A POST-CAUCHY VIEW OF LAGRANGE'S CALCULUS

Adapted from the Lecture Notes of a Minicourse given at the Joint Mathematics Meetings in Louisville, Janauary 1990.

Francesca Schremmer, *West Chester University*
Alain Schremmer, *Community College of Philadelphia*

–1. INTRODUCTION. Many reasons have been given as to why so many (50%) of our students fail calculus: lack of relevance of the subject, poor preparation, lack of motivation have been cited. It is our belief that this is mainly because calculus is just plain difficult: unlike other subjects, algebra included, in which one is introduced progressively, calculus is brutally difficult, right from the start. Limits, continuity, differentiability are introduced in the first weeks of the course and, if those concepts are not mastered immediately, it is impossible for the students to function intelligently. Many of those who do pass the course in fact do so only because, through patience and discipline, they have been able to master some basic techniques. But, really, they failed the subject. Beside the inherent difficulty of calculus, another reason for failure is that unless one knows already calculus and, in fact, a good deal more than that, it is difficult, the way most calculus texts are written now, to see in calculus more than a disconnected set of topics collected there because of a set of common techniques used to study them.

What we propose here is a unified, progressive approach throughout calculus so that we can provide a mathematical environment sufficiently structured that the students can feel secure and, at the same time, take an active part in their learning instead of solving passively excruciatingly boring routine exercises.

The theme throughout is that of local approximation and, from there, attempts at globalization. Roughly speaking: It is generally agreed that first year calculus is about the calculus of functions from R to R. In other words, one wishes to calculate with functions and to understand their behaviour. Because of the great variety of possible behaviors, one borrows from number theory the idea of approximating irrationals by, for example, their truncated decimal expansion. Because the truncated decimal expansion has finitely many terms, it is a finite combination of terms of the form 10^k or $\left(\frac{1}{10}\right)^k$ where k is any integer ≥ 0. Suppose that we truncate the expansion of a positive irrational at the n^{th} decimal, then the error made in the approximation is smaller that $\left(\frac{1}{10}\right)^n$. A better approximation is obtained by truncating one decimal further and the new approximation is equal to the previous one to which is added a multiple of $\left(\frac{1}{10}\right)^{n+1}$, the approximations are "nested". By analogy, given a function f, we will try to approximate it near a point x_0, as well as possible and in a sense which we will make clear later on. We will do so by using functions as simple as possible. At first, we try constant functions, in other words polynomials of degree 0. If, by our standards, a good approximation is possible, then f is continuous at x_0. If an approximation by a polynomial of degree 1 in $(x–x_0)$ is possible, then f is differentiable and so on but the approximation will always be using a finite number of terms, the various approximations are nested in the sense that the n^{th} approximation is the sum of the $(n–1)^{\text{th}}$ approximation and a multiple of $(x–x_0)^n$. The newly added term is smaller by an order of magnitude than the previous one when $x \to x_0$. The gauge functions for our approximations are from the set $\{(x–x_0)^n\}$, n an integer ≥ 0. (Such polynomials are nothing but the Taylor polynomials or jets of various order for f at x_0.) For every approximation, one asks: how well does this approximation reflect: the numerical values, the geometrical and topological properties of f in a neighborhood of x_0? What can be said of those functions which enjoy such approximation at each point of an interval?

But, very quickly, one discovers that some very desirable non-pathological functions, for instance rational functions near their poles, cannot always be well approximated by polynomials. so one is led naturally to enlarging the set of gauge functions to $\{(x-x_0)^n\}$ $n \in \mathbb{Z}$.

One is also very interested by what happens as $x \to \pm\infty$. By letting $t = \frac{1}{x}$, one reduces the problem to the study of the behaviour of f at the origin but, here again, the need to extend the set of gauge functions is very clear: neither e^x nor $\log x$ can be decently approximated by polynomials at ∞. So, depending on which function one wishes to study, one may add to our set of gauge functions, for instance, $x^{\alpha x}$, $(\log x)^{\beta}$, x^{γ}, where $\alpha, \beta, \gamma \in \mathbb{Q}$, together with their products.

This approach of systematically expanding functions goes as far back as Lagrange (1797). His goal was to algebraize calculus and free it from "*any consideration of infinitesimals, vanishing quantities, limits and fluxions and reduce it to the algebraic study of finite quantities*". Even though Lagrange's great contributions to mathematics were in mechanics, Lagrange wanted the calculus to be developed for solving problems in geometry first and only then in mechanics.

One may wonder whether Lagrange, had he not had his vision of derivatives as coefficients in power expansions, could have developed the Calculus of Variations as far as he did. Unfortunately, instead of using finite expansions, Lagrange used infinite formal power series which were not understood until at least a century later (and here we must disagree with Robinson's reading of Lagrange: Lagrange was very aware of the difference between a C^∞ function and an analytic one). What made the matter far worse was that Lagrange thought every continuous function was piecewise C^∞. Even though a modification of his method, the method of asymptotic expansions, which is the one which we present here is easy enough, has all the advantages of Lagrange's approach, is used extensively in ODE, in Mathematical Fluid Mechanics, Differential Topology and Number Theory, it never regained its due place in the exposition of elementary calculus.

