# Derivatives or Differential Calculus?

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

**0. Introduction.** The currently prevailing emphasis in Calculus I on the derivative at the expense of, for instance, differentiability is more the result of teaching habits than of mathematical or pedagogical considerations. For instance, why go from the definition of continuity, a property of *f*, to the definition of the derivative, a function, rather than to the definition of differentiability, a stronger property of *f* ? Or, to put it another way, why start with the definition of the derivative in dimension 1 when it is necessary to start with the definition of differentiability in dimensions larger than 1? 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*" «Anton, 1988 #111» but little explanation is offered and the subsequent treatment of the total derivative is disappointing[1].

In reality, the purpose of real valued functions is to represent the way situations change and that of the *differential* calculus, the "mathematics of change", is to derive *local information* about (mostly gradual) changes[2] from punctual information. The nature of the desired information depends on the situation. If only because a real number has a *sign* and a *magnitude*, the desired information can be *qualitative*—is *f* near $x_0$ positive/negative, increasing/decreasing, concave up/concave down?— or *quantitative*—what is the approximative value, rate of change, acceleration of *f* near $x_0$? But, qualitatively, we might also want to know whether, at $x_0$, $f$ is continuous (resp. differentiable) while, quantitatively, we might ask what the jump (resp. the slope) is. That "fat calculus texts" are not organized along such lines might be precisely at the root of the "calculus crisis".

We hope to show that to study a function by way of its 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 it organizes it, unifies and simplifies it «Gleason, 1967 #34», and, moreover, extends canonically to the Frechet derivative in multi-variable calculus «Flanigan, 1971 #78», Banach Spaces[3] «Dieudonné, 1960 #112», jets in Differential Topology «Bröcker, 1975 #68». Of particular interest for students intending to pursue a career in sciences other than mathematics, physics or engineering, is that the basic ideas are precisely those used in the study of Dynamical Systems. Last but not least, not only are the "naïve" proofs in this setting natural and plausible, but they can also easily be made rigourous.

The benefits of this approach, essentially due to Lagrange, «Lagrange, 1881 #17» and «Lagrange, 1884 #101», have recently 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" «Bivins, 1986 #32». 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, this view seems to elicit a certain amount of unconscious resistance. For

---

[1] There are indeed two aspects to the whole question which are merged in dimension 1 and therefore hard to distinguish. Moreover, the terminology is not consistent when we go from dimension 1 to dimension 2. One aspect is the Gateau or directional derivative, which in dimension 1 is the rarely introduced *sided*-derivative, and the other is the Frechet derivative or differential, which, in dimension 1, is identified to its (1x1) matrix.

[2] Catastrophe Theory is the theory of *abrupt* changes.

[3] In this context, Dieudonné "*aims at keeping as close as possible to the fundamental idea of Calculus, namely 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 correspondance 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.*"

instance, in an article advocating Carathéodory's definition of the derivative «Kuhn, 1991 #113», it is only mentioned as a "*variation*" to be found for instance in «Protter, 1977 #114» and «Boyce, 1988 #115». Even though he observes that "[t]*his approach has the intended additional benefit of making transparent the linear approximation of the tangent line*", the author does not seem to have realized its full implications.

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

**1. Preliminary.** The definition, if not the notion, of function is deceptively simple: A function $f$ can be as simple as a polynomial or as complicated as a fractal. Thus, the general idea when discussing $f(x_0+h)$, the value of $f$ near a point $x_0$, is naturally to separate the *principal part*, that is the part smooth enough to be relevant to the information being sought, from the *remainder*, the part too small to be significant in that regard. We thus distinguish $P^{(n)}(h)$, a polynomial part of degree $n$ in $h = x–x_0$, and a remainder $R^{(n)}(h)$ small enough that, compared to $P^{(n)}(h)$ and for the given purpose, it can be neglected:

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

where $P^{(n)}(h) = A_0 + A_1h + A_2h_2 + ... + A_n{}^n$ 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. For beginning students we just point out that $P^{(n)}(h)$ carries the relevant quantitative information and that $R^{(n)}(h)$ carries the qualitative information that $P^{(n)}(h)$ differs from $f(x_0+h)$ by only a small amount so that we just write

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

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

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

Note that $A_0$ must then be unique, that is that any other constant approximation will yield a worse remainder, and in fact that $A_0 = f(x_0)$. In other words, we are saying that a function is **continous at $x_0$** iff it has a Best Constant Approximation 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 $h\varnothing 0$.

