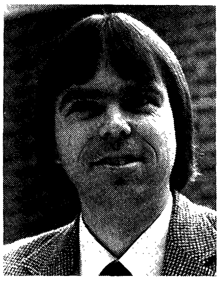
What a Tangent Line is When it isn't a Limit

Irl C. Bivens



Irl Bivens is an Associate Professor of Mathematics at Davidson College, where he has been since 1982. In 1979, he received his Ph.D. from the University of North Carolina at Chapel Hill, under the direction of Professor Patrick Eberlein. Before coming to Davidson, he taught at Pfeiffer College and was a G. C. Evans Instructor of Mathematics at Rice University. His professional interests include differential geometry, mathematical physics, and the mathematics of the undergraduate curriculum. His nonmathematical interests include science fiction, spy novels, B-movies, his wife Patti, and his son Robert.

My mathematical tutors had never shown me any reason to suppose the Calculus anything but a tissue of fallacies.

—Bertrand Russell

Does the relationship between the tangent line and the derivative express a beautiful connection between geometry and analysis? Or is this relationship a mere tautology, resulting purely from the choice of definitions? The use of the tangent line to motivate the definition of the derivative is an almost universal practice. But can one really make sense of the tangent line apart from the derivative? How is, say, the beginning calculus student supposed to know if a given line is tangent to a particular graph?

Since these are such basic questions, suppose we consider how they are dealt with in some of the current calculus textbooks. After looking through a dozen or so, we discover that there are essentially two different approaches to the question of the tangent line.

Type S: A very brief intuitive discussion of the tangent line.

Type T: An extended discussion of the tangent line, starting from an intuitive point of view and ending with a formal definition in terms of the derivative.

As we shall see, each type reflects a particular philosophy towards the notion of the tangent line.

In the *Type S* treatments, we find the purely intuitive, geometrical definitions of the tangent line. For example, the tangent line is that line which takes the "direction of the graph" at a point, or that line which gives the "limiting position" of secant lines. Yet in spite of such "definitions," there is the more or less implicit assumption that students will recognize a tangent line when they see it. In some cases, this assumption can become quite explicit.

> *We do not attempt to define the tangent line to a curve at a point. Your work in geometry has given you an intuitive grasp of this idea which will suffice.* [3]

And indeed, as long as the graphs under consideration are either concave up or concave down, students tend to find the notion of a tangent line to the graph "intuitively obvious." They will even generally accept (perhaps with some mild grumbling) that a tangent line to the graph of a linear function is the graph itself.

Experience shows, however, that tangent lines may exist only in the eye of the beholder. This often becomes manifestly apparent with the consideration of the absolute value function. To many students, it is "intuitively obvious" that the x-axis is tangent to the graph of $y = |x|$ at the origin. After all, the x-axis touches the graph in a single point, and thus, "must" be tangent there. In fact, upon further reflection, it appears that there are quite a few other tangent lines as well (Figure 1).

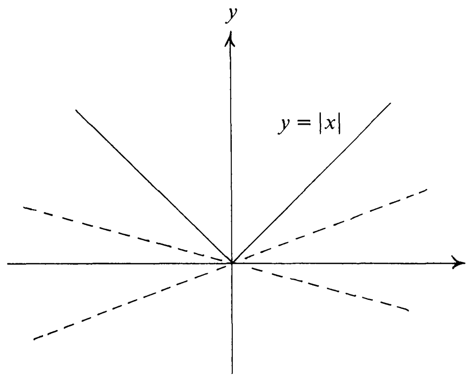


Figure 1. The graph of $y = |x|$.

The task of convincing students that their intuition has led them astray can meet with mixed success. (Following a somewhat lengthy discussion of this point, one of my students remarked: "O.K., I think I've finally got this straight. If a graph has more than one tangent line through a point, then it has no tangent lines through that point.") Likewise, students often feel that the x-axis cannot be tangent to the graph of $y = x^3$ at the origin. To them, it is "obvious" that a tangent line cannot cut across the graph at the point of tangency.

