Elementary Differential Calculus

ALAIN SCHREMMER
Department of Mathematics, Community College of Philadelphia, Philadelphia, PA 19130
FRANCESCA SCHREMMER
Department of Mathematical Sciences, West Chester State University, West Chester, PA 18383

0. Introduction. The currently prevailing emphasis in Differential Calculus on the derivative at the expense of, for instance, differentiability does not result from mathematical or pedagogical considerations. For instance, why should the definition of continuity, a property of f, be followed with the definition of the derivative, a function, rather than with the definition of differentiability, a stronger property of f ? Or, to put it another way, why should Differential Calculus be based on the notion of derivative in dimension 1 when it must be based on that of differentiability in all other dimensions? Typically, students are warned that "[while] *for functions of one variables, the terms "differentiable" and "has a derivative" are synonymous,* [...] *for functions of two variables differentiability is a more stringent requirement than the existence of partial derivatives*" [**1**] but little explanation is offered and the subsequent treatment of the total derivative is generally disappointing[1].

But, to quote Dieudonné [**9**] at some length, "*the fundamental idea of Calculus* [is] *the 'local' approximation of functions by* linear *functions. In the classical teaching of Calculus, this idea is immediately obscured by the accidental fact that, on a one-dimensional vector space, there is a one-to-one correspondence between linear forms and numbers, and therefore the derivative at a point is defined as a* number *instead of a* linear form. *This slavish subservience to the shiboleth of numerical interpretations at any cost becomes much worse when dealing with functions of several variables: one thus arrives, for instance, at the classical formula giving the partial derivatives of a composite function, which has lost any trace of intuitive meaning, whereas the natural statement of the theorem is of course that the (total) derivative of a composite function is the composite of their derivatives, a very sensible formulation when one thinks in terms of linear approximations*."

For real valued functions to represent the way situations change, the *differential* calculus, the "mathematics of change", must derive *local information*—about (mostly) gradual[2] changes—from punctual information. The desired information can be *qualitative*—is f near x_0 positive/negative, increasing/decreasing, concave up/concave down?— or *quantitative*—what is the approximate value, rate of change, concavity of f near x_0? But, qualitatively, we might also want to know whether, at x_0, f is continuous or differentiable while, quantitatively, we might ask what the jump or the slope is. That "fat calculus texts" are topics based and far from being organized along such lines is probably the single most important reason why they have turned into "cookbooks": It is difficult to see how the "*fragmentation problem* [...] *can be resolved by adding an additional semester to the calculus course*" [**2**].

We hope to show that to study functions by way of their local polynomial approximations is considerably more natural than, to quote Lagrange, "*seeing derivatives in isolation*". Specifically, we will argue that the systematic use of polynomial approximations has for the differential study of functions of one real variable much the same advantages that the use of decimal numbers has for the study of real numbers in that: **i.** it organizes, unifies and simplifies it [**13**], **ii.** it extends, naturally, to the Frechet derivative in multivariable calculus [**12**], to Banach Spaces, and to jets in Differential Topology [**6**] and, last but not least, **iii.** not only are the "naïve" proofs in this setting natural and plausible, but they are easily made rigorous.

The benefits of this approach, essentially due to Lagrange [**15**] and [**16**] have begun to be recognized. For instance, I. Bivins has by now received *two* prizes for his article "What a Tangent Line is When it isn't a Limit" [**4**]. The committee's citation for the Polya prize reads in part: "*By defining the tangent line as the best linear approximation to the graph of a function near a point,* [Bivins] *has narrowed the gap, always treacherous to students, between an intuitive idea and a rigorous definition. The subject of this article is* fundamental to the first two years of college mathematics *and should simplify things for students....*" (Emphasis added). Nevertheless, the intuitiveness of the tangent-as-limit-of-a-secant remains often unquestioned. For instance, in an article advocating Carathéodory's definition of the derivative [**14**], the linear approximation definition is merely mentioned as a "*variation*" to be found for instance in [**5**] and [**19**] even though"[t]*his approach has the intended additional benefit of making transparent the linear approximation of the tangent line*".

---

[1] These two aspects, the Gateau or directional derivative, which in dimension 1 is the rarely introduced *sided*-derivative, and the Frechet or total derivative, which, in dimension 1, is identified to its (1x1) matrix, are indeed merged in the conventional treatment of dimension 1 and therefore hard to distinguish. Moreover, the terminology is not consistent when we go from dimension 1 to dimension 2.

