9-1-1983

The Changing Concept of Change: The Derivative from Fermat to Weierstrass

Judith V. Grabiner
Pitzer College

The Changing Concept of Change: The Derivative from Fermat to Weierstrass

First the derivative was used, then discovered, explored and developed, and only then, defined.

JUDITH V. GRABINER
Department of History
University of California, Los Angeles
Los Angeles, CA 90024

Some years ago while teaching the history of mathematics, I asked my students to read a discussion of maxima and minima by the seventeenth-century mathematician, Pierre Fermat. To start the discussion, I asked them, "Would you please define a relative maximum?" They told me it was a place where the derivative was zero. "If that's so," I asked, "then what is the definition of a relative minimum?" They told me, *that's* a place where the derivative is zero. "Well, in that case," I asked, "what is the difference between a maximum and a minimum?" They replied that in the case of a maximum, the second derivative is negative.
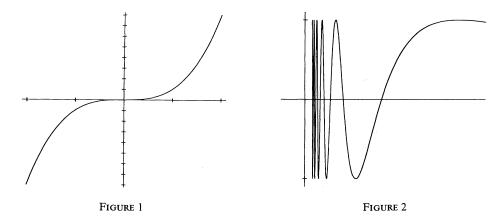
What can we learn from this apparent victory of calculus over common sense?

I used to think that this story showed that these students did not understand the calculus, but I have come to think the opposite: they understood it very well. The students' answers are a tribute to the power of the calculus in general, and the power of the concept of derivative in particular. Once one has been initiated into the calculus, it is hard to remember what it was like *not* to know what a derivative is and how to use it, and to realize that people like Fermat once had to cope with finding maxima and minima without knowing about derivatives at all.

Historically speaking, there were four steps in the development of today's concept of the derivative, which I list here in chronological order. The derivative was first *used*; it was then *discovered*; it was then *explored and developed*; and it was finally *defined*. That is, examples of what we now recognize as derivatives first were used on an ad hoc basis in solving particular problems; then the general concept lying behind these uses was identified (as part of the invention of the calculus); then many properties of the derivative were explained and developed in applications both to mathematics and to physics; and finally, a rigorous definition was given and the concept of derivative was embedded in a rigorous theory. I will describe the steps, and give one detailed mathematical example from each. We will then reflect on what it all means—for the teacher, for the historian, and for the mathematician.

The seventeenth-century background

Our story begins shortly after European mathematicians had become familiar once more with Greek mathematics, learned Islamic algebra, synthesized the two traditions, and struck out on their own. François Vieta invented symbolic algebra in 1591; Descartes and Fermat independently

FIGURE 1

FIGURE 2

invented analytic geometry in the 1630's. Analytic geometry meant, first, that curves could be represented by equations; conversely, it meant also that every equation determined a curve. The Greeks and Muslims had studied curves, but not that many—principally the circle and the conic sections plus a few more defined as loci. Many problems had been solved for these, including finding their tangents and areas. But since any equation could now produce a new curve, students of the geometry of curves in the early seventeenth century were suddenly confronted with an explosion of curves to consider. With these new curves, the old Greek methods of synthetic geometry were no longer sufficient. The Greeks, of course, had known how to find the tangents to circles, conic sections, and some more sophisticated curves such as the spiral of Archimedes, using the methods of synthetic geometry. But how could one describe the properties of the tangent at an arbitrary point on a curve defined by a ninety-sixth degree polynomial? The Greeks had defined a tangent as a line which touches a curve without cutting it, and usually expected it to have only one point in common with the curve. How then was the tangent to be defined at the point $(0,0)$ for a curve like $y = x^3$ (FIGURE 1), or to a point on a curve with many turning points (FIGURE 2)?

The same new curves presented new problems to the student of areas and arc lengths. The Greeks had also studied a few cases of what they called "isoperimetric" problems. For example, they asked: of all plane figures with the same perimeter, which one has the greatest area? The circle, of course, but the Greeks had no general method for solving all such problems. Seventeenth-century mathematicians hoped that the new symbolic algebra might somehow help solve all problems of maxima and minima.

Thus, though a major part of the agenda for seventeenth-century mathematicians—tangents, areas, extrema—came from the Greeks, the subject matter had been vastly extended, and the solutions would come from using the new tools: symbolic algebra and analytic geometry.