The first part of this course will be devoted to a presentation of the mathematical framework free from pedagogical concerns, while the second will present a model for a possible implementation.

0. TOOLBOX. Our approximation will mostly use polynomials. Thus our tools really are mainly in the algebra of polynomials and, in particular, mostly multiplication and division.

We will need:

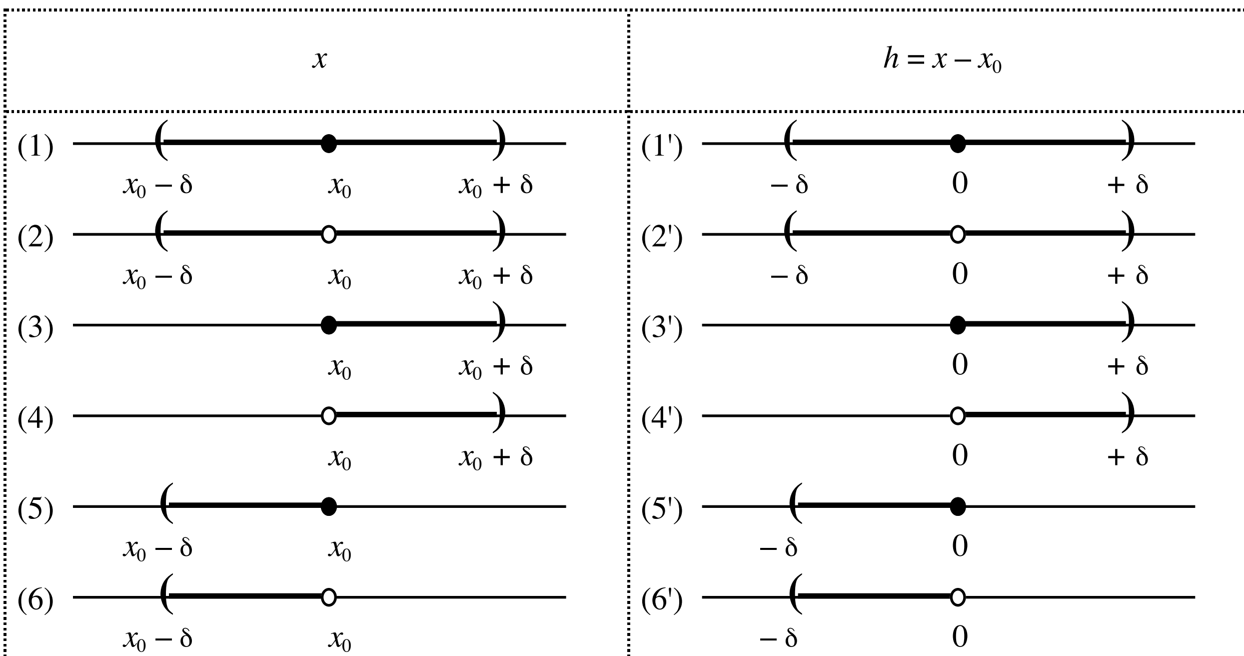
— The binomial theorem

— Two ways of dividing polynomials. We will use one when $|x| \to \infty$ and the other when $x \approx 0$. this parallels the way one divides numbers: any number (assumed >0) is the sum of an integer and of a number between 0 and 1. An integer, when written in base 10 is a combination of powers of 10 ordered by decreased exponents and also decreasing order of magnitude, the first one being the dominant one. For example, $1{,}349 = 1\cdot 10^3 + 3\cdot 10^2 + 4\cdot 10^1 + 9\cdot 10^0$ and $1{,}349 \approx 10^3$. On the other hand, a number between 0 and 1 is written in base 10 as a combination of powers of $\frac{1}{10}$ ordered by increasing exponents but also by decreasing order of magnitude and, again, the first non-zero term is the dominant one; for example: $0.085 = 0\bullet\left(\frac{1}{10}\right)^0 + 0\bullet\left(\frac{1}{10}\right)^1 + 8\bullet\left(\frac{1}{10}\right)^2 + 5\bullet\left(\frac{1}{10}\right)^3$ and $0.085 \approx 8\bullet\left(\frac{1}{10}\right)^2$.

For $|x| \to \infty$, $1 \ll x \ll x^2 \ldots$, hence, to divide two polynomials for $|x| \to \infty$, one orders their terms by decreasing exponents and one performs the division to get a polynomial in x but, if the remainder is not 0, it is often very useful to have also at least one non-zero "decimal". For example,
$\frac{x^2}{x^2+1} = 1 - \frac{1}{x^2} + \frac{1}{x^2(x^2+1)}$ $|x| \to \infty$, where 1 is the principal part, $-\frac{1}{x^2}$ is the "first decimal", and $\frac{1}{x^2(x^2+1)}$ is the remainder.

On the other hand, when $x \approx 0$, $1 \gg x \gg x^2 \ldots$, thus when ordering the terms in decreasing order of magnitude, the exponents are increasing:
$\frac{x^2}{x^2+1} = \frac{x^2}{1+x^2} = x^2 - \frac{x^4}{1+x^2}$ $x \approx 0$, where x^2 is the principal part.

