THE JOURNEY TO A PROOF: IF F' IS POSITIVE, THEN F IS AN INCREASING FUNCTION

CHORLAY, Renaud*

IREM de Paris, 175 rue du Chevaleret, 75013, Paris, France
& IUFM de Paris, 10 rue Molitor, 75016, Paris, France
renaud.chorlay@paris.iufm.fr

ABSTRACT

When moving on from secondary to tertiary education, students are - in most countries - faced with new challenges in terms of proof: all theorems are, from then on, proved by the lecturer (which calls for proof-understanding skills); student are now expected to devise proofs of a more or less formal nature. As a consequence, the issues of proof-understanding and proof-writing have long been focal points in the research on AMT (Advanced Mathematical Thinking). Numerous strategies have been put forward - and sometimes tried out with students -, among which: (1) to distinguish between proof-ideas (or proof-germs) and formal proofs, and have students write formal proofs from informal ones [Downs & Mamona-Downs 2010], (2) to study historical proofs [Robert & Schwartzenberger 1991].

We will present a series of historical texts which lead to the now standard proof of the fact that, for a differentiable function of one real variable, the sign of the derivative determines the variations of the function (on an interval). Several features of this historical file are relevant from a maths-education perspective : (1) it illustrates the role of "*local*" counter-examples (to use Lakatosian terminology), a role which may not be familiar to students (although some students may be used to dealing with "global" counter-examples); (2) the various proofs (or proof-attempts) are based on at least *two* pretty different proof-ideas; (3) even a *proper* (meaning, both intuitive and formal) understanding of the concepts involved in the statement of the theorem may lead to a faulty proof scheme; (4) it helps understand the necessity of such intricate concepts as "uniform upper bound" or the completeness of **R**.

For this workshop, we will provide translations of French and German original sources, with excerpts from Lagrange, Cauchy, Serret, Peano, Darboux and Weierstrass.

Keywords: mathematical analysis, calculus, AMT, proof design, proof analysis.

1 Rationale

In the twentieth century, most tertiary-level textbooks of mathematical analysis prove the following theorem: let f be a differentiable real-valued function defined on an interval, if its derivative f'

*Large parts of this paper derive from joint work with Anne Michel-Pajus, and Philippe Brin. A fundamental starting point was [Dugac 1979].

is positive, then f increases over this interva[1]. Its standard proof is a rather straightforward application of the "mean value theorem"[2] ("*égalité des accroissements finis*" in French, "*Mittelwertsatz*" in German); the proof of which is a rather straightforward application of the "Rolle theorem"Let f be a differentiable real-valued function, defined over some interval [a,b], such that f(a) = f(b). There is a value c between a and b for which de derivative vanishes.[3], which, in turn, depends on the fact that a continuous real-valued function defined on a closed and bounded (i.e. compact) interval has a maximum or a minimum. The latter fact, although quite intuitive, depends on not-so-trivial properties of the set of real numbers (completeness of the metric space, local compactness). Historically speaking, this proof-chain can be found in the textbooks of Jordan [Jordan 1893, 65-67], Stolz [Stolz 1893, 51-], Osgood [Osgood 1912, 26-28]; in his chapter on differential and integral calculus for the famous *Enzyklopädie der mathematischen Wissenschaften* [Voss 1899, 65-66], Voss states in the clearest of ways that the mean value theorem in calculus depends on Weierstrass' theorem on the existence of extrema for continuous functions.
With this example, we can see that the proof of a rather intuitive qualitative fact (namely: if all the tangents point upward, the curve has to move up) requires several layers of sophisticated concepts (differentiability, continuity, properties of the numerical continuum), and a few silly tricks (affine changes of variable). In this workshop, we will present some of the proofs given, over the 19^{th} century, either of this mathematical fact, or of some key points in its proof.
We must stress the fact that this paper was designed for a *reading workshop*: it is by no means a research report on the "history" of that theorem - whatever that may mean. Among other things, we do not aim for a comprehensive overview; the historical connections between the various authors are hardly mentioned; nor are the institutional and intellectual contexts of the various research or teaching programs. Our main goal is to make a few important texts available to an audience of teachers and researchers who are not familiar with the French or the German languages. We hope this selection of texts, and the points we will highlight as we read them, will provide food for thought, and trigger further work, be it classroom work, or more theoretical work in the teaching and learning of (advanced) mathematics.

2 Lagrange's proof (1806)

As we saw earlier, any function $f(x+i)$ can be expanded into the series

$$f(x) + if'(x) + \frac{i^2}{2!}f''(x) + \frac{i^3}{3!}f'''(x) + \cdots ,$$

which naturally goes on to infinity, unless the derived functions vanish, which is the case when f(x) is an entire rational function of x.
As long as this expansion is used for the sole generation of derived functions, it is indifferent whether the series goes to infinity or not; it is also the case, when the expansion is seen as a mere analytical transformation of the function; but if one wants to use it to get the value of the function in particular

[1]For the analysis of a teaching experience, see [Praslon 1994].

[2]Let f be a differentiable real-valued function, defined over some interval $[a,b]$, there exists a value c between a and b such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. Geometrically speaking: on the arc of curve joining the points $(a, f(a))$ and $(b, f(b))$, there is a point where the tangent is parallel to the chord joining the two endpoints.

