A LAGRANGIAN APPROACH TO THE DIFFERENTIAL CALCULUS

Paper presented at the *What is happening with calculus revision* session of the
Joint AMS/MAA Phoenix Meeting, January 1989.

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

Supported in part by NSF Grant USE-8814000

0. Introduction.

The conventional approach to the teaching of the first year calculus is currently based on a conception in which it is made to depend, from the outset, whether explicitly or implicitly, on the notion of limit. That this approach is by no means the only possible one, let alone the most desirable one, has been completely forgotten to the point that limits are now considered to be absolutely indispensible to the comprehension of the calculus.

For the past several years, we have been developing an approach to the differential calculus based on Lagrange's idea that a function can be studied *locally* through its Taylor approximations obtained directly from its definition[1]. We illustrate some of the main points of this approach in the case of polynomial functions but it should be clear that the same ideas apply to "all" other functions. We will briefly indicate how to obtain approximations in "all" cases.

Some of the pedagogical advantages of Lagrange's viewpoint, the single most important of which is that it *empowers* the students by not reducing the calculus to a cookbook, are described in F. Schremmer and A. Schremmer (1989 a). See F. Schremmer and A. Schremmer (in press) for some aspects pertaining to "calculus literacy". F. Mattei & A. Schremmer (1988 a, b, c) are taskbook implementations of Lagrange's approach[2].

1. Best Polynomial Approximations.

Consider a function f which, for the sake of simplicity, we assume to be polynomial, and suppose that we are interested in the numerical values of f when x is *near* x_0. We begin by **localizing** f at x_0, that is by expressing $f(x)$ in terms of $x - x_0$. Suppose that we obtain something like

$$f(x) = 102.7 + 0.473(x - x_0) + 2.17(x - x_0)^2 + 3.14159(x - x_0)^3 + ...$$

or, setting $x = x_0 + h$,

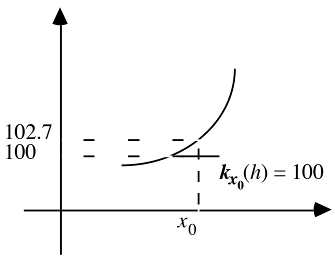
$$f(x_0 + h) = 102.7 + 0.473h + 2.17h^2 + 3.14159h^3 + ...$$

---

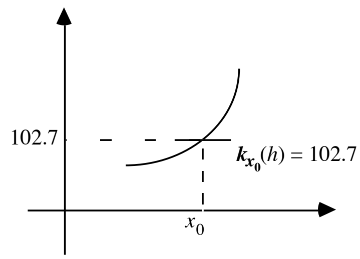
[1] In the January-February 1988 issue of FOCUS, mention is made of a 1987 MAA Award for Expository Excellence, the George Polya Award for articles in THE COLLEGE MATHEMATICS JOURNAL, to Irl. C. Bivens for "What a Tangent Line is When it isn't a Limit" in the March 1986 issue, pages 133-143. The committee's citation is quoted in part as: "*By defining the tangent line as the best linear approximation to the graph of a function near a point, [Bivens] 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). Lagrange's approach is of course just the extension of this idea to best approximations of any degree. In other words, what we are really dealing with are jets of differentiable functions.

[2] The interested reader is invited to write us for copies of any of these.

If we only want to have an idea of the value of $\boldsymbol{f}(x_0 + h)$, we can approximate $\boldsymbol{f}(x_0 + h)$ with **constant** functions, the simplest non-zero functions:

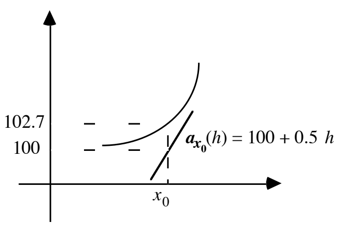


A good constant approximation

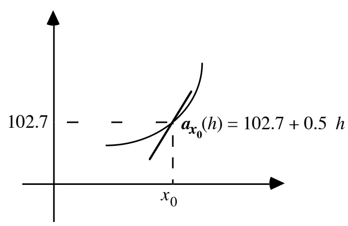


The best constant approximation

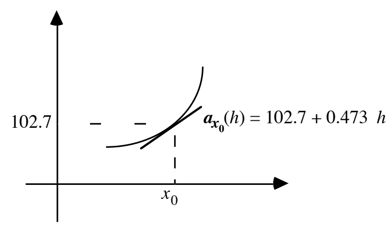
To see how $\boldsymbol{f}(x_0 + h)$ varies, we use **affine** functions, the simplest non-constant functions:



A good affine approximation



A better affine approximation



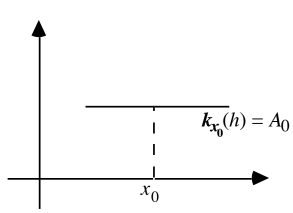
The best affine approximation

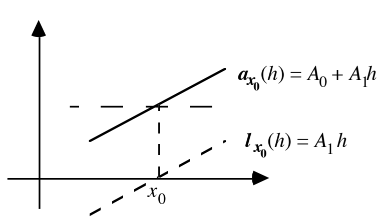
We can also approximate $\boldsymbol{f}(x_0 + h)$ with **quadratic** functions, the simplest non-affine functions. Thus, we might use $\boldsymbol{q}_{\boldsymbol{x}_0}(h) = 100 + 0.5h + 2h^2$ but the *best* quadratic approximation is $\boldsymbol{q}_{\boldsymbol{x}_0}(h) = 102.7 + 0.473h + 2.17h^2$ as the error is now "in the thousandths".

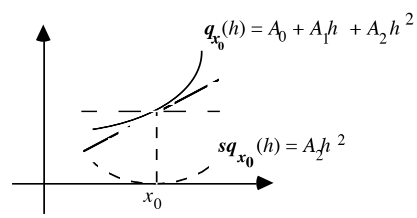
More generally, suppose that we have

$$f(x_0 + h) = \boldsymbol{f}_{\boldsymbol{x}_0}(h) = A_0 + A_1h + A_2h^2 + \ldots + A_nx^n + \ldots$$

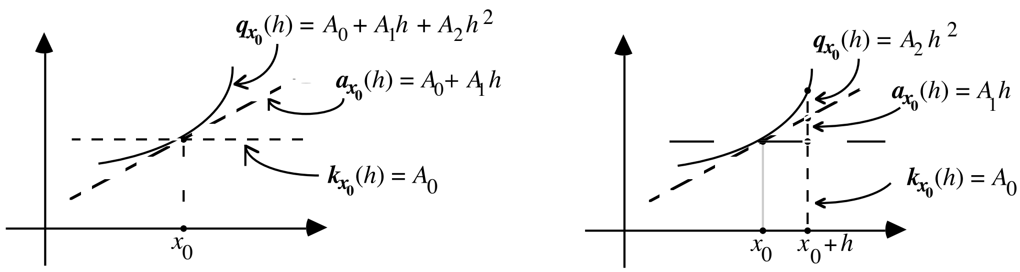
To obtain a **local graph** of a polynomial approximation $\boldsymbol{f}_{\boldsymbol{x}_0}(h)$ of $\boldsymbol{f}(x_0 + h)$ near x_0, we first draw the constant function $\boldsymbol{k}_{\boldsymbol{x}_0}(h) = A_0$, then the linear function $\boldsymbol{l}_{\boldsymbol{x}_0}(h) = A_1h$ using the graph of $\boldsymbol{k}$ as "base line", and then the parabolic function $\boldsymbol{p}_{\boldsymbol{x}_0}(h) = A_2h^2$ using the graph of the affine function $\boldsymbol{a}_{\boldsymbol{x}_0}(h) = A_0 + A_1h$ as base line, etc:







The constant function $\boldsymbol{k}_{\boldsymbol{x}_0}(h) = A_0$ gives the part of $\boldsymbol{f}(x)$ due to the value *at* x_0. The linear function $\boldsymbol{l}_{\boldsymbol{x}_0}(h) = A_1h$ gives the part of $\boldsymbol{f}(x)$ due to the *value* change as x changes away from x_0. The parabolic function $\boldsymbol{p}_{\boldsymbol{x}_0}(h) = A_2h^2$ gives the part of $\boldsymbol{f}(x)$ due to the *value change* change as x changes away from x_0, Altogether, we get



which gives a graphic representation of the polynomial approximation $\boldsymbol{f}_{x_0}(h)$.

2. Qualitative Analysis.

Much of what we do in calculus consists in using information *at* a point to obtain information *near* that point. We first give *elementary* definitions of the usual features of a function and then necessary and sufficient conditions in terms of Taylor polynomial approximations. It is useful to keep in mind that, among all power functions, constant and linear functions are pathological. Here, and again for the sake of simplicity, we assume x_0 to be neither a pole nor at È

