An Introduction to Lagrangian Differential Calculus

by

Francesca Schremmer, *West Chester University* West Chester, PA 19383

Alain Schremmer, *Community College of Philadelphia* Philadelphia, PA 19130

Supported in part by NSF Grant USE-8814000

WCU *Francesca Schremmer has a DES from the Université de Paris and a PhD from the University of Pennsylvania with a thesis on the Dirichlet problem for parabolic equations. Her research interest remains in differential equations, particularly solitons and strange attractors.*

CCP *Alain Schremmer has a DES from the Université de Paris and an ABD in model theory from the University of Pennsylvania. He likes to develop new courses but hopes, one day, to return to singular perturbation theory.*

Introduction.

In the conventional approach to teaching first year calculus, limits are considered, whether explicitly or implicitly, to be absolutely fundamental. But this approach is by no means the only possible one, let alone the most desirable one.

In (Awards, 1988), for instance, mention is made of a 1987 MAA Award for Expository Excellence, the George Polya Award, to Irl. C. Bivens for "What a Tangent Line is When it isn't a Limit", (Bivins, 1986). 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).

In fact, the differential study of functions through their best *polynomial* approximations is an old idea, going back to (Lagrange, 1797), and it is still basic to Perturbation Theory. It has also acquired a new life as Theory of Jets in the local theory of differentiable maps.

We will illustrate here how Lagrange's approach applies to the differential calculus of polynomial functions as, in this case, the approximations are readily obtained. We then shall briefly indicate how to obtain the polynomial approximations in "all" other cases.

Best Polynomial Approximations.

Consider the polynomial function

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n$$

and suppose that we are interested in $f(x)$ when x is *near* x_0. In other words, we want to **expand** $f(x)$ near x_0. To **localize** f at x_0, that is to get a form where the terms are in descending order of magnitude, we express $f(x)$ in terms of $h = x - x_0$:

$$f(x_0 + h) = a_0 + a_1(x_0 + h) + a_2(x_0 + h)^2 + a_3(x_0 + h)^3 + \ldots + a_n(x_0 + h)^n$$

Expanding the binomials and rearranging gives

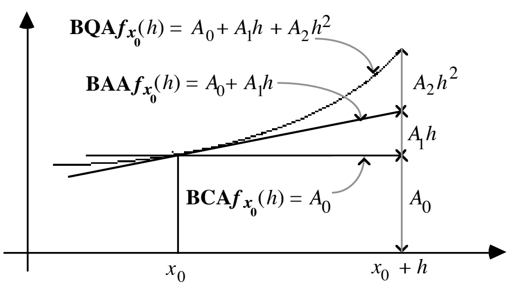
$$f(x_0 + h) = A_0 + A_1h + A_2h^2 + A_3h^3 + \ldots + A_nh^n.$$

where the A_i depend only on f and x_0. In this case then, $f(x_0 + h)$, the value of $f(x)$ *near* x_0, is given exactly by $f_{x_0}(h)$, the **localization** of f *at* x_0 where $f_{x_0}(h) = A_0 + A_1h + A_2h^2 + \ldots + A_nx^n$. Since h is now near 0 and the powers of h in descending orders of magnitude, we can **approximate** the localization to the **principal part** simply by truncating it to the appropriate degree. To approximate at infinity, we truncate $f_{\text{È}}(x)$, the localization at infinity, which is just $f(x)$ in descending powers.

We can approximate $f(x_0 + h)$ with **constant** functions, the simplest non-zero functions. If we approximate $f(x_0 + h)$ by a constant function $\mathbf{CA}f(x_0 + h) = k$, the "error" is $f(x_0 + h) - \mathbf{CA}f(x_0 + h) = [A_0 - k] + \ldots$ and the order of magnitude of the error is the same as that of the approximation. But if we take $k = A_0$, then the "error" will be A_1h, that is smaller than the approximation by an order of magnitude. Thus, the **Best Constant Approximation** of $f(x)$ *near* x_0 is $\mathbf{BCA}f(x_0 + h) = A_0$. We shall just write $f(x_0 + h) = A_0 + \ldots$. At this point, the ellipsis just indicates that A_0 is a best approximation.

We have, similarly, the **Best Affine Approximation** $\mathbf{BAA}f(x_0 + h) = A_0 + A_1h + \ldots$ and the **Best Quadratic Approximation** $\mathbf{BQA}f(x_0 + h) = A_0 + A_1h + A_2h^2 + \ldots$.