[3]Let f be a differentiable real-valued function, defined over some interval $[a,b]$, such that $f(a) = f(b)$. There is a value c between a and b for which de derivative vanishes.

cases, in which case it displays an expression of a simpler form - quantity i having been released from the function - then, since only a given number of terms can be taken into account, it is important to have means to assess the remainder of the series which we neglect, or, at least, to find bounds to the error that we make by neglecting this remainder.

The determination of these bounds is above all important in the application of the Theory of functions to the Analysis of curves, and to Mechanics, so as to impart on this application the rigour of ancient geometry, as can be seen in part two of the *Theory of analytic functions*.

In the solution which I gave in the above mentioned work, I first found the exact expression of the remainder of the series, then determined bounds for that expression. But these bounds can also be found in a more elementary way, which is just as rigorous. For this purpose, we shall establish this general principle, which can be of use in several occasions:

A function which vanishes when the variable vanishes, will, as the variable increases positively, have finite values of the same sign as that of its derived function; or of the opposite sign if the variable increases negatively, as long as the values of the derived function keep the same sign and do not become infinite.

This principle is very important in the theory of functions, since it establishes a general relationship between the state of primitive functions and that of derived functions, and also helps determine bounds[4] for functions for which only the derivatives are known.

We shall prove it rigorously.

Let us consider the function $f(x+i)$, whose general development is

$$f(x) + if'(x) + \frac{i^2}{2!}f''(x) + \cdots .$$

As we saw in the former lesson, the form of the development may be different for some specific values of x; but we saw that, as long as $f'(x)$ is not infinite, the first two terms of the expansion are exact; and that the other terms will, consequently, contain powers of i greater than the first, so that we shall have

$$f(x+i) = f(x) + i[f'(x) + V],$$

V being a function of x and i, which vanishes when $i = 0$.

So, since V vanishes when i vanishes, it is clear that, should i be made to increase from zero through insensible degrees, the value of V would also increase from zero by insensible degrees, either positively or negatively, up to a certain point, after which it may decrease; consequently, one will always be able to assign to i a value such that the corresponding value of V - regardless of the sign - is less than any given quantity, and that for lesser values of i, the values of V are also lesser.

Let D be a given quantity, which may be chosen as small as one pleases; one can always assign to i a value so small that the values of V are bounded by the limits D and $-D$; so, since we have

$$f(x+i) = f(x) + i[f'(x) + V],$$

It follows that the quantity $f(x+i) - f(x)$ will be bound by these two

$$i[f'(x)\pm D].$$

[4]We chose the work "bound" to translate Lagrange's use of the word "*limites*" in this context.

Since this conclusion holds for any value of x, as long as $f'(x)$ is not infinite, it will hold when

$$x+i, x+2i, x+3i, ..., x+(n-1)i$$

are substituted for x; so that one can always choose i positive and small enough for all the quantities

$$\begin{aligned}&f(x+i)-f(x),\\&f(x+2i)-f(x+i),\\&f(x+3i)-f(x+2i),\\&...,\\&f(x+ni)-f(x+(n-1)i),\end{aligned}$$

to be respectively bound between the limits

$$\begin{aligned}&i[f'(x)\pm D],\\&i[f'(x+i)\pm D],\\&i[f'(x+2i)\pm D],\\&...\\&i[f'[x+(n-1)i]\pm D],\end{aligned}$$

taking the same quantity D in all these limits, which is allowable so long as none of the quantities

$$f'(x), f'(x+i), f'(x+2i), ..., f'(x+(n-1)i)$$

is infinite. So, if all these quantities are of the same sign, that is, all positive of negative, it is easy to conclude that their sum, which amounts to

$$f(x+ni)-f(x)$$

is bounded by the sum of the bounds, that is by the quantities

$$if'(x)+if'(x+i)+if'(x+2i)+\cdots+f'[x+(n-1)i]\pm niD.$$

So, if the arbitrary quantity D is chosen less than the sum

$$f'(x)+f'(x+i)+f'(x+2i)+\cdots+f'[x+(n-1)i]$$

divided by n, then, if we do not take into account the sign of this sum, the quantity $f(x+ni)-f(x)$ will necessarily be bound between zero and the sum

$$2i[f'(x)+f'(x+i)+f'(x+2i)+\cdots+f'[x+(n-1)i]].$$

So, if P is the largest positive or negative value of the quantities

$$f'(x), f'(x+i), f'(x+2i), ..., f'(x+(n-1)i),$$

the quantity $f(x+ni)-f(x)$ will be bound between zero and $2niP$.
And yet, since when taking i as small as we wish, n can - at the same time - be taken as large as we

wish, we can assume that in is equal to any given quantity z, positive or negative, since quantity i can be taken positive or negative.
The quantity $f(x+ni) - f(x)$ will thus become $f(x+z) - f(x)$, and can be used to represent any function of z which vanishes for $z = 0$, quantity x being now seen as an arbitrary constant. Similarly, the quantity $f'(x+ni)$ will become $f'(x+z)$, and will represent the derived function of the same function of z, since $f'(x+z)$ is also the derived function of $f(x+z)$, either with respect to x or to z. Hence, one may conclude generally that, if $f'(x+z)$ constantly takes on finite values of the same sign, and if P denotes the largest of these values - regardless of the sign - the primitive function will be bound between 0 and $2zP$; consequently, it will also remain finite, and of the same sign as the derives function if z is positive, or of the opposite sign if z is negative. [Lagrange 1884, 86-89]
This passage clearly demonstrates that Lagrange was not a proponent of a *purely* formal analysis. Of course, in the preface of the *Théorie des fonctions analytiques* (and even in the subtitle), he rejected the notion of limit as a proper foundation and starting point for a systematic development of mathematical analysis. Indeed, he defined the derivative f' of a function f as the coefficient of i in the power series expansion

$$f(x+i) = f(x) + if'(x) + \frac{i^2}{2!}f''(x) + \frac{i^3}{3!}f'''(x) + \cdots$$