Finding maxima, minima, and tangents

We turn to the first of our four steps in the history of the derivative: its *use*, and also illustrate some of the general statements we have made. We shall look at Pierre Fermat's method of finding maxima and minima, which dates from the 1630's [8]. Fermat illustrated his method first in solving a simple problem, whose solution was well known: *Given a line, to divide it into two parts so that the product of the parts will be a maximum.* Let the length of the line be designated B and the first part A (FIGURE 3). Then the second part is $B - A$ and the product of the two parts is

$$A(B - A) = AB - A^2. \tag{1}$$

Fermat had read in the writings of the Greek mathematician Pappus of Alexandria that a problem which has, in general, two solutions will have only one solution in the case of a maximum. This remark led him to his method of finding maxima and minima. Suppose in the problem just stated there is a second solution. For this solution, let the first part of the line be designated as $A + E$; the second part is then $B - (A + E) = B - A - E$. Multiplying the two parts together, we obtain
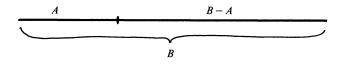
FIGURE 3

for the product

$$BA + BE - A^2 - AE - EA - E^2 = AB - A^2 - 2AE + BE - E^2. \qquad (2)$$

Following Pappus' principle for the maximum, instead of two solutions, there is only one. So we set the two products (1) and (2) "sort of" equal; that is, we formulate what Fermat called the pseudo-equality:

$$AB - A^2 = AB - A^2 - 2AE + BE - E^2.$$

Simplifying, we obtain

$$2AE + E^2 = BE$$

and

$$2A + E = B.$$

Now Fermat said, with no justification and no ceremony, "suppress E." Thus he obtained

$$A = B/2,$$

which indeed gives the maximum sought. He concluded, "We can hardly expect a more general method." And, of course, he was right.

Notice that Fermat did not call E infinitely small, or vanishing, or a limit; he did not explain why he could first divide by E (treating it as nonzero) and then throw it out (treating it as zero). Furthermore, he did not explain what he was doing as a special case of a more general concept, be it derivative, rate of change, or even slope of tangent. He did not even understand the relationship between his maximum-minimum method and the way one found tangents; in fact he followed his treatment of maxima and minima by saying that the same method—that is, adding E, doing the algebra, then suppressing E—could be used to find tangents [8, p. 223].

Though the considerations that led Fermat to his method may seem surprising to us, he did devise a method of finding extrema that worked, and it gave results that were far from trivial. For instance, Fermat applied his method to optics. Assuming that a ray of light which goes from one medium to another always takes the quickest path (what we now call the Fermat least-time principle), he used his method to compute the path taking minimal time. Thus he showed that his least-time principle yields Snell's law of refraction [7] [12, pp. 387–390].

Though Fermat did not publish his method of maxima and minima, it became well known through correspondence and was widely used. After mathematicians had become familiar with a variety of examples, a pattern emerged from the solutions by Fermat's method to maximum-minimum problems. In 1659, Johann Hudde gave a general verbal formulation of this pattern [3, p. 186], which, in modern notation, states that, *given a polynomial of the form*

$$y = \sum_{k=0}^{n} a_k x^k,$$

there is a maximum or minimum when

$$\sum_{k=1}^{n} ka_k x^{k-1} = 0.$$

Of even greater interest than the problem of extrema in the seventeenth century was the finding of tangents. Here the tangent was usually thought of as a secant for which the two points came closer and closer together until they coincided. Precisely what it meant for a secant to "become" a
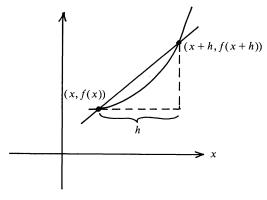
$(x+h, f(x+h))$

$(x, f(x))$

h

x

FIGURE 4

tangent was never completely explained. Nevertheless, methods based on this approach worked. Given the equation of a curve

$$y = f(x),$$

Fermat, Descartes, John Wallis, Isaac Barrow, and many other seventeenth-century mathematicians were able to find the tangent. The method involves considering, and computing, the slope of the secant,

$$\frac{f(x+h) - f(x)}{h},$$

doing the algebra required by the formula for $f(x+h)$ in the numerator, then dividing by h. The diagram in FIGURE 4 then suggests that when the quantity h vanishes, the secant becomes the tangent, so that neglecting h in the expression for the slope of the secant gives the slope of the tangent. Again, a general pattern for the equations of slopes of tangents soon became apparent, and a rule analogous to Hudde's rule for maxima and minima was stated by several people, including René Sluse, Hudde, and Christiaan Huygens [3, pp. 185–186].

By the year 1660, both the computational and the geometric relationships between the problem of extrema and the problem of tangents were clearly understood; that is, a maximum was found by computing the slope of the tangent, according to the rule, and asking when it was zero. While in 1660 there was not yet a general concept of derivative, there was a general method for solving one type of geometric problem. However, the relationship of the tangent to other geometric concepts—area, for instance—was not understood, and there was no completely satisfactory definition of tangent. Nevertheless, there was a wealth of methods for solving problems that we now solve by using the calculus, and in retrospect, it would seem to be possible to generalize those methods. Thus in this context it is natural to ask, how did the derivative as we know it come to be?