The zero function has no sign. So, we define the **sign** of ***f*** *near* x_0 as the way $\boldsymbol{f}(x)$ differs from the zero function on each side of x_0. For example, sign ***f*** near x_0 is (+,–) if $\boldsymbol{f}(x) > 0$ when $x < x_0$ and $\boldsymbol{f}(x) < 0$ when $x > x_0$. To find how $\boldsymbol{f}(x)$ differs from the zero function, we just need the first *non-zero* term in the Taylor approximation. Indeed, because the terms are in order of descending importance, and are so by orders of magnitude, none of the further terms in the Taylor polynomial can affect the sign as given by the first term. Thus, the sign of ***f*** near x_0 is determined by the way the Least Non-Zero Approximation of $\boldsymbol{f}(x)$ differs from the zero function and since the LNZA is "usually" the constant part $A_1 = \boldsymbol{f}(x_0)$, we get the

SIGN THEOREM. When $\boldsymbol{f}(x)$ is neither 0 nor È *at* x_0, $\boldsymbol{f}(x)$ is "even signed"[1]. Moreover,
if $\boldsymbol{f}(x)$ is positive *at* x_0, then the sign of $\boldsymbol{f}(x)$ near x_0 is (+,+) and so $\boldsymbol{f}(x)$ is positive *near* x_0
if $\boldsymbol{f}(x)$ is negative *at* x_0, then the sign of $\boldsymbol{f}(x)$ near x_0 is (–,–) and so $\boldsymbol{f}(x)$ is negative *near* x_0

When $\boldsymbol{f}(x)$ is either 0 or È, $\boldsymbol{f}(x)$ near x_0 can be even or odd signed and the "unusual" points for the sign are thus the poles and the zeros of $\boldsymbol{f}(x)$—whether finite or infinite.

Similarly, we define the **variance** of ***f*** *near* x_0 as the way $\boldsymbol{f}(x)$ differs from its best constant approximation $\boldsymbol{k}_{x_0}(x)$ on each side of x_0. Then we need only approximate $\boldsymbol{f}(x_0 + h)$ to the first *non-constant* term to find the variance of ***f*** near x_0. Thus $\boldsymbol{f}_{x_0}(h) = A_0 + A_n h^n + \ldots$ shows that x_0 is a monotonic point with variance (Ω,Ω) or (œ,œ) or a turning point with variance (œ,Ω) or (Ω,œ) depending on the parity of n and on the sign of A_n. The Least Non-Constant Approximation of $\boldsymbol{f}(x)$ is "usually" the affine part $A_0 + A_1 h$ and since, anticipating on the definition of the derivative, we have $A_1 = \boldsymbol{f}'(x_0)$, we get:

VARIANCE THEOREM. When $\boldsymbol{f}'(x)$ is neither 0 nor È *at* x_0, $\boldsymbol{f}(x)$ is *monotonic*. Moreover,:
if $\boldsymbol{f}'(x)$ is positive *at* x_0, then the variance of $\boldsymbol{f}(x)$ near x_0 is (Ω,Ω) and so $\boldsymbol{f}(x)$ is increasing *near* x_0

---

[1] We use the term for the sake of terminological symmetry.

if $f'(x)$ is negative *at* x_0, then the variance of $f(x)$ near x_0 is (œ,œ) andso $f(x)$ is decreasing *near* x_0

When $f'(x)$ is either 0 or È, $f(x)$ near x_0 can be either monotonic or turning and the "unusual" points for the variance are thus the poles and the zeros of $f'(x)$, that is the critical points of $f(x)$.

From a different viewpoint, we define the **optimization** of f *near* x_0 as the way $f(x_0)$ compares with $f(x)$ on each side of x_0. Thus $f_{x_0}(h) = A_0 + A_n h^n + \ldots$ shows that x_0 is a "saddle point"[1] with optimization (min,max) or (max,min) or an extreme point with optimization (min,min) or (max,max) depending on the parity of n and on the sign of A_n. We get:

OPTIMIZATION THEOREM. When $f'(x)$ is neither 0 nor È *at* x_0, $f(x)$ has a *saddle point*. Moreover,

If $f'(x)$ is positive *at* x_0, then the optimization of $f(x)$ *near* x_0 is (max,min)
If $f'(x)$ is negative *at* x_0, then the optimization of $f(x)$ *near* x_0 is (min,max)

When $f'(x)$ is either 0 or È, $f(x)$ can have either a saddle point or an extreme point and the "unusual" points for optimization are thus the critical points of $f(x)$.

