Descartes's Double Point Method for Tangents: An Old Idea Suggests New Approaches to Calculus

R. Michael Range

Calculus has been around for a long time, evolving into a core requirement for many disciplines. Given the changing and growing clientele, over the past decades there have been extensive and continuing discussions about how to improve the teaching of calculus and the level of understanding of our students. Just remember the "calculus reform" of the 1980s, and—jumping ahead—two recent contributions by K. Stroyan [Str] and F. Quinn [Qui] that have appeared in this column of the *Notices*. Still, the basic *structure* of the introduction to calculus has hardly changed over many years. While the style of presentation has undergone many transformations and—supported by technology—graphing and numerical approaches have become more widespread, tangents and derivatives are invariably introduced via the standard approximation process that involves some version of limits. This makes it necessary to study limits, limit theorems, and continuity to lay the foundations for understanding derivatives. Even if discussed in an intuitive and nontechnical form, these topics present quite a challenge for the majority of our students. Of course limits and infinite processes cannot be avoided. They are central to the subject, and they are what distinguishes calculus and analysis from algebra, which deals only with finite processes. But are they really necessary at the very beginning?

Limits are a subtle and difficult concept that was properly understood and formalized only about 150 years after the origins of calculus. Furthermore, the typical introductory examples make it difficult to understand what is really going on. For example, calculating tangents to the parabola described by $y = x^2$ at the point (a, a^2) leads to $\lim_{x\to a} \frac{x^2-a^2}{x-a}$. Since the obvious answer $0/0$ is meaningless, one uses algebra to cancel the vanishing $x - a$, i.e., $\frac{x^2-a^2}{x-a} = x + a$ for $x \neq a$. Everyone now agrees that the limit is $2a$, the result that is obtained by plugging in $x = a$. Of course the instructor warns that this result, i.e., $\lim_{x\to a}(x + a) = 2a$, requires a proof, since we can't just use $x = a$ in a formula that was derived under the assumption that $x \neq a$. This apparently so simple matter is really quite nontrivial, and it caused a lot of difficulties already in the seventeenth century as calculus was being developed. So it should not surprise us that students still have problems. All other algebraic examples (e.g., $x^n, 1/x, \sqrt{x}$, and so on) examined in an introduction to calculus

R. Michael Range is professor of mathematics at the State University of New York at Albany. His email address is range@math.albany.edu.

DOI: http://dx.doi.org/10.1090/noti1092

follow the same pattern; i.e., the relevant limit is ultimately obtained by plugging $x = a$ into an algebraic expression. After extensive discussions of limits and limit theorems, this result is eventually justified by the *continuity* of the relevant algebraic expressions. Since continuity—visually reinforced through graphing calculators—is such an intuitively obvious property of all the natural functions encountered by the student, it is not surprising that students find it difficult to understand the *need* for limits at this stage and more often than not forget about limits in the final steps of the calculation of derivatives for standard algebraic functions.

Note that the difficulties with $0/0$ disappear if the relevant equation is written in product form $x^2 - a^2 = (x+a)(x-a)$. This latter algebraic identity holds for *all* x. So how do we justify that the value $2a$ of the factor $q(x) = x + a$ at $x = a$ is indeed the desired slope of the tangent? As was already recognized by René Descartes (1596–1650), simple algebra readily provides the answer. Given his deep understanding of algebra, it was clear to Descartes that the property that singles out the tangent line through a point P on a curve is that it intersects the curve at P with *multiplicity greater than one.*[1] (See [Des] and [vSc].) Today algebraic geometers are of course well familiar with this definition of the tangent, but it seems that it has been neglected in the teaching of calculus, deferring instead to the definition of tangents as the limiting position of secants. Returning to the parabola, Descartes's idea is implemented as follows. Let $y - a^2 = m(x - a)$ be the equation of a line through (a, a^2) with slope m. Its points of intersection with the parabola $y = x^2$ are determined by the solutions of $x^2 - a^2 - m(x - a) = 0$, which factors into

$$(x + a - m)(x - a) = 0.$$

The solution $x = a$ has multiplicity 2 if this equation takes the form $(x - a)^2 = 0$; that is, the factor $(x + a - m)$ must also have a zero at $x = a$. This occurs precisely when $m = 2a$. It doesn't get simpler than this.[2]