It is sometimes said that the idea of the derivative was motivated chiefly by physics. Newton, after all, invented both the calculus and a great deal of the physics of motion. Indeed, already in the Middle Ages, physicists, following Aristotle who had made "change" the central concept in his physics, logically analyzed and classified the different ways a variable could change. In particular, something could change uniformly or nonuniformly; if nonuniformly, it could change uniformly-nonuniformly or nonuniformly-nonuniformly, etc. [3, pp. 73–74]. These medieval classifications of variation helped to lead Galileo in 1638, without benefit of calculus, to his successful treatment of uniformly accelerated motion. Motion, then, could be studied scientifically. Were such studies the origin and purpose of the calculus? The answer is no. However plausible this suggestion may sound, and however important physics was in the later development of the calculus, physical questions were in fact neither the immediate motivation nor the first application of the calculus.

Certainly they prepared people's thoughts for some of the properties of the derivative, and for the introduction into mathematics of the concept of change. But the immediate motivation for the general concept of derivative—as opposed to specific examples like speed or slope of tangent—did not come from physics. The first problems to be solved, as well as the first applications, occurred in mathematics, especially geometry (see [1, chapter 7]; see also [3; chapters 4–5], and, for Newton, [17]). The concept of derivative then developed gradually, together with the ideas of extrema, tangent, area, limit, continuity, and function, and it interacted with these ideas in some unexpected ways.

Tangents, areas, and rates of change

In the latter third of the seventeenth century, Newton and Leibniz, each independently, invented the calculus. By "inventing the calculus" I mean that they did three things. First, they took the wealth of methods that already existed for finding tangents, extrema, and areas, and they subsumed all these methods under the heading of two general concepts, the concepts which we now call **derivative** and **integral**. Second, Newton and Leibniz each worked out a notation which made it easy, almost automatic, to use these general concepts. (We still use Newton's $\dot{x}$ and we still use Leibniz's dy/dx and $\int y\, dx$.) Third, Newton and Leibniz each gave an argument to prove what we now call the Fundamental Theorem of Calculus: the derivative and the integral are mutually inverse. Newton called our "derivative" a *fluxion*—a rate of flux or change; Leibniz saw the derivative as a ratio of infinitesimal differences and called it the *differential quotient*. But whatever terms were used, the concept of derivative was now embedded in a general subject—the calculus—and its relationship to the other basic concept, which Leibniz called the integral, was now understood. Thus we have reached the stage I have called *discovery*.

Let us look at an early Newtonian version of the Fundamental Theorem [13, sections 54–5, p. 23]. This will illustrate how Newton presented the calculus in 1669, and also illustrate both the strengths and weaknesses of the understanding of the derivative in this period.

Consider with Newton a curve under which the area up to the point $D = (x, y)$ is given by z (see FIGURE 5). His argument is general: "Assume any relation betwixt x and z that you please;" he then proceeded to find y. The example he used is

$$z = \frac{n}{m+n} ax^{(m+n)/n};$$

however, it will be sufficient to use $z = x^3$ to illustrate his argument.

In the diagram in FIGURE 5, the auxiliary line bd is chosen so that $Bb = o$, where o is not zero. Newton then specified that $BK = v$ should be chosen so that area $BbHK$ = area $BbdD$. Thus ov = area $BbdD$. Now, as x increases to $x + o$, the change in the area z is given by
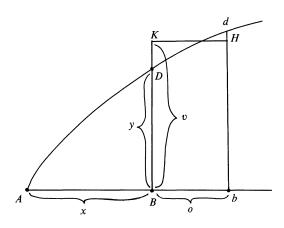

FIGURE 5

$$z(x+o) - z(x) = x^3 + 3x^2 o + 3xo^2 + o^3 - x^3 = 3x^2 o + 3xo^2 + o^3,$$

which, by the definition of v, is equal to ov. Now since $3x^2 o + 3xo^2 + o^3 = ov$, dividing by o produces $3x^2 + 3ox + o^2 = v$. Now, said Newton, "If we suppose Bb to be diminished infinitely and to vanish, or o to be nothing, v and y in that case will be equal and the terms which are multiplied by o will vanish: so that there will remain..."

$$3x^2 = y.$$

What has he shown? Since $(z(x+o) - z(x))/o$ is the rate at which the area z changes, that rate is given by the ordinate y. Moreover, we recognize that $3x^2$ would be the slope of the tangent to the curve $z = x^3$. Newton went on to say that the argument can be reversed; thus the converse holds too. We see that derivatives are fundamentally involved in areas as well as tangents, so the concept of derivative helps us to see that these two problems are mutually inverse. Leibniz gave analogous arguments on this same point (see, e.g. [16, pp. 282–284]).