However, he was also concerned with numerical aspects in which issues of convergence, and lower and upper bounds are of the essence. In particular, he determined upper bounds for the integral remainder in the Taylor-Lagrange expansion, in order to assess the degree of approximation given by a partial series expansion, and to establish convergence in some important cases[5].
In this passage, we can see that Lagrange also had a proper numerical understanding what the value of the derivative at a given point represents, and that he did interpret limits as relationships of dependence between inequalities. For instance, he rephrased "V being a function of x and i, which vanishes when $i = 0$" as "one will always be able to assign to i a value such that the corresponding value of V - regardless of the sign - is less than any given quantity (...)".
However, in spite of the fact that the theorem Lagrange set out to prove is correct, and that the proof relied on a correct numerical understanding of the derivative construed as a limit, something does not sound right in the proof. The modern, 21st-century reader probably feels a bit uneasy when reading the summing argument, and the conclusions derived by passing to the limit. Even if we do not know exactly where things go wrong, we feel too many variables depend on one another in more than one way for the final limiting arguments to be safe and sound.
However, this "gut feeling" of disbelief, this red signal flashing before your eyes as we read the proof, is ascribable to our maths education: we were taught to distrust this kind of reasoning. In §5 of this paper, we will endeavour to shed some light on the historical roots of this distrust.

3 Cauchy's proof (1823)

> Problem. Assuming that the function $y = f(x)$ is continuous relative to x in the neighborhood of specific value $x = x_0$, one asks whether the function increases or decreases as from this value, as the variable itself is made to increase or decrease.

[5]For a recent historical analysis of Lagrange's work, see [Ferraro & Panza 2012].

> Solution. Let Δx, Δy denote the infinitely small and simultaneous increments of variables x and y. The $\Delta y/\Delta x$ ratio has limit $dy/dx = y'$. It has to be inferred that, for very small numerical values of Δx and for a specific value x_0 of variable x, ratio $\Delta y/\Delta x$ is positive if the corresponding value of y' is positive and finite. $(\cdots)$
> This being settled, let's assume function $y = f(x)$ remains continuous between two given limits $x = x_0$ and $x = X$. If variable x is made to increase by imperceptible degrees from the first limit to the second one, function y shall increase every time its derivative, while being finite, has a positive value. [Cauchy 1823, 37]

Unlike Lagrange, Cauchy defined the derivative as a limit; just like Lagrange, he was able to derive proper numerical conclusions from this numerical conception of the derivative. So what makes his argument so different from Lagrange's ? Actually, they do not have the exact same understanding of what the *conclusion* to be reached is; both have implicit definitions[6] of what it is for a function to be increasing, but their definitions do not match exactly. Lagrange's definition is closer to the one we find in today's textbook: a real valued function defined over some interval I is an increasing function if, a and b being any elements of I, $a < b$ implies $f(a) < f(b)$. Lagrange's (implicit) definition reads slightly differently, since he compared the values of f at 0 and at any other *given* value[7].

Cauchy's implicit definition of an increasing function can be rephrased as follows: a real-valued function f defined over some interval I is an increasing function if, a being any element of I, there is a neighbourhood N_a of a such that, for any x in N_a, the order between $f(a)$ and $f(x)$ is the same as that between a and x. Lagrange's definition is global, point-wise, and refers to two (arbitrarily, independently) given points; Cauchy's definition is one in which some local property holds in the neighbourhood of every (arbitrarily) given point. It can be shown - but it takes a little work - that both definitions are actually equivalent from a mathematical viewpoint. However, they differ significantly, both from an epistemological viewpoint (in which, for instance, the difference between local and global properties are put to the fore), and from a cognitive viewpoint [Chorlay 2007, 2011].

The fact that both definitions coincide from a mathematical viewpoint does not imply that proving that the first holds involves the same kind (and amount) of work than proving that the second holds. The information we start with (sign of the derivative) being of the everywhere-local-type[8], a mere rephrasing of the hypotheses leads to Cauchy's definition of increasing functions, hence to the conclusion. Reaching Lagrange's conclusion involves patching up local pieces of information to reach global conclusions, an endeavour which the modern reader knows to be usually tricky.

To conclude this paragraph, we must add that, on other occasions, Cauchy himself used the same kind of reasoning that Lagrange used in the above quoted proof. For instance, in the third volume of his *Cours d'analyse à l'école Polytechnique*, he set out to prove the existence of the solution to an ordinary, first degree differential equation $y' = f(x, y)$ in the neighbourhood of a regular point [Cauchy 1981[9]]. His proof relied to some extent on the same idea as Lagrange's: the derivative provides local affine

[6] We could use the notion of "in-action definitions" [Ouvrier-Buffet 2011]. For an analysis of the notion of functional variation, see [Chorlay 2010]. For a more detailed analysis of Cauchy's proof, see [Chorlay 2007, 2011].