This elementary technique to identify tangents easily extends to polynomials and rational functions, and with a bit more work also to their local inverses and so on, and ultimately to all functions defined by algebraic expressions. Similarly, all standard differentiation formulas are verified in a straightforward manner. It is quite noteworthy that the chain rule—usually viewed as the deepest rule for differentiation—turns out to be most elementary. To summarize the main conclusion, if a is in the domain of an algebraic function f,[3] then there exists a factorization

$$f(x) - f(a) = q(x)(x - a), \tag{1}$$

where q is an algebraic function defined on the domain of f and the value $q(a)$ of q at a is the slope of that unique line through $(a, f(a))$ that intersects the graph of f with multiplicity greater than one. In other words, the value $q(a)$ is the derivative $D(f)(a)$ of f at a. Full details, and more, may be found in [Ran].

How could this algebraic approach be used in the teaching of calculus?

First of all, in the case of polynomials, the basic factorization (1), the related information about zeroes, and the notion of multiplicity of such zeroes is standard high school material. These methods thus provide a simple solution for the tangent problem for all polynomials, i.e., for a deep problem with a long history that goes back to Greek geometers over 2,000 years ago, and that was one of the principal driving forces in the development of calculus in the seventeenth century. Just one simple additional step gives the corresponding result for all *rational* functions. All the standard rules of differentiation can easily be obtained in this setting and—if desired—can be extended to root functions and more. Shouldn't high school teachers take a look at this and use it in their algebra classes?

Next, the direct algebraic approach to derivatives avoids the introduction of deep new concepts involving limits and continuity early on in a context where—as we just saw—they clearly are not necessary and may even cause confusion. If the major part of a first course in calculus ends up focusing on the mechanical aspects of differentiation anyway, primarily involving algebraic functions, shouldn't it help the students to be able to do all that without having to worry about limits?

---

[1] *Descartes actually was interested in the* normal *to a curve. For example, he considered circles that intersect an ellipse at a point P, with center on one of the axes, and noted that such a circle is* tangential *to the ellipse at P precisely when P is a double point of intersection. Algebra allowed him to identify that tangential circle, whose normal at P is then the desired normal to the ellipse. Descartes's expositor F. van Schooten explicitly constructed tangents to a parabola by Descartes's double point method.*

[2] *The details were quite a bit more complicated in the seventeenth century, apparently due to the fact that the* point-slope form *of lines was not used at that time. This required the introduction of a second (distant) point to describe lines. See [Ran] for more details.*

[3] *Critical values of a function f need to be excluded from the domain of its inverse. The condition $f'(a) \neq 0$ is needed to ensure that the inverse has an appropriate factorization and hence is differentiable at the point $b = f(a)$. For example, the appropriate domain of $g(x) = \sqrt{x}$ in calculus is the* open *set of positive numbers.*

Of course the algebraic method reaches its limits when one considers important *non*algebraic functions such as exponential and trigonometric functions. However, it still helps, as it provides motivation for and a direct approach to, continuity, thereby suggesting how to handle more general functions. In more detail, since polynomials are trivially bounded on any finite interval, if f (and hence also q) is a polynomial, the factorization (1) implies the estimate

$$|f(x) - f(a)| \leq K\,|x - a| \text{ for all } x \text{ in an interval centered at } a, \tag{2}$$

for a suitable constant K. Surely this estimate exhibits the essence of continuity, i.e., that $f(x) \to f(a)$ as $x \to a$, in a precise and much stronger form than necessary. Without any additional work (aside from giving the intuitively obvious property a special name), we thus see that every polynomial is continuous. How does this compare to the typical proof in calculus and analysis texts?