The successive approximations of $f(x_0 + h)$ are easily visualized. To obtain a **local graph** of $f_{x_0}(h)$, first graph the constant function A_0, then, using it as "base line", graph the linear function A_1h, then, using the graph of the affine function $A_0 + A_1h$ as base line, graph the parabolic function A_2h^2, then using the graph of the quadratic function $A_0 + A_1h + A_2h^2$ as base line graph the cubic function A_3h^3, etc:



Qualitative Analysis.

Much of what we do in differential calculus consists in extending information *at* a point x_0 into information *near* x_0. It is thus natural that we should expand the function near x_0. We first give *elementary* definitions of the usual features of a function and then necessary and sufficient conditions in terms of the polynomial approximations. Since constant and affine functions are pathological (x^0 and x^1 are the only "straight" power functions) and therefore usually not very good approximations, the general idea is to characterize functions by the way they *differ* from these approximations.

The zero function has no sign. So, we define $\text{sign}_{x_0} f$, the **sign** of f *near* x_0, as the way $f(x)$ differs on each side of x_0 from the zero function. For example, $\text{sign}_0\, x^3 = (-,+)$. Then, to find how $f(x)$ differs from the zero function, we need only to approximate $f(x_0 + h)$ to its **Least Non-Zero Approximation**. Indeed, because the terms of $f_{x_0}(h)$ are in descending order of magnitude, none of the further terms in $f_{x_0}(h)$ can affect the sign as given by the first non-zero term. A **zero** is a point where $A_0 = 0$ and its multiplicity corresponds to the first non-zero A_i. Since, however, the Least Non-Zero Approximation is "usually" the Best Constant Approximation, $f(x_0 + h) = A_0 + ...$, most points are "even signed". Since $A_0 = f(x_0)$, we get that $f(x_0 + h) = f(x_0) + ...$ and the

SIGN THEOREM. If $f(x)$ is neither 0 nor È *at* x_0, $f(x)$ is even signed *near* x_0. Moreover,

if $f(x)$ is positive *at* x_0, then $\text{sign}_x f = (+,+)$ and so $f(x)$ is positive *near* x_0,
if $f(x)$ is negative *at* x_0, then $\text{sign}_{x_0} f = (–,–)$ and so $f(x)$ is negative *near* x_0.
If $f(x)$ is either 0 or È at x_0, $f(x)$ near x_0 can be even or odd signed. The "unusual" points for the sign are thus the poles and the zeros of $f(x)$—whether finite or infinite.

Similarly, a constant function does not vary and we define $\text{var}_{x_0} f$, the **variance** of f *near* x_0, as the way $f(x)$ differs on each side of x_0 from its best constant approximation. For example, $\text{var}_0 –x^2$ = (Ω,œ). Then, to find $\text{var}_{x_0} f$, we need only to approximate $f(x_0 + h)$ to its **Least Non-Constant Approximation**. Depending on the parity of n and on the sign of A_n, $f(x_0 + h) = A_0 + A_n h^n + ...$ 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). A **critical point** is defined as being either a point where $A_1 = 0$ or a pole. Since the Least Non-Constant Approximation of $f(x)$ is "usually" its Best Affine Approximation, $f(x_0 + h) = f(x_0) + A_1 h + ...$, we have that most points are monotonic. Since, anticipating a bit, $A_1 = f'(x_0)$, we get that $f(x_0 + h) = f(x_0) + f'(x_0)h + ...$ and the

VARIANCE THEOREM. If $f'(x)$ is neither 0 nor È *at* x_0, $f(x)$ is monotonic *near* x_0. Moreover,
if $f'(x)$ is positive *at* x_0, then $\text{var}_{x_0} f$ = (Ω, Ω) and so $f(x)$ is increasing *near* x_0,
if $f'(x)$ is negative *at* x_0, then $\text{var}_{x_0} f$ = (œ,œ) and so $f(x)$ is decreasing *near* x_0
If $f'(x)$ is either 0 or È at x_0, $f(x)$ near x_0 can be either monotonic or turning. The "unusual" points for the variance are thus the critical points of $f(x)$.