Newton and Leibniz did not, of course, have the last word on the concept of derivative. Though each man had the most useful properties of the concept, there were still many unanswered questions. In particular, what, exactly, is a differential quotient? Some disciples of Leibniz, notably Johann Bernoulli and his pupil the Marquis de l'Hospital, said a differential quotient was a ratio of infinitesimals; after all, that is the way it was calculated. But infinitesimals, as seventeenth-century mathematicians were well aware, do not obey the Archimedean axiom. Since the Archimedean axiom was the basis for the Greek theory of ratios, which was, in turn, the basis of arithmetic, algebra, and geometry for seventeenth-century mathematicians, non-Archimedean objects were viewed with some suspicion. Again, what is a fluxion? Though it can be understood intuitively as a velocity, the proofs Newton gave in his 1671 *Method of Fluxions* all involved an "indefinitely small quantity o," [14, pp. 32–33] which raises many of the same problems that the o which "vanishes" raised in the Newtonian example of 1669 we saw above. In particular, what is the status of that little o? Is it zero? If so, how can we divide by it? If it is not zero, aren't we making an error when we throw it away? These questions had already been posed in Newton's and Leibniz's time. To avoid such problems, Newton said in 1687 that quantities defined in the way that $3x^2$ was defined in our example were the *limit* of the ratio of vanishing increments. This sounds good, but Newton's understanding of the term "limit" was not ours. Newton in his *Principia* (1687) described limits as "ultimate ratios"—that is, the value of the ratio of those vanishing quantities just when they are vanishing. He said, "Those ultimate ratios with which quantities vanish are not truly the ratios of ultimate quantities, but limits towards which the ratios of quantities decreasing without limit do always converge; and to which they approach nearer than by any given difference, but never go beyond, nor in effect attain to, till the quantities are diminished in infinitum" [15, Book I, Scholium to Lemma XI, p. 39].

Notice the phrase "but never go beyond"—so a variable cannot oscillate about its limit. By "limit" Newton seems to have had in mind "bound," and mathematicians of his time often cite the particular example of the circle as the limit of inscribed polygons. Also, Newton said, "nor... attain to, till the quantities are diminished in infinitum." This raises a central issue: it was often asked whether a variable quantity ever actually reached its limit. If it did not, wasn't there an error? Newton did not help clarify this when he stated as a theorem that "Quantities and the ratios of quantities which in any finite time converge continually to equality, and before the end of that time approach nearer to each other than by any given difference, become ultimately equal" [15, Book I, Lemma I, p. 29]. What does "become ultimately equal" mean? It was not really clear in the eighteenth century, let alone the seventeenth.

In 1734, George Berkeley, Bishop of Cloyne, attacked the calculus on precisely this point. Scientists, he said, attack religion for being unreasonable; well, let them improve their own reasoning first. A quantity is either zero or not; there is nothing in between. And Berkeley characterized the mathematicians of his time as men "rather accustomed to compute, than to think" [2].

Perhaps Berkeley was right, but most mathematicians were not greatly concerned. The concepts of differential quotient and integral, concepts made more effective by Leibniz's notation and by the Fundamental Theorem, had enormous power. For eighteenth-century mathematicians, especially those on the Continent where the greatest achievements occurred, it was enough that the concepts of the calculus were understood sufficiently well to be applied to solve a large number of problems, both in mathematics and in physics. So, we come to our third stage: *exploration and development.*

Differential equations, Taylor series, and functions

Newton had stated his three laws of motion in words, and derived his physics from those laws by means of synthetic geometry [15]. Newton's second law stated: *"The change of motion [our 'momentum'] is proportional to the motive force impressed, and is made in the direction of the [straight] line in which that force is impressed"* [15, p. 13]. Once translated into the language of the calculus, this law provided physicists with an instrument of physical discovery of tremendous power—because of the power of the concept of the derivative.

To illustrate, if F is force and x distance (so $m\dot{x}$ is momentum and, for constant mass, $m\ddot{x}$ the rate of change of momentum), then Newton's second law takes the form $F = m\ddot{x}$. Hooke's law of elasticity (when an elastic body is distorted the restoring force is proportional to the distance [in the opposite direction] of the distortion) takes the algebraic form $F = -kx$. By equating these expressions for force, Euler in 1739 could easily both state and solve the differential equation $m\ddot{x} + kx = 0$ which describes the motion of a vibrating spring [10, p. 482]. It was mathematically surprising, and physically interesting, that the solution to that differential equation involves sines and cosines.

An analogous, but considerably more sophisticated problem, was the statement and solution of the partial differential equation for the vibrating string. In modern notation, this is

$$\frac{\partial^2 y}{\partial t^2} = \frac{T \partial^2 y}{\mu \partial x^2},$$

where T is the tension in the string and μ is its mass per unit length. The question of how the solutions to this partial differential equation behaved was investigated by such men as d'Alembert, Daniel Bernoulli, and Leonhard Euler, and led to extensive discussions about the nature of continuity, and to an expansion of the notion of function from formulas to more general dependence relations [10, pp. 502–514], [16, pp. 367–368]. Discussions surrounding the problem of the vibrating string illustrate the unexpected ways that discoveries in mathematics and physics can interact ([16, pp. 351–368] has good selections from the original papers). Numerous other examples could be cited, from the use of infinite-series approximations in celestial mechanics to the dynamics of rigid bodies, to show that by the mid-eighteenth century the differential equation had become the most useful mathematical tool in the history of physics.