Part of the problem with the *Type S* approach is the lack of an honest definition of the tangent line. To the student, it's a more or less fuzzy notion which can be applied successfully to some examples but which leads to confusion with others. When such confusion arises, the instructor is put into the position of proving his intuition superior to that of the student. This, in turn, may be regarded by the student as the "intimidation school" of mathematics.

Suppose, then, that we consider the *Type T* treatment of the tangent line. Here, the guiding philosophy appears to be Wittgenstein's, "my propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has used them to climb out beyond them." First, the notion of the tangent line is introduced from an intuitive point of view. Then the question of the slope of the tangent line is used to motivate the limit definition of the derivative. Finally, the limit definition of the derivative is used to formally define the tangent line.

If $P(x_0, y_0)$ is a point on the graph of a function f, then the tangent line *to the graph of f at P is defined to be the line through P with slope*

$$m_{\tan} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \quad [1]$$

In effect, the tangent line is defined to be a particular limit. The somewhat circular nature of motivating the derivative via the tangent line and then formally defining the tangent line in terms of the derivative may either be ignored completely or openly acknowledged.

> *What looks like circular reasoning is really motivation for a definition. If we were merely trying to be logical, we would give the definition and be done with it . . . . But then you might justly accuse us of being arbitrary. In the end we are; every definition is arbitrary! But mathematicians are no different from other people in their desire to be understood.* [4]

From the formal point of view, there is of course absolutely nothing wrong with the above definition. (Certainly there is nothing of which Russell would disapprove.) Using it, the instructor can "prove" to the student that the x-axis is not tangent to the graph of the absolute value function. Yet, to many students this is probably a rather unsatisfactory state of affairs. The idea of the tangent line (which initially seemed so geometrical and intuitive) has suddenly receded into the shadows of abstraction. Its behavior is much like that of a faint star which, when looked at directly, disappears.

At any rate, students are likely to take small comfort from a definition of the tangent line which involves limits. Suppose, instead, that we consider a hybrid definition of the tangent line which avoids explicit use of the limit. It is a definition which is based solidly on geometrical intuition, yet which is formal enough to give insight into some fairly subtle problems (Example 5 and Theorem 3).

The Tangent Line as the Best Linear Approximation

What mental processes are going on when someone looks at a graph and quickly sketches in a tangent line? Is that person at an unconscious level actually computing the derivative or finding the "limiting position" of secant lines? (Sort of like the fellow who was surprised to learn he had been speaking prose all his life.) Probably not. Those who were asked to describe their thoughts might say something like: "I need to draw the line which 'clings' most closely to the graph near the point of tangency. A line can't be tangent if another line can be drawn which lies even closer to the graph." In other words, the person is attempting to draw the best linear approximation to the graph of the function. This interpretation is probably the one closest to most people's intuitive understanding of a tangent line. In fact, some textbooks take this point of view (although usually not too seriously) when defining the tangent line intuitively.

> *The tangent line is the line that contains P and 'best approximates' the graph of f near P.* [2]

> *The idea of a tangent to a curve at P as that line which best approximates the curve near P is better, but is still too vague for mathematical precision.* [5]

Let us base a formal definition of the tangent line upon this interpretation, starting with a point $P = (a, f(a))$ on the graph of a function $y = f(x)$. To avoid logical difficulties, let us further assume that the function is defined in some interval about $x = a$.

Definition 1. A line L through $P = (a, f(a))$ will be called a *tangent line* to the graph of f at P if L gives the best *linear* approximation to f near P. More precisely, a line $L = L(x)$ through P is a tangent line to the graph of f at P provided that for any other line $K = K(x)$ through P, there exists a $\delta > 0$ such that

$$|f(x) - L(x)| \leqslant |f(x) - K(x)| \qquad \text{for all} \quad a - \delta < x < a + \delta.$$

Remarks. (1) Although a δ has insidiously crept its way into our definition, this is not an attempt to slip half of a limit by the reader; the δ is forced on us by the fact that tangent lines are determined only by the shape of the graph near the point of tangency. In other words, the question of the existence of tangent lines is a "local" rather than "global" question. (2) Definition 1 is based upon a blend of both geometrical and analytical considerations. However, it is also possible to give purely geometrical definitions of the tangent line [6]–[8].