[2] A theory of *abrupt* changes is provided by Catastrophe Theory.

In any case, Lagrange's approach is simply the extension of this idea to the use, algebraically, of best polynomial approximations and, geometrically, of osculating curves of degree n.

1. Preliminary. A function f can be as simple as a polynomial or as complicated as a fractal. Thus, a natural idea when studying $f(x_0+h)$, the value of f near a point x_0, is to try to separate a *principal part*, that is a part smooth enough to be relevant to the information being sought, from a *remainder*, a part too small to be significant in that regard. We thus distinguish $P^{(n)}(x_0,h)$, a polynomial part of degree n in $h = x-x_0$, and a remainder $R^{(n)}(x_0,h)$ small enough that, compared to $P^{(n)}(x_0,h)$ and for the given purpose, it can be neglected. Henceforth, we shall take the reference to x_0 for granted and just write

$$f(x_0+h) = P^{(n)}(h) + R^{(n)}(h)$$

where $P^{(n)}(h) = A_0 + A_1h + A_2h_2 + \ldots + A_n{}^n$ and $R^{(n)}(h) = \mathrm{o}[h^n]$ which we read as saying that $R^{(n)}(h)$ approaches 0 faster than h^n, that is $\lim_{h\varnothing 0} \dfrac{R^{(n)}(h)}{h^n} = 0$; graphically, this means that the graph of $|R^{(n)}(h)|$ is under the graph of $|h|^n$ in a neighborhood of 0. In other words, we are using *asymptotic* expansions (as opposed to *series* expansions) with the power functions as *gauge* functions [**10**]. For beginning students it is enough to point out that the principal part carries the relevant quantitative information and that the remainder carries only the qualitative information that $P^{(n)}(h)$ differs from $f(x_0+h)$ by a small amount and just write $f(x_0+h) = P^{(n)}(h) + (\ldots)$.

2. Qualitative considerations. Since constant functions have no jump, it is natural to ask what can be said of a function f whose principal part is a constant function, i.e. that is such that $f(x_0+h)$ can be expressed as the sum of a constant part plus a part small enough not to cause a jump:

$$f(x_0+h) = A_0 + \mathrm{o}[1]$$

Note that A_0 is unique as any other constant approximation will yield a worse remainder and that in fact $A_0 = f(x_0)$. So, we say that a function is **continuous at x_0** iff it has a Best Constant Approximation BCA f near x_0. For instance, to show that $f(x) = ax^2 + bx + c$ is continuous at x_0, we write

$$\begin{aligned} f(x_0+h) &= (x_0+h)^2 + a(x_0+h) + c \\ &= [ax_0{}^2 + bx_0 + c] + h[2ax_0 + b + h] \end{aligned}$$

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

Similarly, to show that $f(x) = \dfrac{1}{x-2}$ is continuous at $x_0 \neq 2$, we compute $f(x_0+h) = \dfrac{1}{x_0+h-2}$ 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 $f(x_0+h) = \dfrac{1}{x_0-2} - \dfrac{h}{(x_0-2)(x_0-2+h)} = f(x_0) + \mathrm{o}[1]$ as $h\varnothing 0$.

Thus, continuity, a calculus notion, appears here as a generalization of constancy, a precalculus notion:

DEFINITION. *A function is continuous iff it is locally approximately constant.*

If nothing else, this makes it immediately plausible why a function continuous over a closed interval should be Riemann integrable.

Almost by definition, we have:

THEOREM. *If f is continuous at x_0, then if f is positive (resp. negative) at x_0, then f is positive (resp. negative) near x_0.*

Since the average rate of change of an affine function between two points is independent of these points and therefore constant, it is natural to ask what can be said of a function f whose principal part is an affine function, i.e. that is such that $f(x_0+h)$ can be expressed as the sum of an affine part plus a part small enough not to affect the slope:

$$f(x_0+h) = A_0 + A_1 h + \mathrm{o}[h]$$

Since A_1 too is unique as any other affine approximation will yield a worse remainder, we say that a function f is **differentiable at x_0** iff it has a Best Affine Approximation BAA f near x_0:

DEFINITION. *A function is differentiable iff it is locally approximately affine.*