To facilitate things, we list here the various ways $x \to x_0 \Leftrightarrow x$ is in a neighborhood of $x_0 \Leftrightarrow x$ is in one of the six following types of sets which will be made clear whenever needed.



1. LIMITS AND CONTINUITY. The first topics studied in calculus are limits and continuity. Let us review their definition and recast them in the context of approximation:

(*) $\lim_{x \to x_0} f(x) = L$ iff $\forall \varepsilon>0, \exists \delta>0, \forall x\ 0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$

In other words, $\forall \varepsilon > 0$, $\exists$ a neighborhood of type (2) in which $|f(x) - L| < \varepsilon$ (If one wishes to define one-sided limits one considers neighborhoods of the form (4) or (6) and for continuity, neighborhoods of the form (1), (3) or (5)). This is equivalent to saying:

$$f(x) = \mathrm{L} + \mathrm{R}_0(x_0, x - x_0)$$

where $\left|\mathrm{R}_0(x_0, x - x_0)\right| < \varepsilon$ for any given $\varepsilon > 0$ in a small enough neighborhood of x_0, of type (2) or, letting $x = x_0 + h$, $f(x_0+h) = \mathrm{L}\bullet 1$ + term negligible compared to 1.

Notation. **i.** Let g be a function defined in a neighborhood of x_0, we write $g(x) = \mathrm{o}(1)$ as $x \to x_0$ (x being in one of the six neighborhoods described previously and which neighborhood will be chosen accordingly in each context) if and only if $\lim_{x\,\text{å}\,x_0} g(x) = 0$..

ii. f«g as $x \to \mathrm{x}_0$ if and only iff $\dfrac{f(x)}{g(x)} = \mathrm{o}(1)$, f is said to be negligible compared to g, $x \to 0$.

$\lim_{x\,\text{å}\,x_0} f(x) = L$ iff $f(x_0+h) = \mathrm{L} + \mathrm{R}_0(x_0, h)$ with $\left|\mathrm{R}_0(x_0, h)\right| \ll 1$ as $h \to 0$

or, equivalently,

(**) $\boxed{\lim_{x\,\text{å}\,x_0} f(x) = L \quad \text{iff} \quad f(x_0+h) = \mathrm{L} + \mathrm{o}(1) \text{ as } h \to 0 \text{ in a neighborhood of type (2).}}$

Thus one reads the limit L as the constant term in the one-term expansion of $f(x_0+h)$, $h \to 0$.

This is the definition of limit which we adopt here. It is easy to prove by contradiction that if L exists, it is unique and that it is the **best constant approximation** of f at x_0, by which we mean that in a neighborhood of x_0 any other approximation of f by a constant will yield a larger error than $\mathrm{R}_0(x_0, h)$.

Definition. *f is continuous at x_0 iff $f(x_0+h) = f(x_0) + \mathrm{o}(1)$ as $h \to 0$ in* (1) *or*

iff f admits a Best Polynomial Approximation of degree 0 (BPA$^{(0)}$) *in a neighborhood of type* (1).

Obviously, εs and δs or infinitesimals are hidden in o(1) and whatever rigor one demands in the course is therefore reduced to showing that some functions approach 0 as $h \to 0$. It has been our experience that students react much more positively to (**) than to whatever heuristic approach of (*) one takes. It should also be said, in favor of (**) that many computations are far easier and that they are reduced to manipulations of functions that are o(1) as $h \to 0$. The introduction of o(1) helps the students to develop a sense of what are the various order of magnitude in an expression and that is far more desirable than virtuosity at chopping epsilons. For example,

1. To show that $f(x) = x^2 + 2x - 1$ is continuous at x_0, compute $f(x_0+h)$, organize the terms by decreasing order of magnitude, $h \to 0$, and check that $f(x_0+h) = f(x_0) + \mathrm{o}(1)$ as $h \to 0$:

$$f(x_0+h) = (x_0+h)^2 + 2(x_0+h) - 1$$

$$= x_0^2+2x_0h+h^2 + 2x_0+2h - 1 = [x_0^2 + 2x_0 - 1] + h[2x_0+2 + h].$$

The first term is $f(x_0)$ and, as to the second term, it is easy to see or to prove that it is o(1) as $h \to 0$.

2. To show that $f(x) = \dfrac{1}{x-2}$ is continuous for all $x_0 \neq 2$, compute $f(x_0+h) = \dfrac{1}{x_0-2+h}$ and divide in ascending powers of h:

$$\begin{array}{r|l} & \dfrac{1}{x_0-2} \\ \hline x_0-2+h & 1 \\ & 1 + \dfrac{h}{x_0-2} \\ \hline & -\dfrac{h}{x_0-2} \end{array}$$

so that $f(x_0+h) = \dfrac{1}{x_0-2} - \dfrac{h}{(x_0-2)(x_0-2+h)}$ for $x_0 \neq 2$

The first term is indeed $f(x_0)$ and the second term is o(1), $h \to 0$.