[7] It is however completely equivalent to the $a - b$ definition, since the specific value 0 plays no part in the proof.

[8] We do not need to distinguish here between ≪local≫ and ≪infinitesimal≫ [Chorlay 2007].

[9] These lectures were probably delivered in the 1820s, but were not published by Cauchy nor included in the *Oeuvres complètes*.

approximations of the required function; these affine approximations are to be patched up to form a piece-wise linear function; these are then taken to the limit as the subdivision step tends to zero. For the 21st-century reader, this proof has basically the same flaws as Lagrange's proof: continuity is assumed to be uniform; same for the convergence of the sequence of functions.

4 Bonnet's proof (in J.-A. Serret's textbook, 1868)[10]

Theorem I.- *Let $f(x)$ be a function of x which remains continuous for values of x between two given limits, and which, for these values, has a well-determined derivative $f'(x)$. If x_0 and X denote two values of x between these same limits, the following*

$$\frac{f(X) - f(x_0)}{X - x_0} = f'(x_1),$$

will hold, with x_1 a value between x_0 and X.
Indeed, the ratio

$$\frac{f(X) - f(x_0)}{X - x_0}$$

has, by hypothesis, a finite value; and, if A denotes this value, we will have

$$[f(X) - AX] - [f(x_0) - Ax_0] = 0. \tag{1}$$

Let $\varphi(x)$ denote the function of x defined by the formula

$$\varphi(x) = [f(x) - Ax] - [f(x_0) - Ax_0], \tag{2}$$

then, from equality (1),

$$\varphi(x_0) = 0, \varphi(X) = 0,$$

so that $\varphi(x)$ vanishes for $x = x_0$ and for $x = X$. Let us assume, for instance, that $X > x_0$, and let x increase from x_0 to X; at first, the value of $\varphi(x)$ is zero. If we assume that this function is not everywhere zero, for values of x between x_0 and X, it will have to either begin to increase, thus taking on positive values, or begin to decrease, thus taking on negative values; be it from $x = x_0$, or from some other value of x between x_0 and X. If these values are positive, since $\varphi(x)$ is continuous and vanishes for $x = X$, it is obvious that there will be a value x_1 between x_0 and X such that $\varphi(x_1)$ is greater than or equal to the neighbouring values

$$\varphi(x_1 - h), \quad \varphi(x_1 + h),$$

h being an arbitrarily small quantity. (...)
This, in either cases, the value x1 will be such that the differences

[10] Unfortunately we did not use the 1868 edition, but a later edition. We know the editions do not differ, as far as the quoted passages are concerned (see [Dugac 1979]).

$$\varphi(x_1-h)-\varphi(x_1), \quad \varphi(x_1+h)-\varphi(x_1)$$

Will be of the same sign; consequently, the ratios

$$\frac{\varphi(x_1-h)-\varphi(x_1)}{-h}, \quad \frac{\varphi(x_1+h)-\varphi(x_1)}{h} \tag{3}$$

will be of opposite signs. (...)
Both ratios (3) tend to the same limit when h tends to zero, since we assumed $f(x)$ has a well-determined derivative; hence, so does $\varphi(x)$; besides, these two ratios are of opposite signs: hence their limit is zero. So, one has,

$$\lim \left[\frac{\varphi(x_1+h)-\varphi(x_1)}{h}\right] = 0,$$

or, taking equation (2) into account,

$$\lim \left[\frac{f(x_1+h)-f(x_1)}{h} - A\right] = 0,$$

i.e.

$$A = \lim \frac{f(x_1+h)-f(x_1)}{h} = f'(x_1).$$

Therefore

$$\frac{f(X)-f(x_0)}{X-x_0} = f'(x_1),$$

as we claimed. (...)
Comment.- The above proof is due to Mr. Ossian Bonnet. It should be noticed that no assumptions are made as to the continuity of the derivative $f'(x)$; one merely assumes that it exists and has a well-determined value.
Theorem II.- *If function $f(x)$ is constant for all the values of x between two given limits, the derivative $f'(x)$ vanishes for these values of x. Conversely, if the derivative $f'(x)$ vanishes for all values of x between two limits, the function $f(x)$ has a constant value for the values of x between these limits. (...)*
Theorem III.- *If the derivative $f'(x)$ of function $f(x)$ remains finite for all the values of x between the limits x_0, if $X > x_0$, and if x is made to increase from x_0 to X, the function $f(x)$ will increase as long as the derivative $f'(x)$ will not be negative, and it will decrease as long as $f'(x)$ will not be positive.*
Indeed, since x lies between x_0 and X, the ratio

$$\frac{f(x \pm h)-f(x)}{\pm h}$$

has limit $f'(x)$, which is a finite quantity; so it will of the same sign as that of the limit, for values of h between zero and some sufficiently small positive quantity ε. Consequently, for these values of h, the following will hold

$$f(x-h) < f(x) < f(x+h)$$

if $f'(x)$ is > 0, and

$$f(x-h) > f(x) > f(x+h)$$

if $f'(x)$ is < 0.
Thus, the function $f(x)$ will increase, as from any value of x for which $f'(x)$ is > 0; and decrease, as from any value of x for which $f'(x)$ is < 0. [Serret 1900, 17-22]

In this passage, Serret introduced Ossian Bonnet's proof of the mean value theorem, a proof idea which relied on an affine change of variable and the vanishing of the derivative at a local extremum. The existence of the extremum is not proved (at least when one compares with later rewritings of this proof), but made obvious in the *narrative* style which is so typical of the first half of the 19^{th} century. Strikingly, Serret did not use the mean value theorem to establish the relationship between the sign of f' and the variations of f; he relied on Cauchy's argument, hence on Cauchy's notion of functional variation. However, the mean value theorem was used to establish theorem II. Actually, quite a few textbook writers made the same choice in the second half of the 19^{th} century. For instance, in the very Weierstrassian textbook by Genocchi and Peano, Cauchy's proof is given first; then comes the proof of the mean-value theorem, from which Serret's theorem II is derived [Genocchi & Peano 1889, 43].