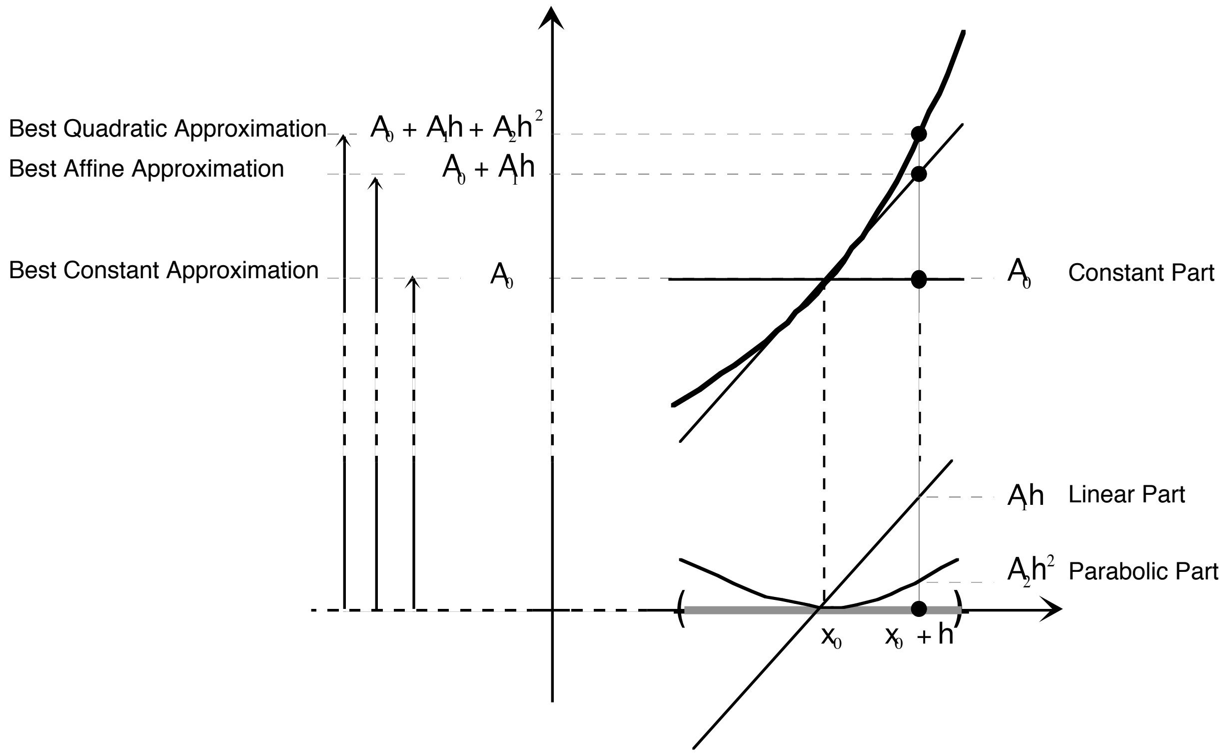
so that, again, a calculus notion, differentiability, appears here as a generalization of a precalculus notion, affinity. More generally,

DEFINITION. *A function is n-differentiable iff it is locally approximately polynomial of degree n.*

What can we say of a function f that is differentiable at x_0? Since affine functions are obviously continuous, f is necessarily continuous at x_0. *Numerically*, we can approximate $f(x_0+h)$ by $f(x_0) + A_1 h$ with an error $\mathrm{o}[h]$ for h small enough, but how small is small enough? We still have no bound on the error made in the approximation so that, from the numerical viewpoint, while better than before, the information is still insufficient. Such a bound will only be given by the Mean Value Theorem. Turning to the *geometry* of the graph, we would like

THEOREM. *If f is differentiable at x_0, then if A_1 is positive (resp. negative) at x_0, then f is increasing (resp. decreasing) near x_0.*

But, from $A_1>0$, if we get that $f(x)$ is larger than $f(x_0)$ in some neighborhood of x_0, we don't get that it is increasing. For that we need to know how A_1 changes in the neighborhood of x_0 which calls for the introduction of the function whose value at x_0 is A_1. Nevertheless, at this point, a local *approximate graph* can plausibly be sketched:



and from this we can see, from a qualitative viewpoint, that, when x_0 is *regular* as in the above figure, the constant part of $f(x_0+h)$ determines whether the function is positive/negative near x_0, the linear part whether it is increasing/decreasing, and the parabolic part whether it is concave up/down. When x_0 is *critical*, that is when one or more part is zero, the information will be provided by the corresponding next non-zero part. For instance, it is clear that, in order for f to have a local extreme at x_0, A_1 has to be 0 and the next non-zero part has to be of even degree.

