*, Oxford, 2002.
[5] J. Keisler, *Elementary calculus, an infinitesimal approach*, Prindle, Weber and Schmidt, 1976, 1986.
[6] J. L. Lagrange, *Théorie des fonctions analytiques, contenant les principes du calcul differentiel*, Gauthier-Villars, 1813.
[7] S. Lang, *Analysis I*, Addison-Wesley, 1976.
[8] H. Levi, *Polynomials, power series and calculus*, D. Van Nostrand, 1968.
[9] E. Picard, *Traité d'analyse*, Gauthier-Villars, 1901.
[10] A. Robert, *Nonstandard analysis*, Wiley, 1988.
[11] A. Schremmer, *How content matters*, Notices of the AMS **60** (2013).

[6] However it is quite annoying still having to read here and there that Lagrange's approach was discarded because "it could only deal with functions *defined by formal series*"—whatever *that* is supposed to mean.

[7] For a treatment appropriate to this approach, see Schremmer [11][pp. 340-342].

[8] Not to mention that it obviates the need for "precalculus" in that it justifies and incorporates the study of the gauges.