Another useful tool was the Taylor series, developed in part to help solve differential equations. In 1715, Brook Taylor, arguing from the properties of finite differences, wrote an equation expressing what we would write as $f(x + h)$ in terms of $f(x)$ and its quotients of differences of various orders. He then let the differences get small, passed to the limit, and gave the formula that still bears his name: the Taylor series. (Actually, James Gregory and Newton had anticipated this discovery, but Taylor's work was more directly influential.) The importance of this property of derivatives was soon recognized, notably by Colin Maclaurin (who has a special case of it named after him), by Euler, and by Joseph-Louis Lagrange. In their hands, the Taylor series became a powerful tool in studying functions and in approximating the solution of equations.

But beyond this, the study of Taylor series provided new insights into the nature of the derivative. In 1755, Euler, in his study of power series, had said that for any power series,

$$a + bx + cx^2 + dx^3 + \cdots,$$

one could find x sufficiently small so that if one broke off the series after some particular

term—say x^2—the x^2 term would exceed, in absolute value, the sum of the entire remainder of the series [6, section 122]. Though Euler did not prove this—he must have thought it obvious since he usually worked with series with finite coefficients—he applied it to great advantage. For instance, he could use it to analyze the nature of maxima and minima. Consider, for definiteness, the case of maxima. If $f(x)$ is a relative maximum, then by definition, for small h,

$$f(x - h) < f(x) \quad \text{and} \quad f(x + h) < f(x).$$

Taylor's theorem gives, for these inequalities,

$$f(x - h) = f(x) - h\frac{df(x)}{dx} + h^2\frac{d^2f(x)}{dx^2} - \cdots < f(x) \tag{3}$$

$$f(x + h) = f(x) + h\frac{df(x)}{dx} + h^2\frac{d^2f(x)}{dx^2} + \cdots < f(x). \tag{4}$$

Now if h is so small that $h\, df(x)/dx$ dominates the rest of the terms, the only way that both of the inequalities (3) and (4) can be satisfied is for $df(x)/dx$ to be zero. Thus the differential quotient is zero for a relative maximum. Furthermore, Euler argued, since h^2 is always positive, if $d^2f(x)/dx^2 \neq 0$, the only way both inequalities can be satisfied is for $d^2f(x)/dx^2$ to be negative. This is because the h^2 term dominates the rest of the series—unless $d^2f(x)/dx^2$ is itself zero, in which case we must go on and think about even higher-order differential quotients. This analysis, first given and demonstrated geometrically by Maclaurin, was worked out in full analytic detail by Euler [6, sections 253–254], [9, pp. 117–118]. It is typical of Euler's ability to choose computations that produce insight into fundamental concepts. It assumes, of course, that the function in question has a Taylor series, an assumption which Euler made without proof for many functions; it assumes also that the function is uniquely the sum of its Taylor series, which Euler took for granted. Nevertheless, this analysis is a beautiful example of the exploration and development of the concept of the differential quotient of first, second, and nth orders—a development which completely solves the problem of characterizing maxima and minima, a problem which goes back to the Greeks.

Lagrange and the derivative as a function

Though Euler did a good job analyzing maxima and minima, he brought little further understanding of the nature of the differential quotient. The new importance given to Taylor series meant that one had to be concerned not only about first and second differential quotients, but about differential quotients of any order.

The first person to take these questions seriously was Lagrange. In the 1770's, Lagrange was impressed with what Euler had been able to achieve by Taylor-series manipulations with differential quotients, but Lagrange soon became concerned about the logical inadequacy of all the existing justifications for the calculus. In particular, Lagrange wrote in 1797 that the Newtonian limit-concept was not clear enough to be the foundation for a branch of mathematics. Moreover, in not allowing variables to surpass their limits, Lagrange thought the limit-concept too restrictive. Instead, he said, the calculus should be reduced to algebra, a subject whose foundations in the eighteenth century were generally thought to be sound [11, pp. 15–16].

The algebra Lagrange had in mind was what he called the algebra of infinite series, because Lagrange was convinced that infinite series were part of algebra. Just as arithmetic deals with infinite decimal fractions without ceasing to be arithmetic, Lagrange thought, so algebra deals with infinite algebraic expressions without ceasing to be algebra. Lagrange believed that expanding $f(x + h)$ into a power series in h was always an algebraic process. It is obviously algebraic when one turns $1/(1 - x)$ into a power series by dividing. And Euler had found, by manipulating formulas, infinite power-series expansions for functions like $\sin x, \cos x, e^x$. If functions like those have power-series expansions, perhaps everything could be reduced to algebra. Euler, in his book *Introduction to the analysis of the infinite* (*Introductio in analysin infinitorum*, 1748), had studied infinite series, infinite products, and infinite continued fractions by what he thought of as purely

algebraic methods. For instance, he converted infinite series into infinite products by treating a series as a very long polynomial. Euler thought that this work was purely algebraic, and—what is crucial here—Lagrange also thought Euler's methods were purely algebraic. So Lagrange tried to make the calculus rigorous by reducing it to the algebra of infinite series.