The graph of BAA $f(x_0+h)$ is the osculating line of degree 1 to the graph of f at x_0, i.e. its tangent. Thus, to obtain the tangent of $f(x) = \dfrac{x^2 - 1}{x - 2}$ at 3, we write $f(3+h) = \dfrac{(3+h)^2 - 1}{3+h - 2} = \dfrac{8+6h+h^2}{1 + h}$ and divide in ascending powers:

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

to get $f(3+h) = 8 - 2h + \dfrac{3h^2}{1 + h}$ where $8 = f(3)$ and, since $\dfrac{3h^2}{1 + h} = o[h]$, $A_1 = -2$. In other words, we see how f looks under x-magnification. We obtain the equation of the tangent by "delocalizing" BAA $f(3+h) = 8 - 2h$ to $T_3 f(x) = 8 - 2(x-3)$ and the equation of the osculating parabola by delocalizing the Best Quadratic Approximation.

3. How to get the principal part. For polynomial functions, we just need the binomial theorem. For rational functions, we need division of polynomials in *ascending* powers to show that, away from its poles, a rational function is locally approximately polynomial. Near ∞, we divide in *descending* powers. This parallels how we divide numbers[3]. Note that the behaviour of negative-power functions near 0 and ∞ is a new one and that, by

---

[3] Any (positive) number is the sum of an integer and of a number between 0 and 1. An integer 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

including them as gauges, we can study rational functions near their poles and ∞ just as easily as near any other point.

If the remainder is not 0, it is often useful to have at least one non-zero "decimal". For example, near ∞, $\frac{x^2}{x^2+1} = 1 - \frac{1}{x^2} + \frac{x^2}{x^2(x^2+1)}$, where 1 is the principal part, $-\frac{1}{x^2}$ is the "first decimal", and $\frac{x^2}{x^2(x^2+1)}$ is the remainder. When x is near 0, we have, for example, $\frac{x^2}{x^2+1} = \frac{x^2}{1+x^2} = x^2 - \frac{x^4}{1+x^2}$, where x^2 is the principal part and $\frac{x^4}{1+x^2}$ is the remainder.

In the case of algebraic (resp. transcendental) functions defined as solutions of functional (resp. differential) equations, the method of undetermined coefficients gives, near the initial point, an *approximate polynomial solution* whose properties announce those of the exact solution. After we obtain an (approximate) addition formula, we can obtain approximate polynomial solutions near other points but this involves passing from the local to the global. We thus get a good approximate local study of transcendental functions except near ∞. Once again, the behaviour of these functions at ∞ is new and makes it necessary to include some of them as gauges. For instance, after we observe that, near ∞, e^x cannot be approximated by any power function, we include it as a gauge to study further functions.

Such a study of the approximate solutions serves as an excellent introduction to the study, assuming their existence [**20**], of the exact solutions carried out as in, for instance, [**11**], [**17**]; the approximate solutions can then be shown to be the Taylor approximations of the exact solutions.

4. Quantitative considerations. The coefficients of $P^{(n)}(h)$ give quantitative as well as qualitative information. For example, A_1 is easily seen to be the instant rate of change of f at x_0: Since

$$A_1 = \frac{f(x_0+h) - f(x) - \mathrm{o}[1]}{h}$$

we have that $A_1 = \lim_{h\varnothing 0} \frac{f(x_0+h) - f(x)}{h}$. In other words, the rate of change of a function is that of its BAA.

But the notion of instant rate is not nearly as primary as is usually taken for granted and the prevalent attitude that consists in looking at $2A_2$ as the instant rate of change of A_1 is not particularly more natural than looking at A_2 as "bending" the graph. While Galileo's principle, for instance, is usually interpreted as saying that, in the absence of force, the velocity does not change, it can also be interpreted as saying, arguably just as naturally, that the trajectory in space-time is rectilinear. A force can thus be interpreted as either causing the velocity to change—i.e. as causing an acceleration—or as causing the trajectory to deviate from rectilinearity—i.e. as causing a local parabolicity which has the advantage of being observable in $f(x_0+h)$ as well as in the graph.

