EDITED BY MELVIN HENRIKSEN AND STAN WAGON

The Derivative á la Carathéodory

STEPHEN KUHN
Department of Mathematics, University of Tennessee, Chattanooga, TN 37403

1. Introduction. Augustin-Louis Cauchy would be pleased. Each year we introduce our elementary analysis students to the notion of the derivative essentially as he gave it to us in 1823 (for details see, e.g. [**4**, **7**]). But there is another, less well known, characterization of the derivative which appears in the last textbook [**2**] written by Constantin Carathéodory (1873–1950). This formulation is not only elegant but useful on both theoretical and pedagogical grounds. The proofs of many important theorems concerning differentiability become significantly easier and, in the process, some of the less enlightening details are rightly submerged.

Carathéodory's formulation also gives us a much clearer view of the fact that continuity is essential for differentiability; indeed, the definition itself contains the necessary continuity. This formulation, which shifts the details from the theory of limits to the theory of continuous functions, requires that our students develop a clearer understanding of continuity than is typical and it demands that we reevaluate this understanding and continually reinforce it.

We believe that Carathéodory's insight deserves to be better known and we hope that the present article will help in that effort. We also believe, however, that since it is somewhat more subtle and less practical from a computational point of view than is the standard definition, Carathéodory's approach should be neither the first nor the only one that students see. But we also believe that all elementary analysis students can appreciate its power and should be exposed to it; in fact even the more perceptive first-year calculus students have something to gain from working with it.

2. Carathéodory's derivative formulation. After the usual definitions and theorems about limits and continuity are presented in a standard elementary real analysis course, the definition of the derivative, essentially as given to us by Cauchy, is given for functions of a single variable. Typically it is presented in both the following forms:

Definition. The value of the derivative of the function f at a is the number

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h},$$

if the limit exists; equivalently,

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

if the limit exists.

Carathéodory realized the importance of the fact that the criterion for continuity is apparent in the second definition above and he proposed the following definition for the derivative at a point; see [**2**, p. 119] or [**8**, p. 58].

CARATHÉODORY'S DEFINITION OF THE DERIVATIVE. The function f, defined on the open interval U, is said to be differentiable at the point $a \in u$ if there exists a function ϕ_a that is continuous at $x = a$ and satisfies the relation $f(x) - f(a) = \phi_a(x)(x - a)$ for all $x \in U$.

We will usually write $\phi(x)$ instead of $\phi_a(x)$, since there seems to be little chance of confusion, but we must remember that the function ϕ depends on the point a. Geometrically, of course, when $x \neq a$, $\phi_a(x)$ is the slope of the secant line through the points $(x, f(x))$ and $(a, f(a))$. This alternative definition emphasizes the fact that the slopes of the secant lines, by way of which we initially arrive at the tangent line, approach the tangent line in a continuous manner; we rarely state this important fact explicitly, but we should.

Clearly all of the subtlety of the limit in the standard definition of the derivative is now crouching just below the surface, hiding in the subtlety of the continuity of the function ϕ.

This definition can be used as is for complex-valued functions of a complex variable; in fact, it is in precisely this context that Carathéodory introduced it in [**2**]. It is interesting to note, however, that this formulation does not appear in Carathéodory's 1918 text [**3**] on real-valued functions.

As pointed out in [**2**] and [**8**], two immediate consequences of this formulation are:

- If f has a derivative at a, then f is continuous at a.
- There is at most one function ϕ satisfying the definition; so if $f'(a)$ exists, $f'(a) = \phi(a)$.

These two corollaries of the definition are both expected and their proofs are trivial, assuming only elementary results about continuous functions. The power of Carathéodory's approach does not become apparent until we apply it to prove some deeper theorems. In the following sections we offer proofs of several standard, basic theorems which we hope will demonstrate the value of this formulation.

3. An indicative example: the chain rule. Shortly after the introduction of his definition of the derivative, Cauchy used it to prove the chain rule (though his proof was incomplete). It seems appropriate then that we do the same with Carathéodory's definition. "How incisive Carathéodory's formulation is may be seen by applying it to the proof of the chain rule for the derivative of a composite function" [**8**, p. 59]. Many of our students will appreciate the pithy elegance of this proof, which is slightly revised from [**8**].