Since the above definition only requires us to compare pairs of lines L and K, testing conjectures about tangent lines is often a straightforward geometrical process.

Example 1. Suppose $f(x)$ is a linear function. Then it is immediately clear from the definition that there is a unique tangent line through any point on the graph of f (namely, the graph itself).

Is it true that tangent lines are unique in general? Our intuition suggests that a graph can have at most one "best" linear approximation at a point. That this is indeed the case may be seen in the following way. Suppose $L = L(x)$ and $K = K(x)$ are *distinct* tangent lines to the graph of f at $P = (a, f(a))$. Then, by definition,

$$|f(x) - L(x)| = |f(x) - K(x)|$$

in some interval about $x = a$. Consequently, $f(x) = 1/2[L(x) + K(x)]$ in this interval, and $f(x)$ is thus a linear function locally. Since tangent lines are unique for linear functions (Example 1), $L = K$. But this contradicts the assumption that L and K are distinct. Therefore, a graph can have at most one tangent line through a given point.

In order to demonstrate the usefulness of Definition 1, we shall now apply it to a number of examples. The intent is *not* to merely give a sequence of formal proofs, but rather to show that this definition together with a careful graph can lead to valid geometrical conclusions.

Example 2. Once again, consider the graph of $y = |x|$ and whether, according to our definition, it has a tangent line at the origin. Let L denote any line which might possibly be a tangent line, and compare this line L with the lines $y = x$ and $y = -x$. For points arbitrarily close to the origin, L gives a strictly worse approximation to the graph than at least one of these two lines. Consequently, the graph of the absolute value function has no tangent line at the origin.

Example 3. Does the graph (Figure 2) of

$$y = \begin{cases} x \sin(1/x), & x \neq 0 \\ 0, & \text{if} \quad x = 0 \end{cases}$$

have a tangent line at the origin? No, exactly the same argument used in Example 2 applies.

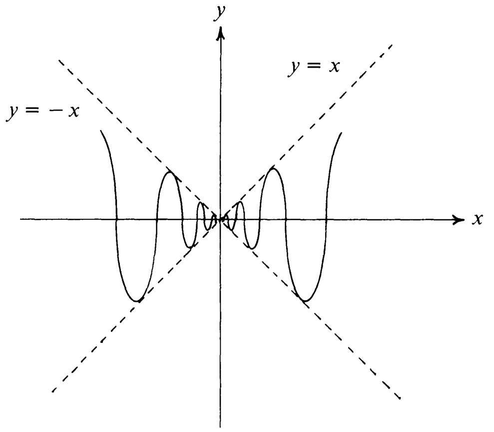


Figure 2.

Example 4. Let's now use our definition to prove that the x-axis is tangent to the graph of $y = x^3$ at the origin. Clearly, the x-axis approximates x^3 better than any line of negative slope through the origin. Therefore, suppose that $K(x) = 2mx$ denotes some arbitrary line K of positive slope $2m$. Since we want to show that in some interval about the origin, x^3 is closer to 0 than to $K(x) = 2mx$, we construct for comparison purposes the line $K_1 = K_1(x) = mx$ (Figure 3).

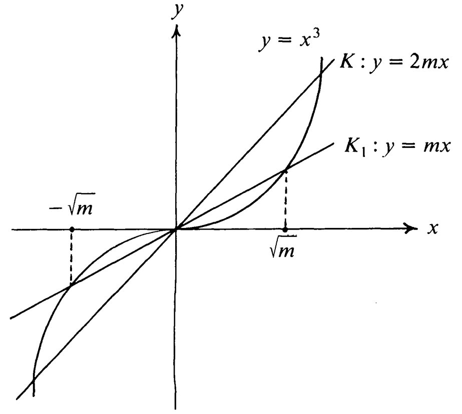


Figure 3.