Finally, we define the **concavity** of f *near* x_0 as the way $f(x)$ differs from its best affine approximation $a_{x_0}(x)$ on each side of x_0. Then we need only approximate $f(x_0 + h)$ to the first *non-linear* term to determine the concavity of f near x_0. Thus, $f_{x_0}(h) = A_0 + A_1 h + \ldots + A_n h^n + \ldots$ shows that x_0 is a "curling" point with concavity ($%,$%) or (^&,^&) or an inflection point with concavity ($%,^&) or (^&,$%) depending on the parity of n and on the sign of A_n. Since the Least Non-Affine Approximation is "usually" the quadratic part $A_0 + A_1 h + A_2 h^2$ and since $A_2 = f''(x_0)/2$, we have:

CONCAVITY THEOREM. When $f''(x)$ is neither 0 nor È *at* x_0, $f(x)$ has a "curling point". Moreover,

if $f''(x)$ is positive *at* x_0, then the concavity of $f(x)$ is ($%,$%) and so $f(x)$ is concave UP *near* x_0
if $f''(x)$ is negative *at* x_0, then the concavity of $f(x)$ is (^&,^&) and so $f(x)$ is concave DOWN *near* x_0

When $f''(x)$ is either 0 or È, $f(x)$ can have either a curling point or an inflection point and the "unusual" points are thus the poles and the zeros of $f''(x)$, that is the critical points of $f'(x)$.

> **Example 1**. Let $f(x) = x^3 - 6x^2 + 9x$. If we want to look at $f(x)$ when x is near $x_0 = 0$, we just write $f(x) = 9x - 6x^2 + x^3 = 9x + \ldots$ which shows that $x_0 = 0$ is a zero of f and that sign f near x_0 is (–,+). If we want to look at $f(x)$ when x is near $x_0 = 3$, we set $x = 3 + h$ and expand which gives $f(3 + h) = (3 + h)^3 - 6(3 + h)^2 + 9(3 + h) = 3h^2 + \ldots$ which shows that 3 is a zero of order 2 , with sign (+,+) , and therefore a minimum.

3. Quantitative Analysis.

Thus far, by looking at the *principal* part of a function, we were able to obtain qualitative information. Here, we don't just ask if f is increasing or concave up near a point x_0 but how much so. We must therefore take into consideration the part that is *small compared to* the principal part and which we represent by the ellipsis ' ... '. Just recognizing the *existence* of this small part allows

---

[1] Here again, we wanted a term for the sake of terminological symmetry but we are not really satisfied with it as it usually implies that x_0 is critical; we use critical saddle in that case.

us, in contrast with the conventional approach, to fully define all the usual notions. For instance, in the conventional approach, we define

$$\lim_{x \to x_0} f(x) = L \qquad \text{iff} \qquad \forall \varepsilon\ \exists \delta \left[0 < |x - x_0| < \delta \ \Rightarrow\ |f(x) - L| < \varepsilon \right]$$

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

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

that is without even the appearance of a definition and with nothing to foster and support an intuition of the meaning of $\lim_{x \to x_0} f(x)$. On the other hand, if we separate in $f(x)$ the part which is finite from the part which is vanishingly small, then, when x approaches x_0, the first part gives the limit because the second part approaches 0. Specifically, we localize $f(x)$ and decompose $f(x_0 + h)$ as L, the constant part plus a remainder $R_0(h)$, that is $f(x_0 + h) = L + R_0(h)$, and then, when x approaches x_0, we get

$$\begin{aligned} \lim_{x \to x_0} f(x) &= \lim_{h \to 0} f(x_0 + h) \\ &= \lim_{h \to 0} [L + R_0(h)] \\ &= \lim_{h \to 0} L + \lim_{h \to 0} R_0(h) \\ &= L + \lim_{h \to 0} R_0(h) \end{aligned}$$

and to say that $\lim_{x \to x_0} f(x) = L$ is to say that $\lim_{h \to 0} R_0(h) = 0$. We thus have

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

which we interpret as saying that, when x is near x_0, $f(x_0 + h)$ is equal to L plus "something small" and this is an *operational* definition[1]. Moreover, note that, in order to prevent students from identifying the limit of a function at a point with the value of the function at that point, it suffices to require *sided* limits and that these are easily obtained by looking at the least non-constant approximation. In the same manner as above, we have

f is **continuous** at x_0 iff $\lim_{x \to x_0} f(x) = f(x_0)$

which, again, once we decide to avoid ε's and δ's, leaves us with

f is **continuous** at x_0 iff

and the notion of continuity becomes essentially a primitive one whose understanding depends solely on the students' intuition of the expression $\lim_{x \to x_0} f(x)$, if any, and on whatever connotations the word 'continuous' may have. On the other hand, localizing and decomposing $f(x_0 + h)$,

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

---

[1] We should keep in mind, though, that we are not dealing with a *series* that is with the limit of an *infinite* sum as this would involve taking the limit of $R_n(h)$ as n approaches ∞. For a treatment of calculus based on power series, see H. Levi, (1968).