Linearization indeed consists in considering intervals of time dt sufficiently small for the force not to have the time to affect rectilinearity noticeably so that trajectories are locally approximately rectilinear. Even if dt is big enough for the force to affect the trajectory, when the force is changing we take dt still small enough that we can assume the force to be constant during that time so that the trajectories are locally approximately parabolic. Thus, apart from better numerical approximations for $f(x_0)$, what we gain by considering approximations of degree higher than 1 is the notion of "bending" of the graph of $f(x)$ at x_0 which, whatever else it is, should be the degree of concavity of the osculating parabola[4].

5. From local to global: Derivatives. By now the need for introducing a function $f\,'$ whose value at x_0 is A_1 is clear: we need to follow how the coefficient A_1 changes with x_0. But, after this, we can adopt either one of two courses: We can define recursively $f^{(n)}(x) = (f^{(n-1)}(x))'$ and define a function f to be n-recursively-differentiable at x_0 iff $f^{(n)}(x_0)$ exists. Alternatively, we can say that f is n-Lagrange-differentiable at x_0 iff f has an osculating polyno-

dominant one. For example, $1{,}349 = 1\bullet 10^3 + 3\bullet 10^2 + 4\bullet 10^1 + 9\bullet 10^0$ 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 in order of decreasing 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 $0.085 \approx 8\bullet\left(\frac{1}{10}\right)^2$.

4 Curvature is the wrong concept here as circular trajectories result from *central* forces.

mial of degree n, that is if there is a polynomial $P^{(n)}(h) = \sum_{k=0}^{k=n} A_k h^{\;n}$ such that $f(x_0+h) = P^{(n)}(h) + \mathrm{o}[h^n]$, $h\varnothing 0$, and then define the k^{th} Lagrange derivative of f, also known as Peano derivative [**7**], to be the function whose value at x_0 is $A_k \bullet k!$[5]. If $f(x)$ admits such an approximation, it must be unique and we must have $A_0 = f(x_0)$, $A_1 = f\,'(x_0)$. It is then natural to ask whether there is a further connection between the two definitions of differentiability and, if so, find the relation between $A_k \bullet k!$ 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 $\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{h^k}{k!}$ as a candidate for the osculating polynomial because its first n derivatives at x_0 agree with those of f at x_0. In fact, we have:

THEOREM. *If* $f^{(k)}(x_0)$ *exists for* $k = 0, 1, \ldots n$, *then* $f(x) = \sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{h^k}{k!} + \mathrm{o}[h^n]$

PROOF. Use L'Hôpital's rule repeatedly on

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

• On the other hand, the existence of an osculating polynomial of degree $n>1$ at x_0 does not insure the existence of any derivative of order >1 at x_0: As a simple counter example [**8**], let

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

Near 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 $0x^2$ is not equal to $\frac{f''(0)}{2!}$ as, from $f'(x) = -x\cos\frac{1}{x} + 3x^2 \sin\frac{1}{x}$, we see that $f''(0)$ does not exist and so cannot be the coefficient of x^2 in the osculating polynomial. But this needs not be a matter of concern as it can be shown that if A_n exists in a neighborhood of x_0 and is bounded either from above or from below, then it *is* the n^{th} recursive-derivative of f: For the purpose of the first year calculus, the two notions are equivalent.

The usual rules are easily proved because quite systematically: In order to find the derivative of $f * g$, for any $*$, we just look for the coefficient of h in $[f * g](x_0+h)$. For instance, here is the proof of the quotient rule:

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

and, by division in *ascending* powers, since $\mathrm{o}[h] = h\bullet\mathrm{o}[1]$,

$$\begin{array}{r|l} & \dfrac{f(x_0)}{g(x_0)} \quad + \quad \dfrac{1}{g(x_0)}\left[f\,'(x_0) - \dfrac{f(x_0)}{g(x_0)}g\,'(x_0)\right]h \\ \hline g(x_0) + g\,'(x_0)h + h\bullet\mathrm{o}(1) & f(x_0) \quad + \quad f\,'(x_0)h \; + h\bullet\mathrm{o}(1 \end{array}$$