Some properties of o(1):

$x - x_0 = o(1)$, as $x \to x_0$

$o(1) \pm o(1) = o(1)$, as $x \to x_0$

$(o(1))^{\alpha} = o(1)$, $\forall \alpha$ positive rational, $x \to x_0$

Let $g(x)$ be a bounded function in a neighborhood of x_0; then $g(x)o(1) = o(1)$, $x \to x_0$ and, in particular, if $c \in R$ then $c \bullet o(1) = o(1)$, $x \to x_0$

$(1+o(1))^{\alpha} = 1 + o(1)$, $\forall \alpha \in Z$, $x \to x_0$

These properties would not be proved in most calculus courses but once the students realize that the o(1) functions behave, as far as their algebra is concerned, like the number 0, they will have understood the point and they will be able to effectively compute with limits.

From the properties of o(1) stated above, it is then straightforward to prove the continuity of the sum, product, quotient and composition of functions as well as the continuity of polynomial and rational functions away from their poles.

Another advantage of the definition (**) is that the theorem:

Theorem. *If f is continuous at x_0 and if $f(x_0) > 0$, then $f(x) > 0$ in an open interval containing x_0*

is obvious when the continuity of f is expressed by $f(x) = f(x_0) + o(1)$ as $x \to x_0$.

Continuity at x_0 is a local property and one would like to study the property of continuous functions on an *interval* (like, for instance, the Intermediate Value Theorem). Clearly, the definition of continuity at x_0 by the local existence of a best constant approximation is of no help, but it can point very clearly where some of the difficulties are in proving a theorem like

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

but offers no help in the proof. Because f is continuous on an interval, say $[a,b]$, $\forall x_0 \in [a,b]$, $f(x_0+h) = f(x_0) + o(1)$. Suppose h is in a neighborhood of 0 whose size depends on x_0, such that $o(1) < \frac{1}{10}$ for example. If one knew that one could cover $[a,b]$ using finitely many of these intervals, say N, then $\left|f(x)-f(a)\right|$ would be bounded by $\frac{N}{10}$ and the theorem would be proved, thus, raising the possibility and desirability of a characterization of a closed bounded interval by the property that one can extract from any open covering a finite one, that is, we need compactness.

2. DIFFERENTIABILITY. We defined earlier the relation «: $f \ll g$, $x \to x_0$ (f is negligible compared to g in a neighborhood of x_0) iff $\frac{f(x)}{g(x)} = o(1)$, $x \to x_0$. Thus, if one wished to improve upon our previous approximation of f at x_0 by a constant function, it makes sense to look for $f(x)$ in the form

$$f(x_0+h) = \beta_0 \bullet 1 + g(x_0,h) + R_1(x_0,h),\ h \to 0$$

where $x = x_0+h$, $\beta_0 \in R$ and the terms are ordered in decreasing order of magnitude: $1 \gg g(x_0,h) \gg R_1(x_0,h)$, $h \to 0$. There are infinitely many choices for $g(x_0,h)$ and for the sake of simplicity one chooses $g(x_0,h)$ of the form $\beta_1 h$, $\beta_1 \in R$ noting that $h \ll 1$, $h \to 0$, and hence $R_1(x_0,h) = h o(1)$ and we define:

Temporary definition. *f is differentiable at x_0 iff there exists $\beta_0, \beta_1 \in R$ such that*

$$f(x) = \beta_0 + \beta_1(x-x_0) + (x-x_0)o(1),\ x \to x_0 \text{ in } (1)$$

or, equivalently,

$$f(x_0+h) = \beta_0 + \beta_1 h + h o(1),\ h \to 0 \text{ in } (1')$$

Note that, if $h \to 0$ in (3') or (5'), then one gets one-sided derivatives.

Because $\beta_1 h + h o(1) = o(1)$, β_0 is the BPA$^{(0)}$ and hence $\beta_0 = f(x_0)$, we have:

Theorem. *If f is differentiable at x_0, then it is continuous at x_0*

Again, by contradiction, one proves that β_1 is unique and that $\beta_0 + \beta_1 h$ is the BPA$^{(1)}$ or Best Affine Approximation (BAA) of $f(x_0+h)$, $h \to 0$. In other words, any other affine approximation will yield a worse remainder. Also,

$$\beta_1 = \frac{f(x_0+h) - f(x_0) - h o(1)}{h}$$

thus

$$\beta_1 = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$$

and one is entitled to call β_1 the **derivative** of f at x_0 and which we denote $f'(x_0)$ and the definitive version of the definition is then

Definition. *f is differentiable at x_0 iff $\exists$ a number $f'(x_0)$ such that*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + (x-x_0)\mathrm{o}(1),\ x\to x_0 \textit{ in } (1)$$

or, equivalently,

$$f(x_0+h) = f(x_0) + f'(x_0)\,h + h\mathrm{o}(1),\ h\to 0 \text{ in } (1')$$

The straight line whose equation is $y = f(x_0) + f'(x_0)(x - x_0)$ is, by definition called the **tangent line** to the graph of f at x_0 also called the **osculating line** to the graph of f at x_0. Practically, to compute the derivative of f at x_0, one writes $f(x_0+h)$ as the sum of polynomial of degree 1 in h and a remainder of the form $h\mathrm{o}(1)$; $f'(x_0)$ is then the coefficient of h.