Lagrange stated in 1797, and thought he had proved, that any function (that is, any analytic expression, finite or infinite) had a power-series expansion:

$$f(x+h) = f(x) + p(x)h + q(x)h^2 + r(x)h^3 + \cdots, \tag{5}$$

except, possibly, for a finite number of isolated values of x. He then defined a new function, the coefficient of the linear term in h which is $p(x)$ in the expansion shown in (5)) and called it the **first derived function** of $f(x)$. Lagrange's term "derived function" (*fonction dérivée*) is the origin of our term "derivative." Lagrange introduced a new notation, $f'(x)$, for that function. He defined $f''(x)$ to be the first derived function of $f'(x)$, and so on, recursively. Finally, using these definitions, he proved that, in the expansion (5) above, $q(x) = f''(x)/2, r(x) = f'''(x)/6$, and so on [**11**, chapter 2].

What was new about Lagrange's definition? The concept of *function*—whether simply an algebraic expression (possibly infinite) or, more generally, any dependence relation—helps free the concept of derivative from the earlier ill-defined notions. Newton's explanation of a fluxion as a rate of change appeared to involve the concept of motion in mathematics; moreover, a fluxion seemed to be a different kind of object than the flowing quantity whose fluxion it was. For Leibniz, the differential quotient had been the quotient of vanishingly small differences; the second differential quotient, of even smaller differences. Bishop Berkeley, in his attack on the calculus, had made fun of these earlier concepts, calling vanishing increments "ghosts of departed quantities" [**2**, section 35]. But since, for Lagrange, the derivative was a function, it was now the same sort of object as the original function. The second derivative is precisely the same sort of object as the first derivative; even the nth derivative is simply another function, defined as the coefficient of h in the Taylor series for $f^{(n-1)}(x+h)$. Lagrange's notation $f'(x)$ was designed precisely to make this point.

We cannot fully accept Lagrange's definition of the derivative, since it assumes that every differentiable function is the sum of a Taylor series and thus has infinitely many derivatives. Nevertheless, that definition led Lagrange to a number of important properties of the derivative. He used his definition together with Euler's criterion for using truncated power series in approximations to give a most useful characterization of the derivative of a function [**9**, p. 116, pp. 118–121]:

$$f(x+h) = f(x) + hf'(x) + hH, \text{ where } H \text{ goes to zero with } h.$$
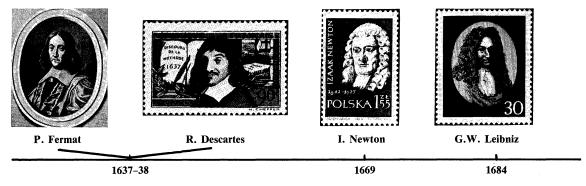
(I call this the *Lagrange property of the derivative*.) Lagrange interpreted the phrase "H goes to zero with h" in terms of inequalities. That is, he wrote that,

$$\text{Given } D, h \text{ can be chosen so that } f(x+h) - f(x) \atop \text{lies between } h(f'(x) - D) \text{ and } h(f'(x) + D). \tag{6}$$

Formula (6) is recognizably close to the modern delta-epsilon definition of the derivative.

Lagrange used inequality (6) to prove theorems. For instance, he proved that a function with positive derivative on an interval is increasing there, and used that theorem to derive the Lagrange remainder of the Taylor series [**9**, pp. 122–127], [**11**, pp. 78–85]. Furthermore, he said, considerations like inequality (6) are what make possible applications of the differential calculus to a whole range of problems in mechanics, in geometry, and, as we have described, the problem of maxima and minima (which Lagrange solved using the Taylor series remainder which bears his name [**11**, pp. 233–237]).

In Lagrange's 1797 work, then, the derivative is defined by its position in the Taylor series—a strange definition to us. But the derivative is also *described* as satisfying what we recognize as the appropriate delta-epsilon inequality, and Lagrange applied this inequality and its nth-order analogue, the Lagrange remainder, to solve problems about tangents, orders of contact between

| P. Fermat | R. Descartes | I. Newton | G.W. Leibniz |

1637–38   1669   1684

Dates refer to these mathematician's major works which

curves, and extrema. Here the derivative was clearly a function, rather than a ratio or a speed.

Still, it is a lot to assume that a function has a Taylor series if one wants to define only *one* derivative. Further, Lagrange was wrong about the algebra of infinite series. As Cauchy pointed out in 1821, the algebra of finite quantities cannot automatically be extended to infinite processes. And, as Cauchy also pointed out, manipulating Taylor series is not foolproof. For instance, e^{-1/x^2} has a zero Taylor series about $x = 0$, but the function is not identically zero. For these reasons, Cauchy rejected Lagrange's definition of derivative and substituted his own.