From the graph, it appears that the x-axis approximates x^3 better than K in the interval $(-\sqrt{m}, \sqrt{m})$. And indeed, in this interval,

$$|x^3| = |x^2|\,|x| \leqslant |2m - x^2|\,|x| = |K(x) - x^3|,$$

which completes the proof.

Example 5. Consider the graph (Figure 4) of

$$f(x) = \begin{cases} x^2\sin(1/x), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

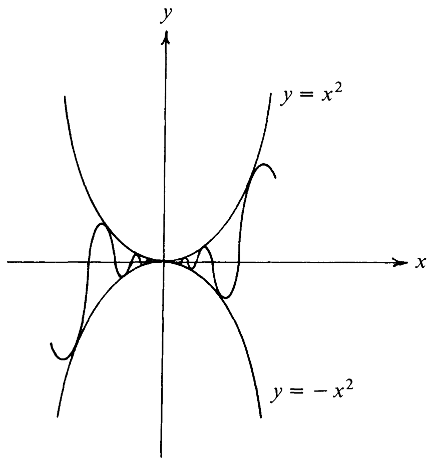


Figure 4.

Most intuitive definitions of the tangent line fare pretty badly when confronted with this example. Yet, we can show without too much difficulty that the x-axis is tangent to this graph at the origin. First, note that the x-axis is tangent at the origin to the graphs of both $y = x^2$ and $y = -x^2$. (This is either "obvious" or can be proved along the lines of Example 4.) Thus, given any line $K = K(x)$ through the origin, there exists an interval about $x = 0$ such that both x^2 and $-x^2$ are at least as close to 0 as to $K(x)$. But since $-x^2 \leqslant f(x) \leqslant x^2$, it immediately follows that in this interval $f(x)$ is at least as close to 0 as to $K(x)$. Since K was arbitrary, we conclude that the x-axis is tangent to the graph of f at the origin.

Remark. The function in Example 5 can be made still harder to visualize. As an exercise, use the above argument to show that the x-axis is tangent at the origin to the graph of

$$f(x) = \begin{cases} x^2\sin(1/x), & x \text{ is irrational} \\ -x^2, & x \text{ is rational} \end{cases}$$

As the above examples illustrate, our definition of the tangent line can be helpful even with "troublesome" graphs. The astute reader will have noticed, however, that all the tangent lines considered have been horizontal. What happens when we must deal with a nonhorizontal tangent line?

Example 6. Does the graph of $y = x^2$ have a tangent line at $(1, 1)$?

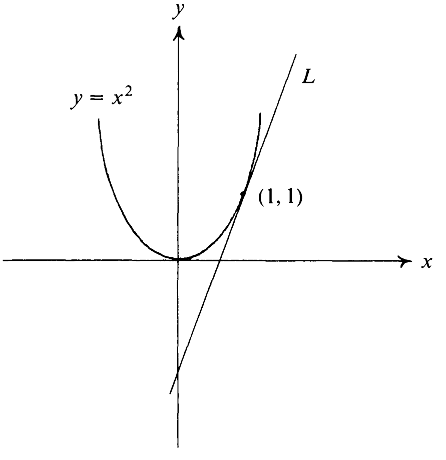


Figure 5.

Certainly from the graph (Figure 5), there appears to be a tangent line at this point. But can we actually prove this using our definition? In order to apply the definition, we first need a candidate for L. But what line should L actually be? It seems that we are now stuck.

The problem, however, does not lie with our definition, but rather with our line-guessing procedure. For a more efficient procedure, we must pursue the implications of Definition 1 a little further.

Tangent Lines and Differentiability

With most formal approaches, the existence of a tangent line and the differentiability of the function are *by definition* one and the same notion. This is not, however, the case with our definition of the tangent line; the connection with the derivative remains to be seen. Could it possibly happen that a function has a "best linear approximation" at a point, yet still not be differentiable there? That this, in fact, *cannot happen* is a consequence of the following result.