5 Proof-analysis and regressive analysis

5.1 Proof-analysis: the role of uniform convergence

We identified in Lagrange's proof a flaw which can be described in several ways: implicit assumption of uniform differentiability; failure to notice that some variable is dependent on some other, while trying to consider the limit of second while leaving the first fixed; exchange, without due caution, of two limiting processes. The same flaws were common to most proofs in analysis which dealt with the numerical aspect of functions (as opposed to formal aspects); to name a few: Ampère's proof of the inequality form of the mean value theorem[11] [Ampère 1806], Cauchy's proof of the same [Cauchy 1823, 44-45[12]], Cauchy's proof that the limit of a sequence of continuous functions is continuous [Cauchy 1821, 120], Cauchy's proof of the existence of a local solution to a first degree differential equation in the neighbourhood of a regular point [Cauchy 1981, chapter 7] etc.
It is well known that the difference between point-wise and uniform[13] continuity (for a function), and point-wise and uniform convergence (for a sequence of functions) was stated in the clearest of

[11]Namely, that $|f(b) - f(a)|$ over $|b - a|$ is less than or equal to the maximal value of $f'(x)$ on $[a, b]$.
[12]The page numbers refer to the *Œuvres*.
[13]"*gleichmässig*" in German.

ways in Weierstrass' Berlin lectures on the foundations of analysis; a distinction which he attributed to his master Gudermann. The awareness of the importance of this distinction spread in the 1870s and 1880s among students, followers or readers of Weierstrass (Hurwitz, Cantor, Schwartz, Dini, Peano, Pincherle, du Bois Reymond, Heine, Thomae etc.). Of course, this awareness was displayed in a new generation of textbooks and research papers; it also spread through criticism of faulty proofs found in papers or textbooks written by the most distinguished mathematicians.
An instance of this is given by Peano's criticism of the proof of the mean value theorem, which he read in the first edition of Jordan's *Cours d'analyse de l'école Polytechnique*. The exchange of letters was published in 1887, in the *Nouvelles annales de mathématiques*:

> Mr. Jordan gave a not quite rigorous proof of the following theorem:
> "Let $y = f(x)$ be a function of x whose derivative remains finite and well-determined when x varies in some interval. Let a and $a + h$ be two values of x in this interval. We will have
>
> $$f(a + h) - f(a) = \mu h,$$
>
> where μ denotes a quantity between the largest and the smallest values of $f'(x)$ in the interval between a and $a + h$."
> Indeed, Jordan writes, let x take on a series of values a_1, a_2, ..., a_{n-1} between a and $a + h$; let us set
>
> $$f(a_r) - f(a_{r-1}) = (a_r - a_{r-1})[f'(a_{r-1}) + \varepsilon_r].$$
>
> Let us now assume that the intermediate values a_1, ..., a_{n-1} are indefinitely multiplied (and brought closer together). The quantities ε_1, ε_2,... will all tend to zero, since ε_r is the difference between $\frac{f(a_r)-f(a_{r-1})}{a_r-a_{r-1}}$ and its limit $f'(a_{r-1})$.
> The latter assumption is not correct; for
>
> $$f'(a_{r-1}) = \lim \frac{f(a_r) - f(a_{r-1})}{a_r - a_{r-1}}$$
>
> when a_{r-1} is assumed to be fixed, and a_r variable and approaching a_{r-1} infinitely closely; but one cannot claim the same when both a_r and a_{r-1} vary, unless the derivative is assumed to be continuous.
> Indeed, for instance, let us set
>
> $$y = f(x) = x^2 \sin \frac{1}{x},$$
>
> with
>
> $$f(0) = 0;$$
>
> its derivative

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

for $x \gtrless 0$, and $f'(0) = 0$, remains finite and well-determined, but it is discontinuous.
Let

$$a = 0, \quad h > 0;$$

let us set

$$a_1 = \frac{1}{2n\pi}, \quad a_2 = \frac{1}{(2n+1)\pi},$$

a_3, a_4, ... any numbers.
One then has

$$\varepsilon_2 = \frac{f(a_2) - f(a_1)}{a_2 - a_1} - f'(a_1);$$

but

$$f(a_1) = 0, \quad f(a_2) = 0, \quad f'(a_1) = -1;$$

hence

$$\varepsilon_2 = 1,$$

which does not tend to zero.
Nearly the same mistake was made by Mr. Hoüel (*Cours de Calcul infinitesimal*, t.I, p.145).
Eventually, I shall add that the formula

$$f(x_0 + h) - f(x_0) = hf'(x_0 + \theta h),$$

Can be established very easily, and without assuming the continuity of f'. [Peano 1884, 45]