Sinilarly, 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 $f(x_0+h) = \frac{1}{x_0-2} - \frac{h}{(x_0-2)(x_0-2+h)} = f(x_0) + \mathrm{o}(1)$ as $h \varnothing 0$.

We thus have:

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

which firmly relates a calculus notion, continuity, to a precalculus notion, constancy.

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[4] between two points $x_1$ and $x_2$ is independant of $x_1$ and $x_2$ and therefore constant, it is natural to ask what can be said of a function whose principal part is an affine function i.e. is such that $f(x_0+h)$ can be expressed as the sum of an affine part plus something small enough not to affect the slope:

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

Since, by contradiction, $A_1$ is also unique, that is any other affine approximation will yield a worse remainder, we are saying that a function is **differentiable at $x_0$** if it has a Best Affine Approximation near $x_0$ and we have:

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

which again firmly relates a calculus notion, differentiability, to a precalculus notion, affinity. More generally,

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

Suppose $f$ is differentiable at $x_0$, what can be obtained from it? Since affine functions are obviously continuous, diffentiability trivially implies continuity. *Numerically*, we can approximate $f(x_0 + h)$ by $f(x_0) + A_1 h$ with an error $h \bullet \mathrm{o}(1)$ for $h$ small enough, but how small is small enough? We still have no bound on the error made in the approximation. So, from the numerical viewpoint, the information is better than before but not sufficient. Concerning the *geometry* of the graph, we would like

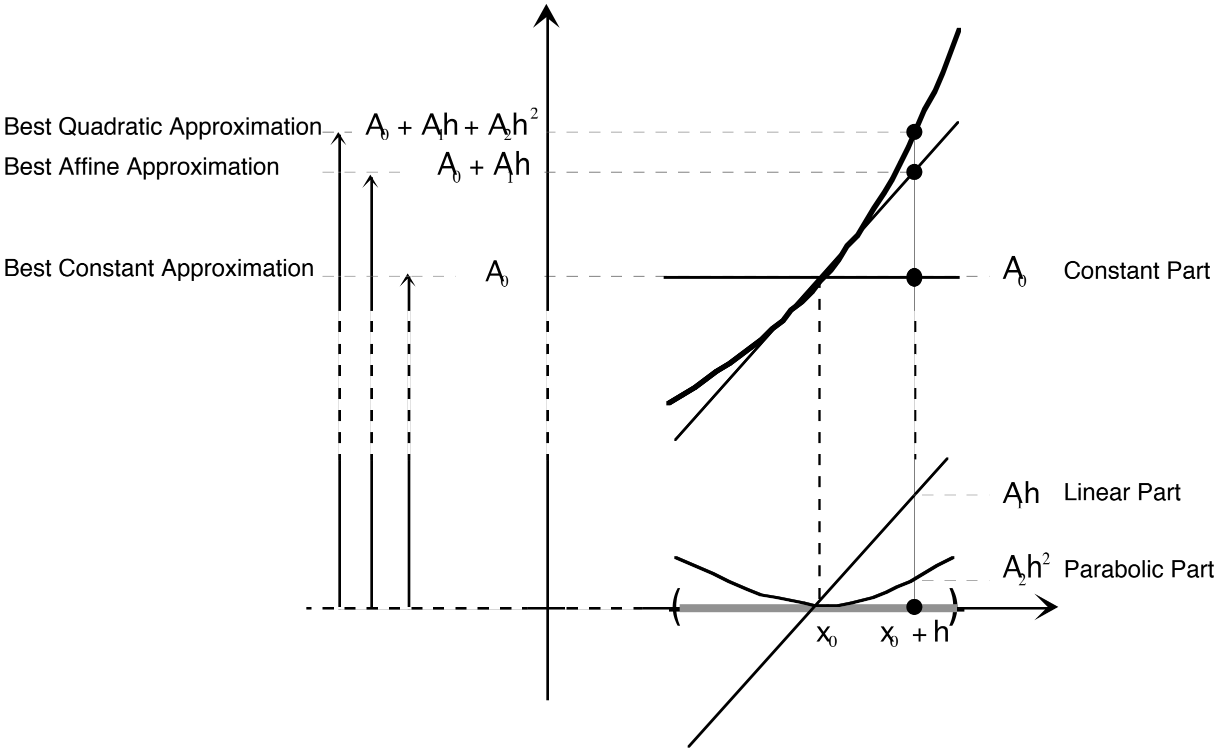
THEOREM. *If $f$ is differentiable at $x_0$, then if $f'$ is positive (resp. negative) at $x_0$, then $f$ is increasing (resp. decreasing) near $x_0$.*