Theorem 1. *The graph of a function $y = f(x)$ has a tangent line L at $P = (a, f(a))$ if and only if $f'(a)$ exists and is equal to the slope of L.*

Proof. Assume that $f'(a)$ exists and let $L(x) = f(a) + f'(a)(x - a)$ denote the line L through $P = (a, f(a))$ having slope $f'(a)$. Denote by $K(x) = f(a) + m(x - a)$ any other line K through P having slope $m \neq f'(a)$. To prove that L is tangent to the graph of f at P, it suffices to show that $|f(x) - L(x)| \leqslant (1/2)|L(x) - K(x)|$ in some interval about $x = a$. By the definition of the derivative, there exists a $\delta > 0$ such that for all $0 < |x - a| < \delta$:

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| \leqslant \tfrac{1}{2}|f'(a) - m|.$$

But then

$$|f(x) - L(x)| \leqslant \tfrac{1}{2}|f'(a) - m|\,|x - a| = \tfrac{1}{2}|L(x) - K(x)|$$

in the interval $(a - \delta, a + \delta)$.

Suppose, conversely, that the line $L(x) = f(a) + m(x - a)$ is tangent to the graph at $(a, f(a))$. We must prove that

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = m.$$

Given $\epsilon > 0$, let $L_1(x) = f(a) + (m + \epsilon)(x - a)$ and $L_2(x) = f(a) + (m - \epsilon)(x - a)$. By our definition of tangent line, there exists $\delta > 0$ such that

$$|f(x) - L(x)| \leqslant |f(x) - L_1(x)| \quad \text{and} \quad |f(x) - L(x)| \leqslant |f(x) - L_2(x)|$$

for all $a - \delta < x < a + \delta$. If $a < x < a + \delta$, then $L_2(x) < L(x) < L_1(x)$, and thus $L_2(x) < f(x) < L_1(x)$. But these inequalities, upon subtracting $f(a)$ and dividing by $(x - a)$, are equivalent to

$$\left| \frac{f(x) - f(a)}{x - a} - m \right| < \epsilon.$$

Since this inequality also holds for $a - \delta < x < a$, we see that $f'(a)$ exists and is equal to m.

The inefficient "line guessing" procedure can now be replaced by the single step: Compute the derivative. (From our point of view, the usual formal approach substitutes a procedure for finding tangent lines in place of the actual definition of the tangent line. This is somewhat like *defining* the roots of a quadratic polynomial to be those numbers given by the quadratic formula.)

Let's return once more to a consideration of the absolute value function. Note that the graph lies completely on one side of each of the dotted lines in Figure 1. With this picture as motivation, we introduce the following.

Definition 2. A line L through point $P = (a, f(a))$ on the graph of $y = f(x)$ is *bounding at P* if the entire graph lies on one side of L. The line L is *locally bounding at P* if there exists an open interval I about $x = a$ such that L is a bounding line for f restricted to I. (Thus, all bounding lines are local bounding lines.)

What connection, if any, does there exist between tangent lines and (local) bounding lines? Does the fact that the absolute value function has more than one local bounding line through the origin have anything to do with its lack of a tangent line there? That this is indeed the case is illustrated by the next result.

Theorem 2. *If $f'(a)$ exists, then f has at most one local bounding line through $P = (a, f(a))$, and it is the tangent to the graph of f at P.*

Proof. If $f'(a)$ is defined, there exists a tangent line $L = L(x)$ to the graph of f at P. Suppose there exists a local bounding line $K = K(x)$ for f at P, and assume without loss of generality that $K(x) \leqslant f(x)$ in some interval about $x = a$. Since L is a tangent line, $K(x) \leqslant L(x)$ in some interval about $x = a$. But L and K intersect at P, and therefore K must coincide with L. To complete our proof, it suffices to observe that f need have *no* local bounding line at P—as, for example, the graph of $f(x) = x^3$ at the origin.

In particular, we can now give a correct version of my student's "nondifferentiability criterion." To wit, if a graph has more than one local bounding line through a given point, then it has no tangent lines through that point.