Definitions, rigor, and proofs

Now we come to the last stage in our chronological list: *definition*. In 1823, Cauchy defined the derivative of $f(x)$ as the limit, when it exists, of the quotient of differences $(f(x + h) - f(x))/h$ as h goes to zero [4, pp. 22–23]. But Cauchy understood "limit" differently than had his predecessors. Cauchy entirely avoided the question of whether a variable ever reached its limit; he just didn't discuss it. Also, knowing an absolute value when he saw one, Cauchy followed Simon l'Huilier and S.-F. Lacroix in abandoning the restriction that variables never surpass their limits. Finally, though Cauchy, like Newton and d'Alembert before him, gave his definition of limit in words, Cauchy's understanding of limit (most of the time, at least) was algebraic. By this, I mean that when Cauchy needed a limit property in a proof, he used the algebraic inequality-characterization of limit. Cauchy's proof of the mean value theorem for derivatives illustrates this. First he proved a theorem which states: *if $f(x)$ is continuous on $[x, x + a]$, then*
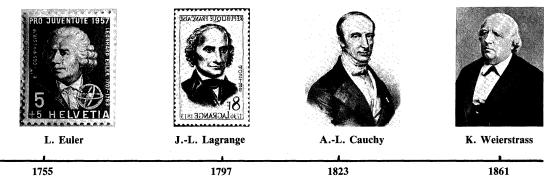
$$\min_{[x, x+a]} f'(x) \leqslant \frac{f(x+a) - f(x)}{a} \leqslant \max_{[x, x+a]} f'(x). \tag{7}$$

The first step in his proof is [4, p. 44]:

> Let δ, ε be two very small numbers; the first is chosen so that for all [absolute] values of h less than δ, and for any value of x [on the given interval], the ratio $(f(x + h) - f(x))/h$ will always be greater than $f'(x) - \varepsilon$ and less than $f'(x) + \varepsilon$.

(The notation in this quote is Cauchy's, except that I have substituted h for the i he used for the increment.) Assuming the intermediate-value theorem for continuous functions, which Cauchy had proved in 1821, the mean-value theorem is an easy corollary of (7) [4, pp. 44–45], [9, pp. 168–170].

Cauchy took the inequality-characterization of the derivative from Lagrange (possibly via an 1806 paper of A.-M. Ampère [9, pp. 127–132]). But Cauchy made that characterization into a definition of derivative. Cauchy also took from Lagrange the name derivative and the notation $f'(x)$, emphasizing the functional nature of the derivative. And, as I have shown in detail elsewhere [9, chapter 5], Cauchy adapted and improved Lagrange's inequality proof-methods to prove results like the mean-value theorem, proof-methods now justified by Cauchy's definition of derivative.

L. Euler	J.-L. Lagrange	A.-L. Cauchy	K. Weierstrass
1755	1797	1823	1861

contributed to the evolution of the concept of the derivative.

But of course, with the new and more rigorous definition, Cauchy went far beyond Lagrange. For instance, using his concept of limit to define the integral as the limit of sums, Cauchy made a good first approximation to a real proof of the Fundamental Theorem of Calculus [9, pp. 171–175], [4, pp. 122–125, 151–152]. And it was Cauchy who not only raised the question, but gave the first proof, of the existence of a solution to a differential equation [9, pp. 158–159].

After Cauchy, the calculus itself was viewed differently. It was seen as a rigorous subject, with good definitions and with theorems whose proofs were based on those definitions, rather than merely as a set of powerful methods. Not only did Cauchy's new rigor establish the earlier results on a firm foundation, but it also provided a framework for a wealth of new results, some of which could not even be formulated before Cauchy's work.

Of course, Cauchy did not himself solve all the problems occasioned by his work. In particular, Cauchy's definition of the derivative suffers from one deficiency of which he was unaware. Given an ε, he chose a δ which he assumed would work for any x. That is, he assumed that the quotient of differences converged uniformly to its limit. It was not until the 1840's that G. G. Stokes, V. Seidel, K. Weierstrass, and Cauchy himself worked out the distinction between convergence and uniform convergence. After all, in order to make this distinction, one first needs a clear and algebraic understanding of what a limit is—the understanding Cauchy himself had provided.

In the 1850's, Karl Weierstrass began to lecture at the University of Berlin. In his lectures, Weierstrass made algebraic inequalities replace words in theorems in analysis, and used his own clear distinction between pointwise and uniform convergence along with Cauchy's delta-epsilon techniques to present a systematic and thoroughly rigorous treatment of the calculus. Though Weierstrass did not publish his lectures, his students—H. A. Schwartz, G. Mittag-Leffler, E. Heine, S. Pincherle, Sonya Kowalevsky, Georg Cantor, to name a few—disseminated Weierstrassian rigor to the mathematical centers of Europe. Thus although our modern delta-epsilon definition of derivative cannot be quoted from the *works* of Weierstrass, it is in fact the *work* of Weierstrass [3, pp. 284–287]. The rigorous understanding brought to the concept of the derivative by Weierstrass is signaled by his publication in 1872 of an example of an everywhere continuous, nowhere differentiable function. This is a far cry from merely acknowledging that derivatives might not always exist, and the example shows a complete mastery of the concepts of derivative, limit, and existence of limit [3, p. 285].