But suppose that $f'(x_0)>0$; we can deduce that $f(x)$ is larger than $f(x_0)$ in some neighborhood of $x_0$ but from that we cannot deduce that $f(x)$ is increasing in the neighborhood of $x_0$ unless, for example and anticipating a bit, we assume that $f'(x)$ is continuous.

Nevertheless, at this point, an approximate *local graph* can be sketched:

---

[4] We prefer this to the term *linear function* for the obvious reason that affine functions are not linear.



and from this we can see, from a qualitative point of view, that, when $x_0$ is regular as in the above figure, the constant part of $f(x_0+h)$ shows 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, one or more part will be missing and the information will be provided by the corresponding next non-zero part. It is also easy to see 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 tangent to the graph of $f$ at $x_0$, i.e. its osculating line of degree 1. 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 $\dfrac{3h^2}{1 + h} = h \bullet o(1)$, $f'(3) = -2$. In other words, what we get is how $f$ appears under magnification. We obtain the equation of the tangent by "delocalizing" BAA $f(x_0+h) = 8 - 2h$ which gives $T_f(x) = 8 - 2(x - x_0)$. We would obtain the osculating parabola in just the same manner by delocalizing the Best Quadratic Approximation of $f$.

**3. How to get $P^{(n)}(h)$ in the case of "most" functions.** For polynomial functions, we just need the binomial theorem. And for rational functions, we just 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[5]. An interesting aspect of this approach is that, by including negative-power functions in the set of *gauges*, we can study rational functions near their poles 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, we use the method of undetermined coefficients to obtain an approximate solution near the initial point. Its properties anounce those of the exact solution. After we obtain an addition formula, we can even obtain approximate solutions near other points but this involves passing from the local to the global. We thus obtain a good local study of transcendental functions except near ∞. But, in a way, transcendental functions can be characterized by the fact that their behaviour at ∞ ia not polynomial which makes it in turn necessary to enlarge again the set of gauges «Dieudonné, 1968/1971 #52».

Such a study of approximate solutions can also serve as an introduction to an exact study done as in «Lang, 1976 #18» or «Finney, 1984 #16» and 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 information as well as qualitative. For example, $A_1$ is easily seen to be the instant rate of change of $f$ at $x_0$: Since

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

we have $A_1 = \lim_{h\varnothing 0} \frac{f(x_0+h) - f(x)}{h}$ and we can define the first derivative of $f$ as the function $f\,'$ whose value at $x_0$ is $A_0$. After this, we can follow either one of two courses: We can define recursively $f^{(n)}(x) = (f^{(n)}(x))'$ and define a function to be $n$-iterated-differentiable at $x_0$ iff $f^{\,(n)}(x_0)$ exists. Or, we can say that $f$ is $n$-Lagrange-Peano-differentiable at $x_0$ iff $f$ has an osculating polynomial of degree $n$, that is a polynomial $P^{(n)}(h) = \sum_{k=0}^{k=n} A_k h^{\,n}$ such that

$$f(x_0+h) = P^{(n)}(h) + o[h^n],\ \ h\varnothing 0$$

and then define the $n^{\text{th}}$ Lagrange-Peano derivative of $f$ to be the function whose value at $x_0$ is $A_k \bullet k!$[6].

If $f(x)$ admits such an approximation, it must be unique and $A_0 = f(x_0)$, $A_1 = f'(x_0)$. 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

---

[5] 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 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 declining 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 $0.085 \approx 8\bullet\left(\frac{1}{10}\right)^2$.