THEOREM 1. CHAIN RULE. *If f is differentiable at the point a and g is differentiable at the point $b = f(a)$, then $h = g \circ f$ is differentiable at a and $h'(a) = g'(f(a))f'(a)$.*

Proof. Since f is differentiable at a there is a function ϕ, continuous at a and defined in an open interval V containing a, with $f(x) - f(a) = f(x) - b = \phi(x)(x - a)$ for all $x \in V$. Similarly we get a function ψ, continuous at b and defined in an open interval U containing a, for which $g(x) - g(b) = \psi(x)(x - b)$ for all $x \in U$.

Then

$$h(x) - h(a) = g(f(x)) - g(f(a)) = \psi[f(x)][f(x) - b]$$
$$= [\psi \circ f(x)]\phi(x)(x - a),$$

for all $x \in V$ with $f(x) \in U$. Since $(\psi \circ f)\phi$ is continuous at a and has the value $g'(f(a))f'(a)$ there, the proof is complete. □

4. Additional examples. Carathéodory's characterization also leads to some sharp, concise proofs of other important theorems and we hope that the examples below will demonstrate the merit of Carathéodory's approach.

THEOREM 2. INVERSE FUNCTION THEOREM. *Let f be continuous and strictly monotonic on an open interval I containing c and assume that $f'(c)$ exists and is nonzero. Then $g = f^{-1}$ is differentiable at $d = f(c)$ and $g'(d) = [f'(c)]^{-1}$.*

Proof. (Our proof here is from [**8**, p. 61].) We have $f(x) - f(c) = \phi(x)(x - d)$ for all x in I with ϕ continuous and $\phi(x) \neq 0$ for all x in I. Let V be the open interval which is the domain of g. Then $y - d = f(g(y)) - d = f(g(y)) - f(c) = \phi[g(y)][g(y) - c]$ for all y in V; thus $[g(y) - c] = (1/\phi[g(y)])(y - d)$ for all y in V. Since g is continuous on V, $1/(\phi \circ g)$ is continuous there as well, and the theorem follows. □

If Carathéodory's alternative were powerful enough to prove theorems such as the two above, but too powerful or cumbersome to handle the earliest calculational results of the subject, we would be less likely to introduce it. In fact, it also manages these earlier theorems with ease and even some grace.

THEOREM 3. LINEARITY AND THE PRODUCT AND POWER RULES. *If f and g are differentiable at a, k is a constant, and n is a natural number, then:*
i) $[kf + g]'(a) = kf'(a) + g'(a)$.
ii) $(fg)'(a) = f(a)g'(a) + g(a)f'(a)$.
iii) *the derivative of $f(x) = x^n$ at a is $f'(a) = na^{n-1}$.*

Proof. We assume that there exists ϕ, ψ, continuous at a and defined on open intervals U and V, respectively, containing a, such that $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in U$ and $g(x) - g(a) = \psi(x)(x - a)$ for all $x \in V$.
i) Then

$$[kf + g](x) - [kf + g](a) = k[f(x) - f(a)] + [g(x) - g(a)]$$
$$= k\phi(x)(x - a) + \psi(x)(x - a)$$
$$= [k\phi(x) + \psi(x)](x - a), \quad \text{for all } x \in U \cap V.$$

ii) For all $x \in U \cap V$,

$$(fg)(x) - (fg)(a) = f(x)g(x) - f(a)g(a)$$
$$= f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)$$
$$= f(x)[g(x) - g(a)] + [f(x) - f(a)]g(a)$$
$$= f(x)\psi(x)(x - a) + \phi(x)(x - a)g(a)$$
$$= [f(x)\psi(x) + \phi(x)g(a)](x - a).$$

iii) Since $x^n - a^n = [x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}](x - a)$ for all x, we have $\phi(x) = x^{n-1} + x^{n-2}a + x^{n-3}a^2 + \cdots + xa^{n-2} + a^{n-1}$ and hence $\phi(a) = na^{n-1}$. □

Finally, we easily derive another basic but very useful result.

THEOREM 4. CRITICAL POINT THEOREM. *Let f be defined on an open interval I containing a. If $f(a)$ is an extreme value then a is a critical point (i.e., $f'(a) = 0$ or $f'(a)$ does not exist).*

Proof. We only prove the theorem for $f(a)$ a maximum and we do this by assuming $f'(a)$ exists and proving that $f'(a) = 0$.