Example. Find the derivative of $f(x) = \dfrac{x^2-1}{x-2}$ at $x_0 = 3$.

$f(3+h) = \dfrac{(3+h)^2-1}{3+h-2} = \dfrac{8+6h+h^2}{1+h}$ and by division in ascending powers,

$$\begin{array}{r|l} & 8 \quad -2h \\ \hline 1+h & 8 \quad +6h \quad +h^2 \\ & 8 \quad +8h \\ \hline & \quad -2h \quad +h^2 \\ & \quad -2h \quad -2h^2 \\ \hline & \qquad +3h^2 \end{array}$$

so that $f(3+h) = 8 - 2h + \dfrac{3h^2}{1+h}$ where $8 = f(3)$ and, since $\dfrac{3h^2}{1+h} = h\mathrm{o}(1)$, $-2 = f'(3)$. The equation of the tangent is $y = 8 - 2(x-x_0)$.

Remark. Few instructors would expect, at this point, their students to be able to solve such a problem without the help of the quotient rule but, in fact this method of solution is faster than when using the quotient rule and more desirable because the students can find easily BAAs without the help of any rule in fact, most of the usual rules can be proved naturally *by the students themselves* and with no trick.

Multiplication Rule:

$$\begin{aligned}[f\bullet g](x_0+h) &= f(x_0+h)\bullet g(x_0+h)\\ &= [f(x_0)+f'(x_0)h+h\mathrm{o}(1)]\bullet[g(x_0)+g'(x_0)h+h\mathrm{o}(1)] \text{ and, by multiplication,}\\ &= f(x_0)g(x_0) + [f'(x_0)g(x_0)+f(x_0)g'(x_0)]h + h(\mathrm{o}(1), \quad h\to 0\end{aligned}$$

Quotient Rule:

$$\frac{f(x_0+h)}{g(x_0+h)} = \frac{f(x_0)+f'(x_0)h+h\mathrm{o}(1)}{g(x_0)+g'(x_0)h+h\mathrm{o}(1)}$$ and, by division in *ascending* powers,

$$\begin{array}{r|llll}
 & \dfrac{f(x_0)}{g(x_0)} & + \dfrac{1}{g(x_0)}\left[f'(x_0)-\dfrac{f(x_0)}{g(x_0)}g'(x_0)\right]h & & \\
\hline
g(x_0)+g'(x_0)h+h\mathrm{o}(1) & f(x_0) & + \quad f'(x_0)h & + h\mathrm{o}(1) & \\
 & f(x_0) & + \quad \dfrac{f(x_0)}{g(x_0)}g'(x_0)h & & + h\mathrm{o}(1) \\
\cline{2-5}
 & & \left[f'(x_0)-\dfrac{f(x_0)}{g(x_0)}g'(x_0)\right]h & + h\mathrm{o}(1) &
\end{array}$$

$$= \frac{f(x_0)}{g(x_0)} + \frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}\,h + h\mathrm{o}(1)$$

Chain Rule:

By the differentiability of f at x_0, $f(x_0+h) = f(x_0)+f'(x_0)h + h\mathrm{o}(1)$, $h\to 0$, and, by the differentiability of g at $f(x_0)$, $g(f(x_0)+k) = g(f(x_0)) + g'(f(x_0))\,k + k\mathrm{o}(1)$, $k\to 0$. Then,

$$g[f(x_0+h)] = g[f(x_0)+k] \text{ where } k = h[f'(x_0)+h\mathrm{o}(1)]$$

Clearly, if $h\to 0$, then so does k and hence if a function is o(1), $h\to 0$, then it is o(1) when $k\to 0$.

$$g[f(x_0+h)] = g(f(x_0)) + g'(f(x_0))\,h[f'(x_0)+\mathrm{o}(1)] + h(f'(x_0)+\mathrm{o}(1))\mathrm{o}(1)$$

$$= g(f(x_0)) + g'(f(x_0))\circ f'(x_0)h + h\mathrm{o}(1),\quad h\to 0.$$

Local Results. Suppose $f(x)$ is differentiable at x_0, what can be obtained from it? More than before: one can approximate $f(x_0+h)$ by $f(x_0)+f'(x_0)h$ with an error $h\mathrm{o}(1)$ for h small enough, but how small is small enough? So our present knowledge, as far as numerical approximation is concerned is better than before but not sufficient. We still have no bound on the error made in the approximation. As far as the geometry of the graph is concerned, we are in the same situation: for example, suppose that $f'(x_0)>0$. One can deduce that $f(x)$ is larger than $f(x_0)$ in some neighborhood of x_0 but one cannot deduce from this that $f(x)$ is increasing in the neighborhood of x_0 unless one assumes, for example, that $f'(x)$ is continuous. But it is easy to see that in order for $f(x)$ to have a local extreme at x_0, $f'(x)$ has to be 0.

But what about the topology of the graph? The answer is given by the

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

Or, 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 (the rectification can be quite cumbersome. For example $f(x) = x + x^3 \sin\frac{1}{x}$, $x \neq 0, f(0) = 0$.)