In fact, very similar objections had been made a few years earlier to Jules Hoüel himself ! In the 1870s, Hoüel and Darboux co-edited the *Bulletin des sciences mathématiques*, and Hoüel asked Darboux for comments on the drafts of his lectures on analysis, which he would eventually publish in 1878. Darboux was well aware of the recent developments in Berlin, and criticized many of the "classical" proofs Hoüel planned to rely on. Unlike the Peano-Jordan exchange of letters, this correspondence remained private, and was partly published in the 1970s and 1980s [Gispert 1983]. Here is Darboux's view on Hoüel's Lagrange-style proof of the mean value theorem, in a letter dated February 4, 1874:

As to §52, which plays a fundamental part in your argument, I also find fault with it, namely: the ε quantities are functions of two variables. For instance, setting $x_3 - x_2 = h$,

$$f(x_2 + h) - f(x_2) = hf'(x_2) + h\varepsilon_2.$$

> Clearly, ε_2 is infinitely small, and is a function of x and h about which *only* the following is known: it tends to zero with h, when x remains fixed; but then, I claim that you do not know what becomes of it when, as h tends to zero, x_2 varies with h, which is the case in your decomposition. For instance, consider
>
> $$\frac{h_2}{x_1 - a + h}.$$
>
> For any x_1, *so long as it remains fixed*, this expression vanishes with h. But if x varies, for instance if we have
>
> $$x_1 = a - h + h^4$$
>
> Then the expression simplifies into $\frac{1}{h^2}$, which becomes infinitely large as h tends to zero.
> I am telling you this for I am deeply convinced that if you stuck more closely to rigour, you would come up with a treatise of infinitesimal calculus of exceptional interest.
> If I were you, I would give up on the theorem on the limit of sums, which is worthless, just as many other things. With the mean value theorem, such as established by Serret, you could build a strong structure. This, along with the definition of the integral, is all one needs. This is how Weierstrass does it, I believe. [Gispert 1983, 89-90]

Needless to say Hoüel failed to be convinced, even though Darboux repeatedly voiced dissent and disbelief. He sent out several other counter-examples, and rephrased the main argument in various ways. Even though he did not use the terms "uniform convergence", the notion was perfectly clear to him, as we can see in this letter, dated January 18, 1875:

> Here is where I find fault with your reasoning, which no one deems rigorous any more[14].
> When setting
>
> $$\frac{f(x+h) - f(x)}{h} - f'(x) = \varepsilon,$$
>
> ε is a function of the two variables x and h which tends to zero when, leaving x fixed, h vanishes. But if x and h vary, as in your proof; even more, if every new subdivision $x_1 - x_0$ generates new ε quantities, I cannot see anything clearly any more, and your proof becomes only seemingly rigorous. (...) You could get out of this predicament in one of two ways, 1. By changing proofs altogether, which I advise you to do[15]. 2. By proving that if a function always admits a derivative between x_0 and x_i, one can find a quantity h such that for all values of x between x_0 and x_i, and all values x_0 and h_1 of h less than some limit value, one has
>
> $$\frac{f(x+h) - f(x)}{h} - f'(x) < \varepsilon,$$

[14] This is an optimistic overstatement from Darboux.
[15] Darboux strongly recommended the Bonnet proof.

> where ε has a value which is fixed but chosen as small as one wishes; which is difficult[16]. [Gispert 1983, 99-100]

5.2 Regressive analysis: the role of the existence theorem for extrema

A critical mind might object to Bonnet's proof of the mean value theorem that it depends on the existence of a maximum or a minimum, an existence which is implicitly taken for granted. It seems clear that if the function is piece-wise monotonous (as seems to be assumed in the text), it will indeed admit either a local maximum or a local minimum; but a differentiable function needs not be piece-wise monotonous, as the ever useful example $f(x) = x^2 \sin \frac{1}{x}$ shows.
In fact, the existence of a maximum can be grounded without piece-wise monotony, or continuous differentiability, as Weierstrass established, for instance in his 1878 lectures on the theory of functions. The following passage has nothing to do a *priori* with calculus. It comes after the construction of the set of real numbers **R** (or, more precisely, the affinely extended real number line $\overline{\mathbf{R}} = [-\infty, +\infty]$) starting from rational numbers. The theorem about nested intervals was established on this basis, as well as the existence theorem of an upper bound (*obere Grenze*) for any non-empty subset of the extended real number system.

> Let a value y correspond to every point $(x_1, \ldots, x_n)$ of some domain; then y is also a variable quantity, hence has a lower and an upper bound; let g denote it. Then, there exists at least one point in the x-domain (that point needs not belong to the defined domain)[17], with the following property: if we consider however small a neighbourhood of that point, and consider the values of y corresponding to that x-domain, then these values of y also have an upper bound, this upper bound being exactly g. Similarly for the lower bound. [Weierstrass 1988, 91]

What we have here is a purely set-theoretic theorem; the function does not have to be continuous; no hypotheses are made on the domain (if it is not closed, the point with the remarkable property may lie on its boundary). We skip the proof, which relies on the definition of the upper bound, and the (now familiar) technique of nested intervals. Weierstrass then turned to an application of this very abstract result in "everyday" analysis:

> One is commonly faced with the question: among the values taken on by some magnitude, is there a maximum or a minimum (maximum or minimum in the absolute sense[18]). Let y be a continuous function of x, $y = f(x)$. Here, x must remain between two given limits a and b. In which circumstances is there a maximum and a minimum for y? There is an upper bound for y. According to our proposition, there must be some point x_0 in the x-domain such that the upper-bound of the values of y for x between $x_0 - \delta$ and $x_0 + \delta$ is

[16] The fact that, if the derivative is continuous, then $\frac{f(x+h)-f(x)}{h}$ *does* tend to $f'(x)$ *uniformly* on every closed and bounded interval was proved, for instance, in the second edition of Jordan's textbook [Jordan 1893, 68].