While optimization is closely related to variance, we prefer to define extremes independantly. So, we define $\text{opt}_{x_0} f$, the **optimization** of f *near* x_0, as the way $f(x)$ compares with the constant function $f(x_0)$ on each side of x_0. For instance, we have that both $\text{opt}_0\ x^1$ and $\text{opt}_0\ x^3$ are equal to (max,min). This notion is also useful at boundary points. To find $\text{opt}_{x_0} f$, we need only to approximate $f(x_0 + h)$ to its Least Non-Constant Approximation. Depending on the parity of n and on the sign of A_n, $f(x_0 + h) = f(x_0) + A_n h^n + ...$ shows that x_0 is a **saddle point** with optimization (min,max) or (max,min) or an **extreme point** with optimization (min,min) or (max,max). Since the Least Non-Constant Approximation is "usually" the Best Affine Approximation, most points are saddle points and this explains why, in the case of differentiable functions, the search for extremes first has to go through a search for critical points:

OPTIMIZATION THEOREM. If $f'(x)$ is neither 0 nor È *at* x_0, $f(x)$ has a saddle *near* x_0. Moreover,
if $f'(x)$ is positive *at* x_0, then $\text{opt}_{x_0} f(x)$ = (max,min),
if $f'(x)$ is negative *at* x_0, then $\text{opt}_{x_0} f(x)$ = (min,max).

If $\boldsymbol{f}'(x)$ is either 0 or È at x_0, $\boldsymbol{f}(x)$ near x_0 can have either a saddle or an extreme. The "unusual" points for optimization are thus the critical points of $\boldsymbol{f}(x)$.

Finally, we define $\text{conc}_{x_0}\, \boldsymbol{f}$, the **concavity** of $\boldsymbol{f}$ *near* x_0, as the way $\boldsymbol{f}(x)$ differs from its Best Affine Approximation on each side of x_0. Then, to find $\text{conc}_{x_0}\, \boldsymbol{f}$, we need only to approximate $\boldsymbol{f}(x_0 + h)$ to its **Least Non-Affine Approximation** of $\boldsymbol{f}(x)$. Depending on the parity of n and on the sign of A_n., $\boldsymbol{f}(x_0 + h) = A_0 + A_1h + \ldots + A_nh^n + \ldots$ shows that x_0 is a **curling** point with concavity ($%,$%) or (^&,^&) or an **inflection** point with concavity ($%,^&) or (^&,$%). Since the Least Non-Affine Approximation is "usually" the Best Quadratic Approximation, $\boldsymbol{f}(x_0 + h) = \boldsymbol{f}(x_0) + \boldsymbol{f}'(x_0)h + A_2h^2 + \ldots$, we have that most points are curling. Since, anticipating a bit further, $A_2 = \dfrac{\boldsymbol{f}''(x_0)}{2}$, we get that $\boldsymbol{f}(x_0 + h) = \boldsymbol{f}(x_0) + \boldsymbol{f}'(x_0)h + \dfrac{\boldsymbol{f}''(x_0)}{2}$ $h^2 + \ldots$ and the

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

if $\boldsymbol{f}''(x)$ is positive *at* x_0, then $\text{conc}_{x_0}\, \boldsymbol{f}(x)$ = ($%,$%) and so $\boldsymbol{f}(x)$ is concave up *near* x_0,

if $\boldsymbol{f}''(x)$ is negative *at* x_0, then $\text{conc}_{x_0}\, \boldsymbol{f}(x)$ = (^&,^&) and so $\boldsymbol{f}(x)$ is concave down *near* x_0.

If $\boldsymbol{f}''(x)$ is either 0 or È, $\boldsymbol{f}(x)$ can have either a curling point or an inflection point. The "unusual" points are thus the "second critical" points of $\boldsymbol{f}(x)$ that is the critical points of $\boldsymbol{f}'(x)$.

Example 1. Let $\boldsymbol{f}(x) = x^3 - 6x^2 + 9x$. Lagrange's approach allows us to investigate the behaviour of $\boldsymbol{f}$ near *any* given point.

To look at $\boldsymbol{f}(x)$ near 0, localize at 0, that is just rearrange in ascending powers: $\boldsymbol{f}_0(x) = 9x - 6x^2 + x^3$. Then, $\boldsymbol{f}_0(x) = 9x + \ldots$ shows that 0 is a zero with $\text{sign}_0\, \boldsymbol{f} = (-,+)$. It also shows that 0 is a monotonic point with $\text{var}_0\, \boldsymbol{f}$ = (Ω,Ω) and a saddle point with $\text{opt}_0\, \boldsymbol{f}$ = (max,min). Finally, $\boldsymbol{f}_0(x) = 9x - 6x^2 + \ldots$ shows that 0 is a curling point with $\text{conc}_0\, \boldsymbol{f}$ = (^&,^&).