Global results. To get these, we again use compactness in proving Rolle's theorem. The Mean Value Theorem then gives us the answers we need as it provides bounds on the error made when we approximate $f(x_0 + h)$ by $f(x_0)$. It also 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, that if $f'(x) > 0$ then $f(x)$ is increasing and with some work gives L'Hôpital's rule.

3. HIGHER ORDER DIFFERENTIABLITY. At this point, one can follow either one of two courses. One can differentiate derivatives and define recursively $f''(x) = (f'(x))'$ and so on and define: f is ***n*-times (iterated) differentiable at** x_0 iff $f^{(n)}(x_0)$ exists. Or, one can pursue the idea of approximation using as gauge functions for the approximation the functions from the set e $=\{(x-x_0)^n\}_{n \text{ integer} \geq 0}$ Note that the set is linearly ordered by « as $x \to x_0$ and that it is closed for multiplication. By analogy with our treatment of differentiablity, one would like to define: f is ***n*-times (Peano) differentiable at** x_0 iff there exists a polynomial of degree n, called the osculating polynomial, such that

$$f(x) = \sum_{k=0}^{k=n} \alpha_k (x - x_0)^k + (x - x_0)^n o(1), x \to x_0$$

or, equivalently,

$$f(x_0 + h) = \sum_{k=0}^{k=n} \alpha_k h^k + k^n o(1), h \to 0$$

and, again, if $f(x)$ admits such an approximation, the approximation is unique and hence $\alpha_0 = f(x_0)$, $\alpha_1 = f'(x_0)$. It is thus natural to ask whether there is a further connection between the the two definitions of differentiability and, if yes, find the relation between α_k and $f^{(k)}(x_0)$. If $f^{(k)}(x_0)$ exists for $k = 0$ to n, then it is reasonable to consider the Taylor polynomial $\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}$ as a candidate for the osculating polynomial because its n first derivatives at x_0 agree with those of f at x_0. In fact, one has:

Theorem. *If* $f^{(k)}(x_0)$ *exists for* $k = 0, 1, \dots n$, *then* $f(x) = \sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!} + (x - x_0)^n o(1)$

Proof. Use L'Hôpital's rule repeatedly on

$$\frac{f(x_0 + h) - \sum_{k=0}^{k=n-1} f^{(k)}(x_0)\frac{h^k}{k!}}{\frac{h^n}{n!}}$$

to get the result.

On the other hand, the existence of an osculating polynomial of degree n with $n>1$ at x_0 does not insure the existence of any derivative of order >1 at x_0.

A simple pathological case:

Let $f(x) = \begin{cases} x^3 \sin\frac{1}{x} & x\neq 0 \\ 0 & x=0 \end{cases}$

For $x\approx 0$, $f(x) = 0 + 0x + 0x^2 + x^2 \bullet x \sin\frac{1}{x}$

where $0 = f(0)$, $0x = f(0)x$ and $x^2 \bullet x\sin\frac{1}{x} = x^2 0(1)$. But is $0x^2$ equal to $\frac{f''(0)}{2!}$?

On the other hand,

$$f'(x) = -x\cos\frac{1}{x} + 3x^2 \sin\frac{1}{x}$$

Consequently, $f''(0)$ does not exist and cannot be the coefficient of x^2 in the osculating polynomial.

This needs not be a matter of concern for us as it can be shown that if α_n exists in a neighborhood of x_0 and is bounded either from above or from below, then it *is* the n^{th} (iterative) derivative of f. For our purposes, the two notions are equivalent.

Local results. Apart from a better numerical approximation for $f(x_0)$, what have we gained by considering higher degree approximations? The notion of curvature of the graph of $f(x)$ at x_0 which should be, whatever it is, the curvature of the osculating parabola.

Classification of critical points. From

$$f(x_0+h) = f(x_0) + f'(x_0)h + f''(x_0)\frac{h^2}{2!} + h^2\, o(1), \text{ as } h \to 0$$

the second derivative test to classify the non-degenerate critical points of f is obvious. In fact, in the context of approximation, the classification of all critical points is clear:

Let x_0 be a critical point of $f(x)$ which we will assume, for simplicity, to be C^∞.

Theorem. *If the first non-zero derivative of f at x_0 is of odd order, x_0 is not a (local) extremum*

If the first non-zero derivative of f at x_0 is of even order, x_0 is a (local) extremum and, if it is positive, x_0 is a (local) minimum and if it is negative, x_0 is a (local) maximum.

Moreover, using the inverse function theorem, one has:

Let f be, for simplicity, a C^∞ function in a neighborhood of x_0, then the graph of f is, up to a smooth reparametrixation of x the graph of its first non-constant, non-zero term in its Taylor expansion.

Global results. Taylor's theorem, like the Mean Value Theorem of which it is the generalization, gives global information on, for example, the concavity of the graph and an estimate of the accuracy of the numerical approximations.

Important Remark. In this context, $\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}$ is *not* to be thought of as the n^{th} partial sum of a Taylor series. When writing

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

the remainder, $(x-x_0)^n R_n(x_0,x-x_0)$, for x_0 fixed, is a function of two variables, x and n. In order to try to make it small, one can do either one of two things:

♦*For fixed n*, we can make $|x-x_0|$ small (this was our point of view)

For example.. By integration by parts, one shows:

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

If $x\geq 0$, 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 one wishes by choosing x close enough to 0.