[17] Indeed, it may lie on the boundary of the first domain.

[18] i.e. not a local maximum or minimum.

> also g. Point x_0 either lies inside $a...b$, or on its border ($x_0 = a$, or $x_0 = b$).
> In the first case, $f(x_0)$ is a maximum. Indeed, $f(x_0)$ must be equal to g: for $f(x) - f(x_0)$ can be made as small as we wish, by choosing an adequately small $|x - x_0|$; on the other hand, since x lies between $x_0 - \varepsilon$ and $x_0 + \varepsilon$, $f(x)$ can be chosen arbitrarily close to g; hence $f(x_0) = g$. (If we had $f(x_0) = g + h$, we would have $f(x) - f(x_0) = f(x) - g - h$, and $f(x)$ could not come arbitrarily close to g if h was not 0).
> If x_0 coincided with either a or b, then we could only claim that $f(a)$ (resp. $f(b)$ is a maximum if $f(x)$ was continuously at a (resp. b) as well. [Weierstrass 1988, 91-92]

A similar proof can be found in other texts, sometimes published before 1878, but all deriving directly or indirectly from Weierstrass' Berlin lectures: [Cantor 1871], [Heine 1872 186], [Darboux 1872]. Fifty years later, the full conceptual clarification of the notion of maximum would still be considered one of the achievements of Weierstrass' work on the foundation of analysis, as is shown by the first lines of Hilbert's famous 1925 paper *On the Infinite*:

> Weierstrass, through the critique elaborated with the sagacity of a master, created a firm foundation for mathematical analysis. By clarifying, among other notions, those of minimum, function, and derivative, he removed the remaining flaws from the calculus, cleansed it of all vague ideas concerning the infinitesimal, and conclusively overcame the difficulties that until then had their roots in the notion of infinitesimal. [Hilbert 1967, 369]

6 Conclusion

Let us attempt to summarize the pretty intricate network of definitions, proof-ideas (or proof-germs), proof-techniques, and proof-analyses displayed in this sample of texts.
At least two *definitions* of what it means for a real-valued function to "increase" can be found in the 19^{th}-century: a point-wise and global definition which can be found in Lagrange; a definition that relies on an everywhere-valid local property, which can be found in Cauchy. If we stick to Cauchy's definition, then the proof of the theorem about the relationship between the sign of f' and the variations of f is pretty trivial. If we want to reach the Lagrange-style conclusion, then much more work is needed, since one has to start from an everywhere-valid local property (sign of f') and reach a global conclusion.
To reach that conclusion, we saw two very different *proof-ideas*, namely Lagrange's and Bonnet's. In the proof we studied, and in quite a few other parts of his work, Lagrange distanced himself from the formal manipulation of formulae (finite or infinite), and engaged in numerical proof: he relied on the correct numerical understanding of the notion of limit; on this basis, he cautiously built networks of inequalities; he finally endeavoured to ground his reasoning on the determination of upper bounds for the errors in a process of affine approximation. In the first half of the 19^{th} century, many proofs of the most important theorems in function theory were written along this line. Distrust of this proof-scheme spread as mathematicians grew aware of the distinction between point-wise and uniform (continuity, convergence). They spread all the more slowly since the theorems were correct,

the building blocks of the proofs showed a proper understanding of the notions at stake, and local counterexamples[19] were hard to find. As Darboux insightfully (but to no avail, as far as Hoüel was concerned) stressed, there were only two ways out of this predicament: either to change proof-germs, or to establish uniformity[20].

For the theorem on which we chose to focus, an alternative proof became available in the 1860s, which relied on a completely different proof-idea; unlike Lagrange's proof, it did not rely on what the derivative of a function at a point *is* (a limit, which provides some local affine approximation), but on a *property* of the derivative (stated in the mean value theorem). Some elements of Bonnet's proof were later seen as insufficiently grounded, in particular the existence of a minimum or a maximum; in the 1890s, mathematicians such as Jordan or Stolz used Weierstrass' analysis of the set-theoretic properties of the real line to back up that weaker step in Bonnet's proof.

References

- Ampère, A.-M., 1806, "Recherche sur quelques points de la théorie des fonctions dérivées qui conduisent à une nouvelle démonstration de la série de Taylor, et à l'expression des termes qu'on néglige lorsqu'on arrête cette série à un terme quelconque", *Journal de l'Ecole Polytechnique* **13**, 148-181.

- Cantor, G., 1871, "Notiz zu dem Aufsatze: Beweis (···)", *Journal für die reine und angewandte Mathematik* 73, 84-86.