To look at $\boldsymbol{f}(x)$ near 1, localize at 1: $\boldsymbol{f}(1 + h) = (1 + h)^3 - 6(1 + h)^2 + 9(1 + h) = 4 - 3h^2 + h^3$. Then, $\boldsymbol{f}_1(h) = 4 + \ldots$ shows that $\boldsymbol{f}(x)$ is positive near 1 and $\boldsymbol{f}_1(h) = 4 - 3h^2 + \ldots$ shows that 1 is critical point, a turning point with $\text{var}_1\, \boldsymbol{f}$ = (Ω,œ), an extreme point with $\text{opt}_1\, \boldsymbol{f}$ = (max,max) and a curling point with $\text{conc}_1\, \boldsymbol{f}$ = (^&,^&).

To look at $\boldsymbol{f}(x)$ near 3, localize at 3: $\boldsymbol{f}(3 + h) = (3 + h)^3 - 6(3 + h)^2 + 9(3 + h) = 3h^2 + h^3$. Then, $\boldsymbol{f}_3(h) = 3h^2 + \ldots$ shows that 3 is a zero with $\text{sign}_3\, \boldsymbol{f} = (+,+)$, a turning point with $\text{var}_3\, \boldsymbol{f}$ = (œ,Ω), an extreme point with $\text{opt}_3\, \boldsymbol{f}$ = (min,min) and a curling point with $\text{conc}_3\, \boldsymbol{f}$ = ($%,$%).

To look at $\boldsymbol{f}(x)$ near È, localize at È: $\boldsymbol{f}_{\text{È}}(x) = +x^3 - 6x^2 + 9x$ (it is the "default" localization). The local graph of $\boldsymbol{f}_{\text{È}}(x) = x^3 + \ldots$ near È shows that È is a pole with $\text{sign}_{\text{È}}\, \boldsymbol{f} = (+,-)$, a monotonic point with $\text{var}_{\text{È}}\, \boldsymbol{f}$ = (Ω, Ω), a saddle point with $\text{opt}_{\text{È}}\, \boldsymbol{f}$ = (max,min) and an inflection point with $\text{conc}_{\text{È}}\, \boldsymbol{f}$ = ($%,^&). (Note that +È is the *left* side of infinity and that –È is the *right* side of infinity.)

Quantitative Analysis.

Just by looking at the principal part of a function, we were able in the previous section to obtain qualitative information. Here, we don't just ask if $\boldsymbol{f}$ is increasing or concave up near a point x_0 but how much so. We must therefore take into consideration the **remainder**, that is the part that is *small compared to* the principal part and which we represented by the ellipsis ' ... '. Just recognizing the *existence* of this small part allows us, in contrast with the conventional approach, to *fully* define all the usual notions. For instance, in the conventional approach, we define

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

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

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

that is without even the appearance of a definition and with nothing to foster and support any intuition of the meaning of $\lim_{x\emptyset x_0} f(x)$. Since the "intuitive" idea of closeness provides the students with no working definition, in particular with no way to *find* limits, it seems hardly worth the usual effort.

By contrast, in Lagrange's approach, we have

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

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* a working definition as it allows us to *find* limits. In fact, we can find *sided* limits just as easily by looking at the graph of the Least Non-Constant Approximation $f(x_0 + h) = L + A_n h^n + \ldots$ instead of just that of the Best Constant Approximation $f(x_0 + h) = L + \ldots$

Continuity is then defined as

$$\boxed{f \text{ is } \textbf{continuous} \text{ at } x_0 \qquad \text{iff} \qquad f(x_0 + h) = f(x_0) + \ldots}$$

which we read as: f is continuous at x_0 iff f near x_0 is approximately equal to f at x_0. So, A CONTINUOUS FUNCTION IS A FUNCTION THAT, LOCALLY, IS APPROXIMATELY CONSTANT which again is a working definition. Indeed, f has a noticeable jump at x_0 if and only if $f(x_0 + h)$ is noticeably different from $f(x_0)$ on at least one side of x_0.