---

[5] According to Lagrange, this is the way Newton had first proceeded except that he had omitted the $k!$ and that, annoyed to have been corrected by one of the Bernoullis, he started afresh.

$$\frac{f(x_0) \quad + \quad \frac{f(x_0)}{g(x_0)}\, g'(x_0)h \quad + h\bullet o(1}{\left[f'(x_0) \; - \frac{f(x_0)}{g(x_0)}\, g'(x_0)\right] \; h \quad + h\bullet o(1}$$

which gives

$$\left[\frac{f}{g}\right](x_0+h) \;\; = \;\; \frac{f(x_0)}{g(x_0)} \; + \frac{f'(x_0)g(x_0) \; - f(x_0)g'(x_0)}{g(x_0)^2} \; h \; + \; h\bullet o(1) \; .$$

Here is the proof of the chain rule:

By the differentiability of f at x_0: $f(x_0+h) = f(x_0) + f'(x_0)h + h\bullet o[1]$, $h \varnothing 0$, and by the differentiability of g at $f(x_0)$: $g\big(f(x_0) + k\big) = g\big[f(x_0)\big) + g'\big(f(x_0)\big]\bullet k + k\bullet o(1)$, $k \varnothing 0$. Then,

$$\begin{aligned}
g[f(x_0+h)] \;\; &= \;\; g\big(f(x_0+ k)\big) \qquad\qquad \text{where} \qquad\qquad h\bullet\big[\, f'(x_0) + h\bullet o[1]\big] \\
&= g\big(f(x_0)\big) + g'\big(f(x_0)\big)\bullet k + k\bullet o[1] \\
&= g\big(f(x_0)\big) + g'\big(f(x_0)\big)\bullet h\bullet\big[f'(x_0) + o[1]\big] + h\big[\, f'(x_0) \; + \; o[1]\big] \;\; \bullet o[1] \\
&= g\big(f(x_0)\big) + g'\big(f(x_0)\big)\bullet f'(x_0)\; h + g'\big(f(x_0)\big)\bullet h\bullet o[1] + h\big[\, f'(x_0) \; + \; o[1]\big] \;\; \bullet o[1]
\end{aligned}$$

where the remainder is plausibly small (With beginning students, remainders can be denoted with just ellipses (...)). For a *proof*, if $h \varnothing 0$, then so does k and hence if a function is $o_h[1]$, then it is also $o_k[1]$ and the remainder is $h\bullet o_k[1]$.

Looking at the *topology* of the graph, we have 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:*

$$(f^{-1}(f(x)))'|_{x=x_0} \;\; = \frac{1}{f'(x_0)}$$

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: For example, $f(x) = x + x^3 \sin\frac{1}{x}$, $x \neq 0, f(0) = 0$.) To show that f' is differentiable, we check that $f^{-1}\big[\, f(x_0)+k\big]$ is approximately affine. We have

$$\begin{aligned}
f(x_0) + k \;\; &= f(x_0+h) \\
&= f(x_0) + f'(x_0)h + h\bullet o[1] \qquad\qquad \text{with} \qquad\qquad h = \frac{k}{f'(x_0)} \; - \frac{h\bullet o[1]}{f'(x_0)}
\end{aligned}$$

Then,

$$\begin{aligned}
f^{-1}\big[\, f(x_0)+k\big] \;\; &= x_0 + h \\
&= f^{-1}\big[\, f(x_0)\big] \;\; + \;\; \frac{1}{f'(x_0)} \; k \;\; - \; \frac{h}{f'(x_0)} \;\; \bullet o[1]
\end{aligned}$$

in which the remainder $-\frac{h}{f'(x_0)} \;\; \bullet o[1]$ is plausibly small.

For a *proof*, the remainder must be shown to be $k{\bullet}\mathrm{o}_k[1]$. From $k = h\big(f\,'(x_0)+\mathrm{o}_h[1]\big)$, $h\varnothing 0$ implies $k\varnothing 0$ and, since $f\,'(x_0) \neq 0$, $k\varnothing 0$ implies $h\varnothing 0$ so that $\mathrm{o}_h[1]$ iff $\mathrm{o}_k[1]$. Then, $\dfrac{h}{f\,'(x_0)}\ {\bullet}\mathrm{o}_h(1) = \dfrac{k{\bullet}\mathrm{o}_k[1]}{f\,'(x_0)\big(f\,'(x_0)+\mathrm{o}_k[1]\big)} = k{\bullet}\mathrm{o}_k[1]$.