It is even possible to establish a partial converse to Theorem 2. What makes this converse (Theorem 3) somewhat surprising is that if the words "bounding line" are replaced by the words "local bounding line," then Theorem 3 is no longer true (Exercise 3). This is the case in spite of the fact that differentiability is a purely local phenomenon.

Theorem 3. *Let $y = f(x)$ denote a continuous function defined on some open interval. If through every point P on the graph of f there exists a unique bounding line, then f is differentiable on this interval.*

Proof. In view of Theorem 1, it suffices to show that each bounding line is also a tangent line. Suppose then that at some point P on the graph of $f(x)$ the bounding line L is *not* tangent. After composing f with suitable reflections and translating, we may assume without loss of generality that $P = (0,0)$, line L has nonnegative slope m, and that the graph of f lies on or above L. Since L is not a tangent line, there must exist a line $K = K(x)$ through the origin and a sequence of points $x_n \to 0$ such that $(x_n, f(x_n))$ and (x_n, mx_n) lie on *opposite* sides of K (Exercise 2). We will suppose the slope of K is strictly less than that of L, leaving the case of greater slope to the reader. Therefore, $K(x) = (m - \epsilon)x$ for some $\epsilon > 0$, and it follows that $x_n < 0$ and $f(x_n) > K(x_n) = (m - \epsilon)x_n$ for all n. Let m_n denote the slope of the unique bounding line $L_n(x) = m_n(x - x_n) + f(x_n)$ through $(x_n, f(x_n))$. We claim that no subsequence of $\{m_n\}_{n=1}^{\infty}$ converges to m. Suppose some subsequence does converge to m. Then, by passing to this subsequence, we may assume $m_n \to m$. But then, $|m - m_n|$ can be made arbitrarily small for large enough n. In particular,

$$L_n(0) = -m_n x_n + f(x_n) > (m - m_n - \epsilon)x_n > 0 = f(0)$$

for large n. On the other hand, the continuity of f guarantees that $(x_n, f(x_n)) \to (0,0)$ so that for large enough n

$$\begin{aligned} f(x_1) - L_n(x_1) &> (m - \epsilon)x_1 - [m_n(x_1 - x_n) + f(x_n)] \\ &= (m - m_n - \epsilon)x_1 + m_n x_n - f(x_n) > 0. \end{aligned}$$

Consequently, for large n the points $(0,0)$ and $(x_1, f(x_1))$ lie on opposite sides of L_n, a contradiction to L_n as a bounding line (Figure 6).

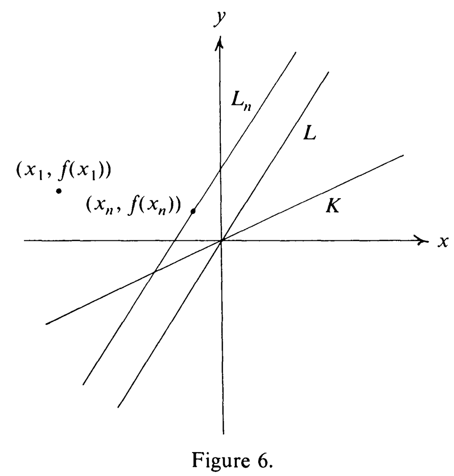


Figure 6.

This establishes our claim that no subsequence of $\{m_n\}_{n=1}^{\infty}$ converges to m. Therefore, $|m_n - m|$ must stay bounded away from 0 for large n, and it follows that $L_n \cap L = (z_n, mz_n) \to (0,0)$, since

$$z_n = \frac{f(x_n) - mx_n}{m - m_n} + x_n .$$

Because K is not a bounding line, there must exist a point $(a, f(a))$ strictly between K and L (Figure 7).

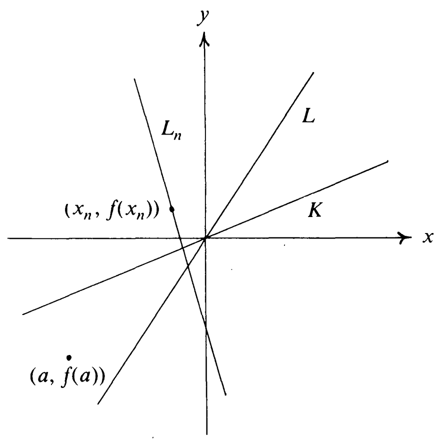


Figure 7.