♦*For fixed x*, we can try to make R_n small by letting $n\to\infty$ which leads to analytic functions theory. The theory is not local anymore as one is approximating f is a fixed neighborhood of x_0.

4. ASYMPTOTIC EXPANSIONS. To summarize, we have been "gauging" functions locally by way of **gauge functions** from the set g = $\{(x-x_0)^n\}$, n integer≥ 0 and x near x_0 which is linearly ordered by « and closed for multiplication. The choice of g was a natural one because of its simplicity but other than that was arbitrary and, in fact, it fails to allow us register certain functions like

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

In order to study such functions one has to choose a different set of finer functions.

More generally,

Definition. *Let* g $=\{\phi_n\}_{n\in\mathbb{N}}$ *be a set of functions linearly ordered by «, closed for multiplication. Hence,* $\phi_{n+1}=\phi_n o(1)$ $x\to x_0$. *The expansion* $\sum_{k=1}^{k=n} a_k\phi_k(x)$ *is an* **asymptotic expansion** *with n term for f near* x_0 *iff* $f(x)=\sum_{k=1}^{k=n} a_k\phi_k(x)+\phi_n(x)o(1)$, $x\to x_0$.

One then proves easily that if f has an asymptotic expansion, it is unique.

Example. If $f(x)$ is a n-times differentiable function in a neighborhood of x_0 then $\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}$ is an asymptotic expansion of f near x_0.

The set of gauges g = $\{(x-x_0)^n\}$, n integer≥0 is not closed for division and this explains why, for example, one cannot obtain an asymptotic expansion of a rational function near one of its poles. To remedy this situation, we just take g = $\{(x-x_0)^n\}$, $n \in \mathbb{Z}$ and, for example, to study the behaviour of $f(x) = \dfrac{2x-3}{x^2-1}$ at $x = 1$, one expands

$$f(1+h) = \frac{-1+2h}{2h+h^2} = -\frac{1}{2h} + \frac{5}{2(2+h)}$$

or

$$f(x) = -\frac{1}{2(x-1)} + \frac{5}{2(x+1)}$$

The study at ±∞ is done similarly. Computing a limit as $x \to \pm\infty$ is useful but often by simple means one can obtain more information. For example, as $x \to \pm\infty$, $\dfrac{3x^2+2}{x^2-1} = 3 + \dfrac{5}{x^2} + \dfrac{5}{x^2(x^2-1)}$, not only gives the existence of the horizontal asymptote $y = 3$ for the graph but shows how the graph sits with respect to it.

In some important cases, the set of gauge functions g = $\{(x-x_0)^n\}$, $n \in \mathbb{Z}$ does not provide satisfactory approximations. For example, e^x is very poorly appproximated by polynomials at $\pm\infty$ and thus the set of gauge functions has to be modified accordingly in each case.

In short, the elementary differential calculus can be seen as being mostly the local study of functions through their asymptotic expansions with respect to the set of gauges $\{(x-x_0)^n\}$, $n \in \mathbb{Z}$. The choice of this set of gauges yields particularly simple computations: in general, the computation of each term of an expansion involves a separate computation sometimes very difficult but when expanding with respect to $\{(x-x_0)^n\}$, $n \in \mathbb{Z}$, all the coefficients are easily obtained as soon as $f(x)$ has been obtained.

5. INTEGRAL CALCULUS. When analyzing the contents of a standard freshman *integral* calculus course, one notices that, after the introduction of the Riemann integral, most of the course is devoted to techniques and applications of various sorts. In fact, the Riemann integral is the one new idea of mathematical importance. Most textbooks motivate its study historically by the area problem. We would rather follow Picard in motivating the relation between the antiderivative and the definite integral. We translate here from his Traité d'Analyse:

"*Integral Calculus was born the day one asked the question: given f(x), does there exist a function whose derivative is f(x), in other words a function which satisfies*

$$\frac{dy}{dx} = f(x) \qquad (1)$$

This question was at first answered by a geometrical interpretation which, even though it had no value in itself, helped greatly with the solution of the problem: One graphs first the function f then one considers the area bounded by this curve, the x*-axis and two parallels to the y-axis, one fixed, the other one variable. One then shows that the area, considered as a function of the x-intercept x of the second parallel is a function of x having f(x) as derivative. It is clear that, unless one assumes that the notion of area is given, the problem has not been solved rigorously. We assume f continuous. The following considerations lead naturally to the algebraic expression which* *plays a fundamental role in the Integral Calculus. Assume, for a moment, the existence of a function y satisfying* (1), *with* $y(a) = y_0$ *and* $y(b) = Y$. *Subdivide the interval* $[a,b]$ *in n intervals and let* $x_1, x_2, \ldots, x_{n-1}$, *be the x-coordinate of the subdividing points. Let* $y_1, y_2, \ldots, y_{n-1}$ *be the corresponding values for y. If the interval* $x_1 - a$ *is small enough, the quotient* $\frac{y_1 - y_0}{x_1 - a}$ *is very close to* $f(a)$ *and we have the following equations which hold only approximately:*

$$\begin{gathered} y_1 - y_0 = (x_1 - a)f(a) \\ y_2 - y_1 = (x_2 - x_1)f(x_1) \\ \vdots \\ Y - y_{n-1} = (b - x_{n-1})f(x_{n-1}) \end{gathered}$$

Adding them up, we obtain:

$$Y - y_0 = (x_1 - a)f(a) + (x_2 - x_1)f(x_1) + \ldots + (b - x_{n-1})f(x_{n-1})$$

This holds only approximately but, hopefully, the approximation will get better and better as the number of intervals increases and the length of each one goes to 0. *We are thus led, given a continuous function f to study the sum* $(x_1 - a)f(a) + (x_2 - x_1)f(x_1) + \ldots + (x_2 - x_1)f(x_1)$."