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

$\sum_{k=0}^{k=n} f^{(k)}(x_0)\frac{h^k}{k!}$ 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, 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 is $0x^2$ equal to $\frac{f''(0)}{2!}$ ? 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 to us 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 $n^{\text{th}}$ iterative-derivative of $f$. For the purpose of the first year calculus, the two notions are equivalent.

The usual rules are proved easily and naturally[7]: In order to find the derivative of $f * g$, for any $*$, we just expand $[f * g](x_0+h)$ and look for the coefficient of $h$. Here is, for instance 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,

$$\begin{array}{r|llr}
 & \frac{f(x_0)}{g(x_0)} & + \frac{1}{g(x_0)}\left[f'(x_0) - \frac{f(x_0)}{g(x_0)}g'(x_0)\right] h & \\
\cline{2-4}
g(x_0) + g'(x_0)h + h \bullet \mathrm{o}(1) & f(x_0) \; + & f'(x_0)h & + h \bullet \mathrm{o}(1) \\
 & f(x_0) \; + & \frac{f(x_0)}{g(x_0)} g'(x_0)h & + h \bullet \mathrm{o}(1) \\
\cline{2-4}
 & & \left[f'(x_0) - \frac{f(x_0)}{g(x_0)}g'(x_0)\right] h & + h \bullet \mathrm{o}(1)
\end{array}$$

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

and here is the proof of the chain rule: By the differentiability of $f$ at $x_0$, $f(x_0+h) = f(x_0) + f'(x_0)h + h \bullet 0[1]$, $h \varnothing 0$, and, by the differentiability of $g$ at $f(x_0)$, $g\big(f(x_0) + k\big) = g\big[f(x_0)\big) + g'\big(f(x_0)\big] \bullet k + k \bullet \mathrm{o}(1)$, $k \varnothing 0$. Then,

---

[7] In fact, mostly *by the students themselves.*

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

If $h\emptyset 0$, then so does $k$ and hence if a function is o[1] when $h\emptyset 0$, then it is also o[1] when $k\emptyset 0$.

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

$$= g\big(f(x_0)\big) + \boldsymbol{g'\big(f(x_0)\big)\bullet f'(x_0)}h + h\bullet \mathrm{o}[1],\ h\emptyset 0.$$

In the case of beginning students, we just replace the little o's by ellipses.

We now look at the *topology* of the graph:

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 $\xi$, 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, $f(x) = x + x^3\sin\frac{1}{x}$ , $x \neq 0, f(0) = 0$.)

Apart from a better numerical approximation for $f(x_0)$, what we have gained by considering higher degree approximations is the notion of curvature of the graph of $f(x)$ at $x_0$ which should be, whatever else it is, the curvature of the osculating parabola. Also, in order to classify critical points we get from

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

that the second derivative test to classify the non-degenerate critical points of $f$ is trivial. In fact, in the context of approximation, the classification of all critical points is clear: Let $x_0$ be a critical point of $f$ which we will assume, for simplicity, to be $C^\infty$.

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:

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

**5. From local to global.** We would like to obtain properties of continuous functions on an interval (e.g. 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.*

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 possibility and desirability of a characterization of a closed bounded interval by the property that, from any open covering, a finite one can be extracted. We thus need compactness to prove 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 L'Hôpital's rule.

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 $n^{\text{th}}$ 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, 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 $\infty$ as $n\varnothing\infty$, 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 $n\varnothing\infty$ 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.** 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 «Picard, 1901 #30» 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*

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

*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:*

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

*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* $(x_1 - a)f(a) + (x_2 - x_1)f(x_1) + \dots + (x_2 - x_1)f(x_1)$."

Among other things, this 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 and to the study of dynamical systems à la Poincaré.

**6. Dynamical Systems.** 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). On the other hand, 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.

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 applications, our 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, it is a completely natural continuation of the differential calculus where, for instance, Ln$x$ is studied as the solution of the Initial Value Problem :

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

Naturally, just as before, 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" and that its local behaviour needs to be studied only in the neighborhood of the singular points. Once again, this is a behaviour that is strongly reminiscent of results stressed earlier on.

REFERENCES.