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

so that, to say that $\lim_{x \to x_0} \boldsymbol{f}(x) = \boldsymbol{f}(x_0)$ is to say that $\lim_{h \to 0} \boldsymbol{R_0}(h) = 0$, and we have

f is **continuous** at x_0 iff $\boldsymbol{f}(x_0 + h) = \boldsymbol{f}(x_0) + \ldots$

which we can interpret as: $\boldsymbol{f}(x_0+h)$ is approximately equal to $\boldsymbol{f}(x_0)$: A CONTINUOUS FUNCTION IS A FUNCTION THAT IS LOCALLY APPROXIMATELY CONSTANT which again is an operational definition. We then have, similarly,

f is **differentiable** at x_0 iff $\boldsymbol{f}(x_0 + h) = \boldsymbol{f}(x_0) + lh + \ldots$ for some l

which we can interpret as: $\boldsymbol{f}(x_0+h)$ is approximately equal to $\boldsymbol{f}(x_0)$ plus a linear term: A DIFFERENTIABLE FUNCTION IS A FUNCTION THAT IS LOCALLY APPROXIMATELY AFFINE[1].

Example 2. Let $\boldsymbol{f}(x) = x^3 - 6x^2 + 9x$. To obtain the equation of the tangent to the graph of ***f*** near 2, localize, that is set $x = 2+h$ to obtain $\boldsymbol{f}(2+h) = \boldsymbol{f}_2(h) = (2 + h)^3 - 6(2 + h)^2 + 9(2 + h) = [8 + 12h + \ldots] - 6[4 + 4h + \ldots] = -7 + 3h \ldots$ so that the best affine approximation of $\boldsymbol{f}_2(h)$ is $\boldsymbol{a}_2(h) = +2 - 3h$. We get the *global* equation of the tangent by "delocalizing" $\boldsymbol{a}_2(h)$: $\boldsymbol{t}_2(x) = \boldsymbol{a}_2(x-2) = +2 - 3(x - 2) = -3x + 8$.

We call the coefficient of the linear term **linear rate of change** and we then define the **derivative** of a function ***f*** as the function ***f'*** whose value at x_0 is the linear rate of change of ***f*** at x_0. Thus, whenever we are able to decompose a function ***f*** at a point x_0 as $\boldsymbol{f}(x_0 + h) = \boldsymbol{f}_{x_0}(h) = A_0 + A_1 h + \ldots$, we obtain the value of the derivative of ***f*** at x_0 as the coefficient A_1. This makes it quite simple to obtain the derivative of a function "from first definition".

THEOREM. $(x^n)' = n \cdot x^{n-1}$

Proof. Localize $\boldsymbol{f}(x) = x^n$ at x_0. By the binomial expansion theorem,

$$\boldsymbol{f}_{x_0}(h) = (x_0 + h)^n = x_0{}^n + n \cdot x_0{}^{n-1} \cdot h + [n \cdot (n-1)/2] \cdot x^{n-2} \ldots,$$

so that the linear rate of change is $n \cdot x_0{}^{n-1}$ which gives $\boldsymbol{f}'(x) = n \cdot x^{n-1}$.

At first glance, though, it seems that there is something unsatisfactory about using the *linear* rate of change A_1 and that what we really want to use is the *instant* rate of change $\lim_{x \to x_0} \Delta x/\Delta y$. In the case of an affine function there is no difficulty since the average rate of change between any two points x_1 and x_2 is independent of x_1 and x_2 and equal to A_1. So, whatever the definition of $\lim_{x \to x_0}$,

[1] Observe that this is in fact the way differentiability is defined in higher dimensions. See, for instance, Williamson, Crowell and Trotter (1968).

the instant rate of change is equal to the linear rate. We then *define* the instant rate of change of a function as that of its best affine approximation! Incidentally, the direct proof that the instant rate of change as usually defined is the same as the linear rate of change is trivial:

$$\begin{aligned}
\lim_{x \to x_0} \Delta y/\Delta x &= \lim_{h \to 0} \{[\boldsymbol{f}_{x_0}(h) - \boldsymbol{f}_{x_0}(0)] \,/\, h\, \} \\
&= \lim_{h \to 0} \{[A_0 + A_1 h + A_2 h^2 + A_3 h^3 \ldots - A_0] \,/\, h\, \} \\
&= \lim_{h \to 0} \{[A_1 h + A_2 h^2 + A_3 h^3 \ldots] \,/\, h\, \} \\
&= \lim_{h \to 0} \{A_1 + A_2 h + A_3 h^2 \ldots \} \\
&= A_1
\end{aligned}$$