Conclusion. We see once more that calculating with functions using their approximations makes matters far more natural than, to use Lagrange's expression, "*seeing derivatives in isolation*". Calculus is about calculating on functions.

Bibliography

1. I. C. Bivins. What a Tangent Line is When it isn't a Limit. *The College Mathematics Journal*. **17**(2) (1986) 133-143.
2. N. Bleistein and R. Handelsman A. "Asymptotic expansions of integrals." 1975 Holt, Rinehart and Winston. New York.
3. T. Bröcker and L. Lander. "Differentiable germs and catastrophes." London Mathematical Society Lecture Notes. 1975 Cambridge University Press. London.
4. A. M. Bruckner and J. J. Leonard. "Derivatives." Papers in Analysis. Mathematical Association of America. Washington D. C.
5. J. D. Cole. "Perturbation Methods in Applied Mathematics." A Blaisdell Book in Pure and Applied Mathematics. Carrier ed. 1968 Blaisdell. Waltham.
6. A. Denjoy. Sur l'intégration des coefficients différentiels d'ordre supérieur. *Fundamenta Mathematicae*. **25** (1916) 273-326.
7. J. Dieudonné. "Infinitesimal Calculus." 1968/1971 Hermann/Houghton Mifflin Company. Boston.
8. R. G. Douglas. Steps toward a lean and lively calculus. *Toward a lean and lively calculus*. (1986). Douglas Ed. The Mathematical Association of America. No. 6.
9. Dudley. Review of "Calculus with Analytic Geometry" by George F. Simmons. *The American Mathematical Monthly*. **95**(9) (1988) 888-892.
10. A. Erdélyi. "Asymptotic Expansions." 1956 Dover. New York.
11. R. L. Finney and D. R. Ostbey. "Elementary differential equations with linear algebra." 1984 Addison Wesley. Reading.
12. W. Freed. "Infinitesimal Calculus." 1986 Concordia College. Edmonton.
13. A. Gleason. The geometric content of advanced calculus. *CUPM Geometry Conference*. (1967). Durst Ed. Mathematical Association of America. 2. 53-119.
14. E. Goursat. "A course in Mathematical Analysis." 1904 Dover–Ginn. New York.

15. R. W. Hamming. Review of 'Toward a Lean and Lively Calculus'. *The American Mathematical Monthly*. **95**(5) (1988) 466-471.
16. E. Hille. "Analysis." Blaisdell Books in Pure and Applied Mathematics. Springer ed. 1964 Blaisdell. New York.
17. H. J. Keisler. "Elementary Calculus." 1976 Prindle, Weber & Schmidt. Boston.
18. J. L. Lagrange. "Théorie des fonctions analytiques." 1797 Gauthier-Villars. Paris.
19. S. Lang. "Analysis I." 1976 Addison Wesley. Reading.
20. H. Levi. "Polynomials, Power Series and Calculus." 1968 Van Nostrand. Princeton.
21. C. C. MacDuffee. Objectives in Calculus. *American Mathematical Monthly*. **54** (1947) 335-337.
22. F. Mattei and A. Schremmer. Differential Calculus, a Lagrangian Approach. Vol. 1 and 2. 1988.
23. M. Z. Nashed. "Some remarks on variations and differentials." Papers in Analysis. Mathematical Association of America. Washington D. C.
24. H. Oliver. The exact Peano Derivative. *Transactions of the American Mathematical Society*. **76** (1954) 444-456.
25. E. Picard. "Traité d'Analyse." Cours de la Faculté des Sciences de Paris. 1901 Gauthier–Villars. Paris.
26. T. Poston and I. Stewart N. "Taylor expansions and catastrophes." Research Notes in Mathematics. 1976 Pitman Publishing. London.
27. A. Robinson. "Non-standard analysis." 1966 North Holland. Amsterdam.
28. F. Schremmer and A. Schremmer. The Differential Calculus as Language. *Bulletin of Science, Technology and Society*. **8** (1988) 411-418.
29. F. Schremmer and A. Schremmer. Integrated Precalculus-Differential Calculus: A Lagrangian Approach. *The AMATYC Review*. **11**(1 (Part 2)) (1989) 28-31.
30. F. Schremmer and A. Schremmer. An introduction to Lagrangian differential calculus. *AMATYC Review*. (1990)
31. J. W. Young and F. M. Morgan. "Elementary Mathematical Analysis." 1922 MacMillan. New York.