Since the factorization (1) and the *local* boundedness remain correct for all algebraic functions (as indicated earlier, this requires just a bit more work), so does the estimate (2). It therefore follows immediately that all algebraic functions are continuous as well at all points of their domains. We get all this in a precise form without any formal mention of limits. Instead, it is the estimate (2) that follows from the algebraic factorization that *motivates* the idea of continuity and—implicitly—of limits.

By combining the algebraic derivative (obtained via multiplicities) with continuity, one sees that the factor q in the factorization (1) satisfies $q(x) \to q(a)$ as $x \to a$; i.e., for $x \neq a$ one obtains

$$\frac{f(x) - f(a)}{x - a} = q(x) \to q(a) = D(f)(a) = f'(a) \text{ as } x \to a.$$

One therefore recognizes that the exact algebraic derivative can also be captured by an *approximation process*. In other words, the concept of derivative as the limit of certain difference quotients results as the culmination of Descartes's algebraic approach to derivatives. This is the critical new insight that needs to be implemented when studying nonalgebraic functions. Wouldn't it help our students to meet limits only at this point, that is, in a context where they really are needed? For example, as the derivative of $E_2(x) = 2^x$ at 0 cannot be captured by any finite familiar explicit formula, it must be described by an elusive limit $0.69314\ldots$ that eventually is identified with $\ln 2$.

Finally, the preceding discussion shows that if a is a point in the domain of the algebraic function f, then

$$f(x) - f(a) = q(x)(x - a), \quad \text{where } q \text{ is continuous at } a. \tag{3}$$

Since in this case the factor q is algebraic as well, continuity holds even in the strong version given by the estimate (2). However, if q is only assumed to be continuous in the most general sense, as expressed by $q(x) \to q(a)$ as $x \to a$, then statement (3) is clearly equivalent to the standard definition of differentiability of f at a via limits of difference quotients, with $D(f)(a) = \lim_{x\to a} q(x) = q(a)$. Should the property identified in (3) be used as the primary definition of differentiability? Just try to prove the chain rule based on this definition, and you will recognize one of its major advantages. Furthermore, this formulation relates directly to the fundamental idea that differentiability is equivalent to good local *linear* approximation: just rearrange (3) in the form

$$f(x) - [f(a) + q(a)(x - a)] = [q(x) - q(a)](x - a).$$

Lastly, this definition generalizes in a most natural way to functions and maps of several variables, allowing for an equally simple proof of the chain rule as in one variable. All of this is hardly new but unfortunately not widely known. In fact, this definition was introduced by C. Carathéodory [Car] already in the middle of the last century, and it has been used successfully by him and other authors in Germany. (See [Ran] for more details and other references.)

I hope that these remarks will convince the reader that there are alternatives to the standard limit-based introduction to calculus. Students in beginning calculus courses and in analysis courses have reacted favorably to this approach. I believe that anyone teaching calculus or involved in writing introductory calculus or analysis texts should think about the questions raised here. As for myself, I have been working on an introduction to calculus that builds upon the ideas discussed here.

References

[Car] C. Carathéodory, *Funktionentheorie*, Birkhäuser, Basel, 1950.

[Des] R. Descartes, *La Géométrie. The Geometry of René Descartes,* reprint of the 1637 Leiden edition, with English translation by D. E. Smith and M. L. Latham, Open Court Publ., Chicago, 1925, reprinted by Dover, New York, 1954.

[Qui] F. Quinn, What should students get from calculus? (And how can we provide it?), *Notices AMS* **59** (6) (2012), 839–841.

[Ran] R. M. Range, Where are limits needed in calculus? *Amer. Math. Monthly* **118** (5) (2011), 404–417.

[vSc] F. van Schooten, *Geometria à Renato Des Cartes Anno 1637 Gallicè edita*, Amsterdam 1659–1661.

[Str] K. Stroyan, Why do so many students take calculus? *Notices AMS* **58** (8) (2011), 1122–1124.