The usual rules are also proven quite easily. For instance, the chain rule goes as follows. Assume that $\boldsymbol{f}$ is differentiable at x_0:

$$\boldsymbol{f}(x_0 + h) = \boldsymbol{f}(x_0) + \boldsymbol{f}'(x_0)h + \ldots$$

and that $\boldsymbol{g}$ is differentiable at $\boldsymbol{f}(x_0)$:

$$\boldsymbol{g}(\boldsymbol{f}(x_0) + k) = \boldsymbol{g}(\boldsymbol{f}(x_0)) + \boldsymbol{g}'(\boldsymbol{f}(x_0))k + \ldots$$

Then, to show that $\boldsymbol{g}\circ\boldsymbol{f}$ is differentiable at x_0, we must evaluate $[\boldsymbol{g}\circ\boldsymbol{f}](x_0 + h)$.

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

where $k = \boldsymbol{f}'(x_0)h + \ldots$ so that

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

which shows that $[\boldsymbol{g}\circ\boldsymbol{f}]'(x_0) = \boldsymbol{g}'(\boldsymbol{f}(x_0))\boldsymbol{f}'(x_0)$.

Higher derivatives can be defined *inductively* as usual but also *directly*. For example, the second derivative of $\boldsymbol{f}(x) = x^n$ is the derivative of $\boldsymbol{f}'(x) = n \cdot x^{n-1}$. Localize at x_0. By the binomial expansion theorem,

$$\boldsymbol{f}'(x_0 + h) = n \cdot (x_0 + h)^{n-1} = n \cdot [x_0^{n-1} + (n-1) \cdot x_0^{n-2} \cdot h + \ldots\ ] = n \cdot x_0^{n-1} + n \cdot (n-1) \cdot x_0^{n-2} \cdot h + \ldots\ ]$$

so that the linear rate of change of $\boldsymbol{f}'$ is $n \cdot (n-1) \cdot x_0^{n-2}$ which gives $\boldsymbol{f}''(x) = n \cdot (n-1) \cdot x^{n-2}$.

Observe, though, that this is twice the quadratic coefficient in $f(x_0 + h)$ so that we can *define* the second derivative, as well as all the others, directly from the polynomial approximation. That is, we can write

$$f(x_0 + h) = f(x_0) + f'(x_0)h + f''(x_0)h^2/2 + f^{(3)}(x_0)h^2/2 + \dots + f^{(n)}(x_0)h^n/n! + \dots$$

4. Remainder Theorems.

Various theorems take a simple, natural meaning which is to provide information about the remainder. When, for instance, f is differentiable, the Mean Value Theorem gives us information on how fast the remainder approaches 0 by saying that the remainder $R_0(h)$ in $f(x_0 + h) = f(x_0) + R_0(h)$ is of the form $h \cdot f'(c)$ with c between x_0 and x and this is indeed the form in which the Mean Value Theorem is actually used:

$$f(x_0 + h) = f(x_0) + h \cdot f'(c) \text{ with } c \text{ between } x_0 \text{ and } x.$$

Actually, the Mean Value Theorem is a special case of Taylor's formula with remainder which, significantly, is actually due to Lagrange and which is also often called the Extended Mean Value Theorem: when f is $\mathbf{C}^{n+1}$, the remainder $R_n(h)$ in

$$f(x_0 + h) = f(x_0) + h \cdot f'(x_0) + \dots + h^n \cdot f^{(n)}(x_0)/n! + h^{n+1} \cdot R_n(h)$$

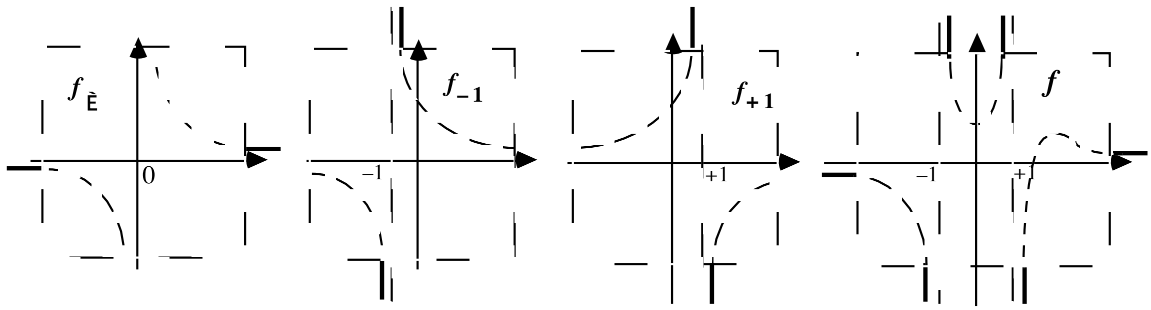
is of the form $R_n(h) = f^{(n+1)}(c)/(n+1)!$ with c between x_0 and x:

5. Applications.

The main applications of the differential calculus are optimization and graphing. Extremes are found by analyzing critical points but here we can do it in several ways: as with any point, we can expand the function f itself or we can expand the *derivative* of f *near* x_0 and then recover from its sign *near* x_0 the information about the variance of f *near* x_0. We can also look at the *second derivative* whose sign at x_0 gives the concavity *near* x_0 and, even if f'' is 0 at x_0, we can expand it to get its sign *near* x_0 and therefore the concavity of f *near* x_0. We give an example of graphing.