- Cauchy, A.-L., 1821, *Cours d'analyse de l'Ecole royale Polytechnique*, 1ère 이부분 출력 *partie Analyse Algébrique*, Paris : chez Debure frères = Cauchy, A.-L., Œuvres Complètes, série 2, tome III. Paris : Gauthiers-Villars, 1882-1974.

- Cauchy, A.-L., 1823, *Résumé des Leçons données à l'Ecole royale Polytechnique sur le Calcul infinitésimal* = Cauchy, A.-L., *Œuvres Complètes*, série 2, tome IV. Paris : Gauthiers-Villars, 1882-1974.

- Cauchy, A.-L., 1981, *Equations différentielles ordinaires - cours inédits* (edited by C. Gillain), Saint-Laurent: Etudes vivantes.

- Chorlay, R., 2007, "La multiplicité des points de vue en Analyse élémentaire comme construit historique", in *Histoire et enseignement des mathématiques : erreurs, rigueurs, raisonnements*, E. Barbin & D. Bénard (eds.), Lyon: INRP, pp.203-227.

- Chorlay, R., 2010, "From historical analysis to classroom work: function variation and long-term development of functional thinking", in *Proceedings of CERME 6* (Congress of the European Society for Research in Mathematics Education), Lyon: INRP édition électronique.

- Chorlay, R., 2011, "the multiplicity of viewpoints in elementary function theory : historical and didactical perspectives", *in Recent Developments on Introducing a Historical Dimension in Mathematics Education*, V. Katz and K. Tzanakis (eds.), Washington D.C.: The Mathematical Association of America, pp.55-63.

[19] Meaning: a counterexample to a step in a proof, not to the theorem itself.

[20] This second option was used to ground Cauchy's proof of the existence of solutions for 1st degree ODEs.

- Darboux, G., 1872, “Sur un théorème relatif à la continuité des fonctions”, *Bulletin des sciences mathématiques et astronomiques* **3**, 307-313.
- Downs, M., & Mamona-Downs, J., 2010, “Necessary Realignment from Mental Argumentation to Proof Presentation”, in *Proceedings of CERME 6* (Congress of the European Society for Research in Mathematics Education), Lyon: INRP édition électronique.
- Dugac, P., 1979, *Histoire du théorème des accroissements finis* (polycopié), Paris : Université Pierre et Marie Curie.
- Dugac, P., 2003, *Histoire de l’analyse*. Paris: Vuibert.
- Ferraro, G., & Panza, M., 2012, “Lagrange’s theory of analytical functions and his ideal of purity of method”, *Archive for History of Exact Sciences* **66(2)**, 95-197.
- Genocchi, A., & Peano, G., 1889, *Differentialrechnung und Grundzüge der Integralrechnung* (Translated into German by G. Bohlmann & A. Schepp, from the 1884 Italian original), Leipzig: Teubner.
- Gispert, H., 1983, “Sur les fondements de l’analyse en France”, *Archive for History of Exact Sciences* **28**, 37-106.
- Heine, E., 1872, “Die Elemente der Funktionenlehre”, *Journal für die reine und angewandte Mathematik* **74**, 172-188.
- Hilbert, D., 1967, On the infinite (translated from the German original by S. Bauer-Mengelberg), in *From Frege to Gödel: A Sourcebook in Mathematical Logic, 1879-1931*, J. van Heijenoort (ed.), Cambridge (Mas.): Harvard University Press, pp.367-392.
- Jordan, C., 1893, *Cours d’analyse de l’école Polytechnique*, tome premier (2ème édition, entièrement refondue), Paris: Gauthier-Villars.
- Lagrange, J.-L., 1884, *Leçon sur le calcul des fonctions* (reprint of the second édition (1806)), in J. Serret (ed.) *Oeuvres Complètes de Lagrange* vol.14, Paris: Gauthier-Villars.
- Osgood, W., 1912, *Lehrbuch der Funktionentheorie* (2te Auflage), Leipzig: Teubner.
- Ouvrier-Buffet, C., 2011, “A Mathematical Experience Involving Defining Processes: In-Action Definitions and Zero-Definitions”, *Educational Studies in Mathematics* **76**, 165-182.
- Peano, G., 1884, “Extrait d’une lettre de M. le Dr. J. Peano”, *Nouvelles annales de mathématiques* **3** (3ème série), 45-49.
- Praslon, F., 1994, *Analyse de l’aspect Méta dans un enseignement de Deug A concernant le concept de dérivée. Etude des effets sur l’apprentissage* (mémoire de D.E.A.), Paris: Université Paris Diderot.
- Robert, A., & Schwartzenberger, R., 1991, “Research in Teaching and Learning Mathematics at an Advanced Level”, in D. Tall (ed.), *Advanced Mathematical Thinking*, Boston: Kluwer Academic Press, pp. 127-139.
- Serret, J.-A., 1900, *Cours de calcul différentiel et intégral*(5$^{\text{ème}}$ édition), Paris: Gauthier-Villars.

– Stolz, O., 1893, *Grundzüge der Differential- und Integralrechnung* (vol.1), Leipzig: Teubner.

– Voss, A., 1899, "Differential- und Integralrechnung", in *Enzyklopädie der mathematischen Wissenschaften* II.1.1, Leipzig: Teubner.

– Weierstrass, K., 1988, *Einleitung in die Theorie der analytischen Funktionen, Vorlesung, Berlin 1878*, P. Ulrich (ed.), Braunschweig: D.M.-V. & Vieweg.