For some open subinterval U of I we have a function ϕ, continuous at a, such that $f(x) - f(a) = \phi(x)(x - a)$ for all $x \in U$. Let $c \in U$, $c < a$. Then $f(c) - f(a) = \phi(c)(c - a)$, and since $f(a)$ is a maximum in U, $f(c) \leq f(a)$. So $f(c) - f(a) \leq 0$ and $c - a < 0$; hence $\phi(c) \geq 0$. Similarly, for $d \in U$, $d > a$, we have $f(d) - f(a) \leq 0$ and so $\phi(d) \leq 0$.

Since $\phi(c) \geq 0$ for all $c \in U$, $c < a$, and $\phi(d) \leq 0$ for all $d \in U$, $d > a$, the continuity of ϕ at a forces us to conclude that $\phi(a) = 0$. □

It is surprising that so few elementary analysis texts even mention this alternate view of the derivative. Of ten undergraduate/graduate analysis texts we examined only Rosenlicht **[12]**, makes use of it, although he does not mention Carathéodory by name; he imbeds it in his proof of the chain rule for one variable and brings it out in the open for partial differentiation.

There are several analysis texts, including Protter and Morrey **[10]**, which offer $f'(a) + r(x; a)$, where $r(x; a) \to 0$ as $x \to a$, instead of $\phi_a(x)$, and use it to good advantage in the proofs of several theorems. This approach has the intended additional benefit of making transparent the linear approximation of the tangent line.

The simplification of additional proofs using this definition and the extension of it to functions of several variables would make interesting projects for enterprising analysis students.

5. Carathéodory's approach in first-year calculus? It is clear that Carathéodory's formulation demands too much of most first-year calculus students to make its presentation worthwhile. Nevertheless, it can be of value, for example, for students in honors calculus courses. The chain rule will serve as a good representative example of the relative improvements in such courses afforded by the proofs given here.

Many current standard calculus texts (**[11]** is typical in this regard) relegate the more delicate details of the proof to an appendix, which will be read by few students. At the extreme, **[5]** puts the entire proof in the appendix. At the other extreme, **[9]** gives a $1\frac{1}{4}$-page "simpler proof that is valid for many functions" and then another $1\frac{3}{4}$-page complete proof (both proofs are within the same section of the text!). Neither of these routes is necessary, however. Using Carathéodory's definition we get a very clean, understandable proof of the chain rule which can be presented to perceptive first-year students in its entirety without interrupting the flow of the text.

Of 18 standard calculus texts we examined, only one, Gillett **[6]**, refers to Carathéodory's formulation of the derivative at all (exercise 66, p. 127), but no mention is made of Carathéodory himself. There are a few calculus texts, including Boyce and DiPrima [1], which present the variation mentioned at the end of the previous section.

(Note: Young [13] contains interesting biographical information on Carathéodory, from the viewpoint of one of his students and strong admirers.)

Acknowledgement. Thanks to the referee for several helpful suggestions.

REFERENCES

1. William Boyce and Richard DiPrima, Calculus, Wiley, New York, 1988.
2. Constantin Carathéodory, Theory of Functions of a Complex Variable, vol. 1, Chelsea, New York, 1954.
3. ______, Vorlesungen über Reele Functionen, Chelsea Publishing Company, New York, 1968.
4. C. H. Edwards, Jr., The Historical Development of the Calculus, Springer-Verlag, New York, 1979.
5. Robert Ellis and Denny Gulick, Calculus with Analytic Geometry, Harcourt Brace Jovanovich, New York, 1978.
6. Philip Gillett, Calculus and Analytic Geometry, third ed., D. C. Heath, Lexington, Mass., 1988.
7. Judith V. Grabiner, The Origin of Cauchy's Rigorous Calculus, MIT Press, Cambridge, Mass., 1981.
8. Maurice Heins, Complex Function Theory, Academic Press, New York, 1968.
9. Louis Leithold, The Calculus with Analytic Geometry, fifth ed., Harper and Row, New York, 1986.
10. M. H. Protter and C. B. Morrey, A First Course in Real Analysis, Springer-Verlag, New York, 1977.
11. Edwin Purcell and Dale Varburg, Calculus with Analytic Geometry, fifth ed., Prentice-Hall, Englewood Cliffs, N.J., 1987.
12. Maxwell Rosenlicht, Introduction to Analysis, Scott Foresman, Glenview, Ill., 1968.
13. Laurence Young, Mathematicians and Their Times, North-Holland, Amsterdam, 1981.