Similarly, and exactly in the way differentiability is defined in higher dimensions, see for instance (Williamson, Crowell, & Trotter, 1968), we have

$$\boxed{f \text{ is } \textbf{differentiable} \text{ at } x_0 \qquad \text{iff} \qquad f(x_0 + h) = f(x_0) + lh + \ldots \text{ for some } l}$$

which we read as: f is differentiable at x_0 iff f near x_0 is approximately equal to f at x_0 plus a term of the order of $|x - x_0|$. So, A DIFFERENTIABLE FUNCTION IS A FUNCTION THAT, LOCALLY, IS APPROXIMATELY AFFINE.

Example 2. Let $f(x) = x^3 - 6x^2 + 9x$. To obtain the equation of the tangent to the graph of f near 2, localize to obtain $f(2+h) = f_2(h) = (2 + h)^3 - 6(2 + h)^2 + 9(2 + h) =$

[8 + 12h + ...] – 6[4 + 4h + ...]+ 9(2 + h) = 2 – 3h + ... so that the best affine approximation of $\boldsymbol{f}_2(h)$ is $\mathbf{BAA}\boldsymbol{f}_2(h) = 2 - 3h$. We get the *global* equation of the tangent by "delocalizing" $\mathbf{BAA}\boldsymbol{f}_2(h)$: $\boldsymbol{t}_2(x) = \mathbf{BAA}\boldsymbol{f}_2(x-2) = 2 - 3(x-2) = -3x + 8$.

We call **linear rate of change** the coefficient A_1 of the linear term. We then *define* the **derivative** of a function $\boldsymbol{f}$ as the function $\boldsymbol{f}'$ whose value at x_0 is the linear rate of change of $\boldsymbol{f}$ at x_0: $\boldsymbol{f}'(x_0) = A_1$. This makes it quite simple to obtain the derivative of a function "from first definition".

Example 3. To obtain the derivative of x^n, localize $\boldsymbol{f}(x) = x^n$ at x_0. By the binomial expansion theorem,

$$\boldsymbol{f}(x_0 + h) = (x_0 + h)^n = {x_0}^n + n \cdot {x_0}^{n-1} \cdot h + \dots ,$$

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 a loss of intuition in using the *linear* rate of change A_1 and that, from a physical point of view, what is really intuitive is the *instant* rate of change $\lim_{x \to x_0} \Delta y / \Delta x$. In the case of an affine function however 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 is equal to A_1. So, in this case, by any definition of $\lim_{x \to x_0}$, the instant rate of change is equal to the linear rate. But then, in the general case, we just say that the instant rate of change of a function must be the same as that of its best affine approximation which is A_1:

$$\begin{aligned}
\lim_{x \to x_0} \frac{\Delta x}{\Delta y} &= \lim_{h \to 0} \frac{f_{x_0}(h) - f_{x_0}(0)}{h} \\
&= \lim_{h \to 0} \frac{A_0 + A_1 h + A_2 h^2 + A_3 h^3 + \dots - A_0}{h} \\
&= \lim_{h \to 0} \frac{A_1 h + A_2 h^2 + A_3 h^3 + \dots}{h} \\
&= \lim_{h \to 0} \{A_1 + A_2 h + A_3 h^2 + \dots\} \\
&= A_1
\end{aligned}$$

The usual rules are also quite easily proven by looking at the linear rate of change in the expansion of the appropriate function.

Example 4. To get the quotient rule, expand $\left[\dfrac{\boldsymbol{f}}{\boldsymbol{g}}\right](x_0 + h)$ by dividing in *ascending* powers:

$$\left[\frac{\boldsymbol{f}}{\boldsymbol{g}}\right](x_0 + h) = \frac{\boldsymbol{f}(x_0 + h)}{\boldsymbol{g}(x_0 + h)} = \frac{\boldsymbol{f}(x_0) + \boldsymbol{f}'(x_0)h + \dots}{\boldsymbol{g}(x_0) + \boldsymbol{g}'(x_0)h + \dots} = \frac{\boldsymbol{f}(x_0)}{\boldsymbol{g}(x_0)} + \frac{\boldsymbol{f}'(x_0)\boldsymbol{g}(x_0) - \boldsymbol{f}(x_0)\boldsymbol{g}'(x_0)}{[\boldsymbol{g}(x_0)]^2}\, h + \dots$$