In order to classify critical points we get from

$$f(x_0+h) = f(x_0) + f\,'(x_0)\,h + \frac{f''(x_0)}{2!}\ h^2 + \mathrm{o}[h^2],\ \ h\varnothing 0$$

that the second derivative test to classify the non-degenerate critical points of f is trivial as is, in this context, the classification of all critical points: Let x_0 be a critical point of f 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, we have:

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

6. Global considerations. We would also like to obtain properties of continuous functions on an interval (e.g. the Intermediate Value Theorem). Clearly, having defined continuity at x_0 by the local existence of a best constant approximation will not help here but does point very clearly where the difficulty lies in proving a theorem like

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

Because f is continuous on an interval, say $[a,b]$, $\forall x_0 \in [a,b], f(x_0+h) = f(x_0) + \mathrm{o}(1)$. Suppose h is in a neighborhood of 0, whose size depends on x_0, such that $\mathrm{o}(1) < \frac{1}{10}$ for example. If we knew that we could cover $[a,b]$ using finitely many of these intervals, say N, then $\big|f(x)-f(a)\big|$ would be bounded by $\frac{N}{10}$ and the theorem would be proved. This raises the question as to whether, from any open covering of a closed bounded interval, we can extract a finite one and, indeed, we do need compactness to prove Rolle's theorem.

In this context, the Mean Value Theorem is seen as a remainder theorem, that is as providing us with 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 $h{\bullet}f\,'(c)$ 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:

$$f(x_0+h) = f(x_0) + f\,'(x_0)h + \ldots + \frac{f^{(n)}(x_0)}{n!}{\bullet}h^n + R^{(n)}(h)$$

in which $R^{(n)}(h) = \dfrac{f^{(n+1)}(c)}{(n+1)!}\ h^{n+1}$ 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 and, with some work, L'Hôpital's rule.

IMPORTANT REMARK. Even if first-year calculus students are not likely to make this confusion, we should stress that $\displaystyle\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{(x-x_0)^k}{k!}$ is *not* to be thought of in this context 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, we can do either one of two things:

→ *For fixed n*, we can make $|x-x_0|$ small (this was our viewpoint). For example, by integration by parts, we have:

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

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 $n\emptyset\infty$, 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 $n\emptyset\infty$ which leads to analytic functions theory. The theory is not local anymore as we are approximating *f* is a *fixed* neighborhood of x_0.

6. Integral calculus. In standard freshman *integral* calculus courses, the Riemann integral is the one new idea of mathematical importance and most textbooks motivate its study by the need to *define* the notion of area. In the present context, it is more natural to follow Picard [**18**] in motivating the relation between the antiderivative and the definite integral:

"*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*

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

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:*

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

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* $(x_1-a)f(a) + (x_2-x_1)f(x_1) + \ldots + (x_2-x_1)f(x_1)$." (Our translation).

In other words, stressing the "antiderivative aspect" is more important, *at this point* and for most calculus students, than stressing the "measure theoretic aspect" which, in practice, devolves into teaching integration techniques. Among other things, the antiderivative aspect leads directly to the indefinite integral $\int f$ consisting of all antiderivatives which, in multivariable calculus, leads to the notions of potential and exactness of forms.

7. Dynamical Systems. Of particular interest for students intending to pursue a career in sciences other than mathematics, physics or engineering, is the fact that the basic attitudes developed by the above treatment are precisely those needed in the study of Dynamical Systems, the first place where real applications become feasible, and which can therefore follow immediately the differential calculus sketched above.

Roughly speaking, the study of Differential Equations can be pursued from three different points of view: **i.** The *computational view point* where one searches for solutions in closed forms and/or series solutions, **ii.** The *numerical view point* where one develops algorithms to compute solutions numerically and **iii.** The *qualitative viewpoint* where one studies the geometrical features of the solutions. Until Poincaré, the computational aspect dominated almost completely the subject. In fact, it still holds a great attraction for many applied mathematicians and has even recently provided some spectacular insights in some long standing problems (e.g. Korteweg-de Vries equation, solitons). But, even in the rare case when a solution in closed form or a series solution can be found, and when the problem is therefore usually considered to have been solved, the solution is usually in so complicated a form as to necessitate difficult qualitative methods of investigation to describe its behaviour. In contrast, a direct study from the equation and the phase portrait is often quite feasible [**3**].