Historical development versus textbook exposition

The span of time from Fermat to Weierstrass is over two hundred years. How did the concept of derivative develop? Fermat implicitly used it; Newton and Liebniz discovered it; Taylor, Euler, Maclaurin developed it; Lagrange named and characterized it; and only at the end of this long period of development did Cauchy and Weierstrass define it. This is certainly a complete reversal of the usual order of textbook exposition in mathematics, where one starts with a definition, then explores some results, and only then suggests applications.

This point is important for the teacher of mathematics: the historical order of development of the derivative is the reverse of the usual order of textbook exposition. Knowing the history helps us as we teach about derivatives. We should put ourselves where mathematicians were before Fermat, and where our beginning students are now—back on the other side, before we had any concept of derivative, and also before we knew the many uses of derivatives. Seeing the historical origins of a concept helps motivate the concept, which we—along with Newton and Leibniz—want for the problems it helps to solve. Knowing the historical order also helps to motivate the rigorous definition—which we, like Cauchy and Weierstrass, want in order to justify the uses of the derivative, and to show precisely when derivatives exist and when they do not. We need to remember that the rigorous definition is often the end, rather than the beginning, of a subject.

The real historical development of mathematics—the order of discovery—reveals the creative mathematician at work, and it is creation that makes doing mathematics so exciting. The order of exposition, on the other hand, is what gives mathematics its characteristic logical structure and its incomparable deductive certainty. Unfortunately, once the classic exposition has been given, the order of discovery is often forgotten. The task of the historian is to recapture the order of discovery: not as we think it might have been, not as we think it should have been, but as it really was. And this is the purpose of the story we have just told of the derivative from Fermat to Weierstrass.

References

[1]   Margaret Baron, Origins of the Infinitesimal Calculus, Pergamon, Oxford, 1969.

[2]   George Berkeley, The Analyst, or a Discourse Addressed to an Infidel Mathematician, 1734. In A. A. Luce and T. R. Jessop, eds., The Works of George Berkeley, Nelson, London, 1951 (some excerpts appear in [**16**, pp. 333–338]).

[3]   Carl Boyer, History of the Calculus and Its Conceptual Development, Dover, New York, 1959.

[4]   A.-L. Cauchy, Résumé des leçons données à l'école royale polytechnique sur le calcul infinitésimal, Paris, 1823. In Oeuvres complètes d'Augustin Cauchy, Gauthier-Villars, Paris, 1882- , series 2, vol. 4.

[5]   Pierre Dugac, Fondements d'analyse, in J. Dieudonné, Abrégé d'histoire des mathématiques, 1700–1900, 2 vols., Hermann, Paris, 1978.

[6]   Leonhard Euler, Institutiones calculi differentialis, St. Petersburg, 1755. In Operia omnia, Teubner, Leipzig, Berlin, and Zurich, 1911- , series 1, vol. 10.

[7]   Pierre Fermat, Analysis ad refractiones, 1661. In Oeuvres de Fermat, ed., C. Henry and P. Tannery, 4 vols., Paris, 1891–1912; Supplement, ed. C. de Waard, Paris, 1922, vol. 1, pp. 170–172.

[8]   _____, Methodus ad disquirendam maximam et minimam et de tangentibus linearum curvarum, Oeuvres, vol. 1, pp. 133–136. Excerpted in English in [**16**, pp. 222–225].

[9]   Judith V. Grabiner, The Origins of Cauchy's Rigorous Calculus, M. I. T. Press, Cambridge and London, 1981.

[10]   Morris Kline, Mathematical Thought from Ancient to Modern Times, Oxford, New York, 1972.

[11]   J.-L. Lagrange, Théorie des fonctions analytiques, Paris, 2nd edition, 1813. In Oeuvres de Lagrange, ed. M. Serret, Gauthier-Villars, Paris, 1867–1892, vol. 9.

[12]   Michael S. Mahoney, The Mathematical Career of Pierre de Fermat, 1601–1665, Princeton University Press, Princeton, 1973.

[13]   Isaac Newton, Of Analysis by Equations of an Infinite Number of Terms [1669], in D. T. Whiteside, ed., Mathematical Works of Isaac Newton, Johnson, New York and London, 1964, vol. 1, pp. 3–25.

[14]   _____, Method of Fluxions [1671], in D. T. Whiteside, ed., Mathematical Works of Isaac Newton, vol. 1, pp. 29–139.

[15]   _____, Mathematical Principles of Natural Philosophy, tr. A. Motte, ed. F. Cajori, University of California Press, Berkeley, 1934.

[16]   D. J. Struik, Source Book in Mathematics, 1200–1800, Harvard University Press, Cambridge, MA, 1969.

[17]   D. T. Whiteside, ed., The Mathematical Papers of Isaac Newton, Cambridge University Press, 1967–1982.