Now observe that $z_n < 0$ for large n. (No positive z_n can belong to the domain of f, for then L_n would separate $(0,0)$ and points of the form $(x, f(x))$ for $x > z_n$. If z_n were equal to 0, then L_n would separate $(a, f(a))$ and points of the form $(x, f(x))$ for $x > 0$.) But then for large n, $(a, f(a))$ and $(0,0)$ lie on opposite sides of L_n, a contradiction. (The proof of this is a straightforward inequalities argument.) Therefore, the line K cannot exist and the bounding line L is, in fact, a tangent line.

Conclusions and Exercises

The interpretation of the tangent line as the "best linear approximation" can be used with profit in the beginning college calculus course. The extent to which this can actually be done, is of course, dependent upon the time available and the nature of the course. A minimal treatment would be to give Definition 1, apply it intuitively to several examples, and then motivate the derivative in the usual way. Theorem 1 could either be stated without proof or omitted completely. This would probably be an appropriate course of action for students who are seeing calculus for the first time. With students who have had some high school calculus, or with an honors calculus class, Definition 1 could be used more rigorously since such students would be more apt to appreciate the interplay between the geometry of the tangent line and the limit definition of the derivative. A proof of Theorem 1 could be developed and some of the exercises given below considered.

Whatever the approach taken, it's always worthwhile to demonstrate that definitions and concepts are not etched in stone, but are susceptible to a variety of interpretations.

The following exercises are offered for those who wish to pursue these ideas a little further.

(1) Prove directly from Definition 1 (i.e., without using Theorem 1) that if the graph of $f(x)$ has a tangent line at $(a, f(a))$, then $f(x)$ is continuous at $x = a$.

(2) Prove that $L = L(x)$ is tangent to the graph of $f(x)$ at $P = (a, f(a))$ if and only if given any line K through P there exists an interval about $x = a$ such that in this interval $(x, f(x))$ and $(x, L(x))$ lie on the same side of K.

(3) Prove the assertion made immediately prior to the statement of Theorem 3. (Hint: Construct a counterexample using the upper half of the function given in Example 3. Stretch all inflection points into line segments.)

REFERENCES

1. Howard Anton, *Calculus with Analytic Geometry (2nd ed.)*, John Wiley and Sons, 1984.
2. D. J. Foulis and M. A. Munem, *Calculus with Analytic Geometry (2nd ed.)*, Worth Publishing Inc., 1984.
3. John B. Fraleigh, *Calculus with Analytic Geometry*, Addison-Wesley, 1981.
4. Philip Gillett, *Calculus and Analytic Geometry (2nd ed.)*, D. C. Heath and Company, 1984.
5. Edwin Purcell and Dale Varberg, *Calculus with Analytic Geometry (4th ed.)*, Prentice-Hall, 1984.
6. E. L. Roetman, "Tangent Planes and Differentiation," Mathematics Magazine 43 (Jan. 1970) 1–7.
7. H. Thurston, "Tangents: An Elementary Survey," Mathematics Magazine 42 (Jan. 1969) 1–11.
8. H. Thurston, "On the Definition of the Tangent Line," American Mathematical Monthly 71 (December 1964) 1099–1103.

It is an error to believe that rigor in proof is an enemy of simplicity. On the contrary we find it confirmed by numerous examples that the rigorous method is at the same time the simpler and the more easily comprehended. The very effort for rigor forces us to find out simpler methods of proof.

DAVID HILBERT
Mathematical Problems; Bulletin American Mathematical Society, Vol. 8, p. 441

The American Mathematical Association of Two-Year Colleges will hold its annual convention in San Francisco, CA on November 13–16, 1986. For further information, contact Hal Anderson at Santa Rosa Junior College, 1501 Mendocino Ave., Santa Rosa, CA 95401.