Example 3. Consider the function $f(x) = (x-2) / (x^2-1)$. To graph f, we approximate f near its essential points, ∞ and the poles -1 and $+1$.
$f_\infty(x) = 1/x + \dots$,
$f(-1 + h) = f_{-1}(h) = 3/2h + \dots$, so that, from a global viewpoint, near -1, $f(x) \approx 3/2(x+1)$
$f(+1 + h) = f_{+1}(h) = -1/2h + \dots$, so that, from a global viewpoint, near $+1$, $f(x) \approx -1/2(x-1)$

We then sketch the local graphs and interpolate smoothly:



$f_{È}(x) \approx 1/x$ $\qquad$ $f_{-1}(h) \approx 3/2h$ $\qquad$ $f_{+1}(h) = -1/2h$ $\qquad$ Essential graph

Thus $\boldsymbol{f}$ must have a minimum somewhere between –1 and +1, a zero somewhere right of +1, a maximum somewhere right of the zero and an inflection somewhere right of the maximum.

6. Fundamental Theorem.

We consider the following initial value problem in terms of finite differences: given a function $\boldsymbol{f}(x)$, find the value at x_1 of a function $\boldsymbol{F}(x)$ such that $\boldsymbol{F}'(x) = \boldsymbol{f}(x)$ given $\boldsymbol{F}(x_0)$. If we assume the existence of an antiderivative $\boldsymbol{F}(x)$, we have immediately from our definition of differentiability[1]:

$$\begin{aligned} \boldsymbol{F}(x_0 + h) - \boldsymbol{F}(x_0) &= \boldsymbol{F}'(x_0)h + h\mathbf{o_1}[1] \\ &= \boldsymbol{f}(x_0)h + h\mathbf{o_1}[1] \end{aligned}$$

Then, taking $h = \dfrac{x_1 - x_0}{n}$, we continue step by step until we reach $x_1 = x_0 + nh$:

$$\begin{aligned} \boldsymbol{F}(x_0 + 2h) - \boldsymbol{F}(x_0 + h) &= h\boldsymbol{f}(x_0 + h) + h\mathbf{o_2}[1] \\ \boldsymbol{F}(x_0 + 3h) - \boldsymbol{F}(x_0 + 2h) &= h\boldsymbol{f}(x_0 + 2h) + h\mathbf{o_3}[1] \end{aligned}$$

....

$$\boldsymbol{F}(x_0 + nh) - \boldsymbol{F}(x_0 + (n-1)h)) = h\boldsymbol{f}(x_0 + (n-1)h) + h\mathbf{o}_{\boldsymbol{n}}[1]$$

Adding and cancelling on the left, we get:

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

This is always true but the term $h\sum_{i=1}^{i=n}\mathbf{o}_{\boldsymbol{i}}[1]$ is extremely complicated to evaluate. *So,* we have a good reason to let n approach È. Clearly, for $\boldsymbol{f}(x)$ smooth enough, $h\sum_{i=1}^{i=n}\mathbf{o}_{\boldsymbol{i}}[1]$ approaches 0 as n approaches È and thus we obtain

$$\boldsymbol{F}(x)\Big|_{x_0}^{x_1} = \lim_{n.È}\sum_{i=0}^{i=n-1}\boldsymbol{f}(x_0 + ih) = \lim_{n.È}\sum_{i=0}^{i=n-1}\boldsymbol{f}(x_i) ,$$

---

[1] In comparing small quantities, it is convenient to introduce Landau's "little oh" notation. Given two functions $\boldsymbol{f}$ and $\boldsymbol{g}$ with $\boldsymbol{g}(0) \neq 0$, if $\lim_{hâ0} \boldsymbol{f}(h)/\boldsymbol{g}(h) = 0$, that is if, as h approaches 0, $\boldsymbol{f}(h)$ approaches 0 *faster* than $\boldsymbol{g}(h)$, we shall say that $\boldsymbol{f}(h) = \mathbf{o}[\boldsymbol{g}(h)]$ near 0.