Example 5. To get the chain rule, expand $[\boldsymbol{g}\circ\boldsymbol{f}](x_0 + h) = \boldsymbol{g}(\boldsymbol{f}(x_0 + h))$.
By the differentiability of $\boldsymbol{f}$ at x_0:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{f}(x_0 + h)) &= \boldsymbol{g}(\boldsymbol{f}(x_0) + \boldsymbol{f}'(x_0)h + \dots) \\ &= \boldsymbol{g}(\boldsymbol{f}(x_0) + k) \qquad \text{where } k = \boldsymbol{f}'(x_0)h + \dots \end{aligned}$$

and by the differentiability of $\boldsymbol{g}$ at $\boldsymbol{f}(x_0)$

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

The n^{th} derivative can be defined *inductively* as usual but also *directly* from the coefficient of h^n in $\boldsymbol{f}(x_0 + h)$.

Example 6. Inductively, the second derivative of $\boldsymbol{f}(x) = x^n$ is the derivative of $\boldsymbol{f}'(x) = n\text{Ù}x^{n-1}$. Localize at x_0. By the binomial expansion theorem,

$$\begin{aligned} \boldsymbol{f}'(x_0 + h) &= n\text{Ù}(x_0 + h)^{n-1} = n\text{Ù}\left[x_0^{\,n-1} + (n-1)\text{Ù}x_0^{\,n-2}\text{Ù}h + \dots\right] \\ &= n\text{Ù}x_0^{\,n-1} + n\text{Ù}(n-1)\text{Ù}x_0^{\,n-2}\text{Ù}h + \dots \end{aligned}$$

and the linear rate of change of $\boldsymbol{f}'$ is $n\text{Ù}(n-1)\text{Ù}x_0^{\,n-2}$. Note that this is twice the coefficient of h^2 in $\boldsymbol{f}(x_0 + h)$ which is the **quadratic rate of change** of $\boldsymbol{f}$.

Altogether then, polynomial approximations are just Taylor expansions:

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

Note that even in the general case, we are not dealing with a *series*, that is with the limit of an *infinite* sum, as this would involve taking the limit of the remainder $\boldsymbol{R}_n(h)$ as n approaches È. For a treatment of calculus based on power series, see (Levi, 1968).

But since "all" functions encountered in first year calculus are C^n, this implies that they should, a priori, be approximable by polynomial functions. And, indeed, nothing in the preceding depended on $\boldsymbol{f}$ being a polynomial function other than the way in which we obtained the approximations.

Approximation of NonPolynomial Functions.

We now indicate how to get the polynomial approximation of "all" other functions. In the case of rational functions, we obtain the approximations by division of polynomials in ascending powers near 0 and descending powers near È. Observe that we can also approximate rational functions near their poles; the only difference is that the approximation will be a Laurent-polynomial. See (Schremmer, & Schremmer, In preparation a) and Example 8 below. For "all" other functions, we obtain the approximations by the method of undetermined coefficients from the functional equation, algebraic or differential, of which they are the solution. See (Schremmer, & Schremmer, In preparation b). In fact, even the rigorous treatment is much simpler that way than the conventional one. See for instance Sections 4-1, 2, and 3 in (Lang, 1976) or Section 4-8 and exercise 3 in (Finney, & Ostbey, 1984).

It is interesting however to check that the polynomial approximations already have many of the properties of the exact solution. Also, note that there is no need for L'Hôpital's rule.

Example 7. Consider the following *complete* list of examples of applications of L'Hôpital's rule taken from a popular textbook.

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

With the exception of **2.** and **6.**, the limits are obvious as soon as we replace the functions by their polynomial approximations. For **2.** we first localize at $\pi/2$ and for **6.** we first set $h = 1/x$.

Applications.

Among the main *mathematical* applications of the differential calculus are optimization and graphing. Extremes are found by analyzing critical points but here we can do this in several ways. As at any point, we can expand the function f itself. Or we can expand the *derivative* of f *near* x_0 and recover from its sign *near* x_0 the information about the variance of f *near* x_0. Or we can look at the *second derivative* whose sign *at* x_0 gives the concavity *near* x_0 and, if f'' is 0 at x_0, we can expand f'' to get its sign *near* x_0 and therefore the concavity of f *near* x_0.

To graph a rational function, we just approximate f near its **essential** points, that is near È and the poles.