When studying differential systems of the form $\frac{dx}{dt} = f(x,t)$, where $x:\mathrm{R}^n \varnothing \mathrm{R}$, $n = 1, 2, 3$, generally arising directly from modelling applications, the goal is to obtain as much *information* as possible on its *solutions* and to present this information graphically. This involves going from a local viewpoint to a global viewpoint. As such, this is a completely natural continuation of the way, for instance, Lnx is studied here, that is as the solution of the Initial Value Problem :

$$\begin{cases} x' = \dfrac{1}{t} \\ x(1) = 0 \end{cases}$$

Naturally, we start with linear systems which we then use to approximate non-linear ones near their singular points because it is quickly realized that, locally and away from singular points, the flow of a dynamical system can always be "rectified" so that its local behaviour needs to be studied only in the neighborhood of the singular points. Once again, this is strongly reminiscent of results stressed earlier on and the approach is therefore quite a natural one.

In conclusion, the treatment of differential calculus sketched above, followed by such a study of Dynamical Systems, constitutes a self contained, compact, conceptually satisfying sequence that is completely accessible to "just plain folks", given that it ultimately rests only on algebraic skills. Yet, it is intellectually uncompromising and, as such, it should afford prospective mathematics majors a solid basis on which to proceed towards advanced calculus.

REFERENCES.

1. H. Anton. "Calculus with analytic geometry." 1988 John Wiley & Sons. New York.
2. H. Anton. In Defense of the Fat Calculus Text. *UME Trends*. **2**(6) (1991) 1.
3. M. Artigue and V. Gautheron. "Systèmes Differentiels, Etude Graphique." 1983 CEDIC/F. Nathan. Paris.
4. I. C. Bivins. What a Tangent Line is When it isn't a Limit. *The College Mathematics Journal*. **17**(2) (1986) 133-143.
5. W. Boyce and R. DiPrima. "Calculus." 1988 Wiley. New York.
6. T. Bröcker and L. Lander. "Differentiable germs and catastrophes." London Mathematical Society Lecture Notes. 1975 Cambridge University Press. London.
7. A. M. Bruckner and J. J. Leonard. "Derivatives." Papers in Analysis. Mathematical Association of America. Washington D. C.
8. A. Denjoy. Sur l'intégration des coefficients différentiels d'ordre supérieur. *Fundamenta Mathematicae*. **25** (1916) 273-326.
9. J. Dieudonné. "Foundations of Modern Analysis." Pure and Applied Mathematics. Eilenberg ed. 1960 Academic Press.
10. J. Dieudonné. "Infinitesimal Calculus." 1968/1971 Hermann/Houghton Mifflin Company. Boston.
11. R. L. Finney and D. R. Ostbey. "Elementary Differential Equations with Linear Algebra." 1984 Addison Wesley. Reading.
12. F. J. Flanigan and J. L. Kazdan. "Calculus Two: Linear and Non-Linear Functions." 1971 Prentice-Hall. Englewood Cliffs.

13. A. Gleason. The geometric content of advanced calculus. *CUPM Geometry Conference*. (1967). Durst Ed. Mathematical Association of America. 2. 53-119.
14. S. Kuhn. The Derivative à la Carathéodory. *The American Mathematical Monthly*. **98**(1) (1991) 40-44.
15. J. L. Lagrange. "Théorie des fonctions analytiques." Oeuvres de Lagrange. Serret ed. 1881 Gauthier-Villars. Paris.
16. J. L. Lagrange. "Leçons sur le calcul des fonctions." Oeuvres de Lagrange. Serret ed. 1884 Gauthier-Villars. Paris.
17. S. Lang. "Analysis I." 1976 Addison Wesley. Reading.
18. E. Picard. "Traité d'Analyse." Cours de la Faculté des Sciences de Paris. 1901 Gauthier–Villars. Paris.
19. M. H. Protter and C. B. Morrey. "A First Course in Real Analysis." 1977 Springer-Verlag. New York.
20. P. Renz. "Steps toward a rethinking of the foundations and purposes of introductory calculus." Toward a lean and lively calculus. Douglas ed. 1986 The Mathematical Association of America. Washington, D.C.