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

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

7. Approximation of Non-Polynomial Functions.

To conclude, we should observe that nothing in the preceding depended on f being a polynomial function other than we approximated by truncation. The only question that remains is to indicate how to get the polynomial approximation of "all" functions". In the case of polynomial functions, we obtained these by truncating binomial expansions. In the case of rational functions, we obtain them by division of polynomials in both ascending order near 0 and descending powers near È. For "all" other functions, we obtain the approximation by the method of undetermined coefficients from the functional equation, algebraic or differential, of which they are the solution[1]. See F. Schremmer and A. Schremmer (1989 b) for a more detailed exposition.

It is interesting to observe that the polynomial approximations already have many of the properties of the exact solution.

Example 4. To get $e^a e^b = e^{a+b}$, multiply $1 + a + a^2/2! + a^3/3! + ...$ by $1 + b + b^2/2! + b^3/3! + ...$; this gives

$$\begin{aligned} e^a Ù e^b = 1 &+ a + a^2/2! + a^3/3! + ... \\ &+ b + ab + a^2b/2! + ... \\ &+ b^2/2! + ab^2/2! + ... \\ &+ b^3/3! + ... \end{aligned}$$

And since ' ... ' stands for finite remainders and *not* for infinite tails, we can commute additions to get:

$$\begin{aligned} e^a Ù e^b &= 1 + a + b + a^2/2! + ab + b^2/2! + a^3/3! + a^3 + a^2b/2! + ab^2/2! + b^3/3! + ... \\ &= 1 + (a + b) + (a^2 + 2ab + b^2)/2! + (a^3 + 3a^2b + 3ab^2 + b^3)/3! + ... \\ &= 1 + (a + b) + (a + b)^2/2! + (a + b)^3/3! + ... = e^{a+b} \end{aligned}$$

Finally, note that we have no need for L'Hôpital's rule.

Example 5. Consider the following *complete* list of examples of applications of L'Hôpital's rule taken from a popular textbook.

1. $\lim_{x å 0} \sin x / x$ **2.** $\lim_{x å \pi/2} [1 - \sin x]/\cos x$ **3.** $\lim_{x å 0} [e^x - 1]/x^3$

---

[1] Even the rigorous treatment is much simpler that way than the conventional one. See for instance Sections 4-1, 2, and 3 in S. Lang (1976) or Section 4-8 and exercise 3 in R. L. Finney and D. R. Ostbey (1984) .

4. $\lim_{x\to 0} [1 - \cos x]/x^2$ **5.** $\lim_{x\to 0} e^x/x^2$ **6.** $\lim_{x\to +\infty} x^{-4/3}/\sin(1/x)$
7. $\lim_{x\to 0} \tan x/x^2$

Now, with the exception of **2.** and **6.**, the limits are obvious as soon as we replace the functions by their polynomial approximations as found by the method of indeterminate coefficients. But, of course, this is exactly what L'Hôpital's rule does. For **2.** and **6.** it would seem that we have a problem since the polynomial approximations for trigonometric functions are only good near 0. However, for **2.** we naturally localize, that is we set $x = \pi/2 + u$ and for **6.** we naturally invert, that is we set $v = 1/x$ after which we can use the polynomial approximations.

References.

Finney, R. L. - Ostbey, D. R. (1984). *Elementary differential equations with linear algebra*. Reading: Addison Wesley.

Lang, S. (1976). *Analysis I*. Reading: Addison-Wesley.

Levi, H. (1968). *Polynomials, Power series and calculus*. Princeton: Van Nostrand.

Mattei, F. & Schremmer, A. (1988 a). *Calculus I, a Lagrangian approach*. (4th ed.) Christiana PA: MetaMath.

Mattei, F. & Schremmer, A. (1988 b). *Differential calculus , a Lagrangian approach (Vol. 1)*. Christiana PA: MetaMath.

Mattei, F. & Schremmer, A. (1988 c). *Differential calculus , a Lagrangian approach (Vol. 2)*. Christiana PA: MetaMath.

Schremmer, F., & Schremmer, A. (in press). The Differential Calculus as Language. *Bulletin of Science, Technology and Society*.

Schremmer, F. – Schremmer, A. (1989 a). *A Lagrangian, integrated precalculus–differential calculus*.Manuscript submitted for publication.

Schremmer, F., & Schremmer, A. (1989 b). *Calculus as approximation theory: The Lagrangian view*. Manuscript submitted for publication.

Williamson, R. E., Crowell, R. H., & Trotter, H. B. (1968). *Calculus of vector functions*. Englewood Cliffs: Prentice-Hall.