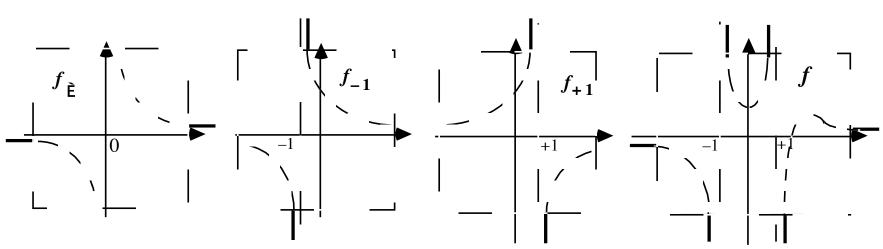
Example 8. Consider the function $f(x) = \dfrac{x-2}{x^2-1}$ whose poles are -1 and $+1$.

$f_È(x) = \dfrac{x + \dots}{x^2 + \dots} = \dfrac{1}{x} + \dots$ by division in descending powers.

$f(-1 + h) = \dfrac{-3 + \dots}{-2h + \dots} = \dfrac{3}{2h} + \dots$ by division in ascending powers.

$f(+1 + h) = \dfrac{-1 + \dots}{2h + \dots} = \dfrac{-1}{2h} + \dots$ by division in ascending powers.

We then sketch the local graphs and interpolate smoothly:



$$f_{\text{È}}(x) = \frac{1}{x} + \ldots \quad f_{-1}(h) = \frac{3}{2h} + \ldots \quad f_{+1}(h) = -\frac{1}{2h} + \ldots \qquad \text{Essential graph}$$

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.

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 $\boldsymbol{F}(x_0)$.

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 $\lim_{h \varnothing 0} \boldsymbol{f}(h)/\boldsymbol{g}(h) = 0$, 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.

If we then assume the existence of an antiderivative $\boldsymbol{F}(x)$, we have immediately from our definition of differentiability:

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

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:

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

This is always true but since we have no information on the various remainders $h\text{Ù}\mathbf{o}_{\boldsymbol{i}}[1]$, we have no way to evaluate the term $\sum_{1}^{n} \mathbf{o}_{i}[1]$. *So,* we have a good reason to let n approach È. Clearly, for $\boldsymbol{f}(x)$ smooth enough, $h\sum_{1}^{n} \mathbf{o}_{i}[1]$ approaches 0 as n approaches È and thus we obtain

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

where the Riemann sum $\sum_{0}^{n-1} \boldsymbol{f}(x_i)$ can *then* be easily interpreted geometrically as the approximation of $\int_{x_0}^{x_1} \boldsymbol{f}(x)\, dx$, the area under the graph of $\boldsymbol{f}$. We thus have, (Picard, 1901):

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

Note. Some of the pedagogical advantages of Lagrange's viewpoint are discussed in (Schremmer, & Schremmer, 1989). For some aspects pertaining to "calculus literacy", see (Schremmer, & Schremmer, 1988). Finally, (Mattei, & Schremmer, 1988) is a "task" implementation of Lagrange's approach. Interested readers are invited to write for (p)reprints.

References.

Bivins, I. C. (1986). What a Tangent Line is When it isn't a Limit., *The College Mathematics Journal,* **17,** 133-143.

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

Lagrange, J. L. (1797). *Théorie des fonctions analytiques*. Paris: Gauthier-Villars.

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

Levi, H. (1968). *Polynomials, Power Series and Calculus*. Princeton: Van Nostrand.

Mattei, F., & Schremmer, A. (1988). *Differential Calculus, a Lagrangian Approach. Vol. 1 and 2*. Christiana: MetaMath, Inc.

Picard, E. (1901). *Traité d'Analyse*. (2nd ed.). Paris: Gauthier–Villars.

Schremmer, F., & Schremmer, A. (1988). The Differential Calculus as Language. *Bulletin of Science, Technology and Society*, **8,** 411-418.

Schremmer, F., & Schremmer, A. (In preparation a). Polynomial approximations of rational functions.

Schremmer, F., & Schremmer, A. (In preparation b). Polynomial approximations of functions defined by equations.

Schremmer, F., & Schremmer, A. (1989). Integrated Precalculus-Differential Calculus: A Lagrangian Approach. *The AMATYC Review*, **11**, 1 (Part 2), 28-31.

The 1987 MAA Awards for Expository Excellence. (1988, January-February). *FOCUS*, p. iii.

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