A Geometric Approach to Calculus

Terry R. McConnell
Syracuse University

September 23, 2022

Abstract

Calculus without the limit of difference quotients

1 Introduction

This paper is a continuation of [1], which explored an approach to calculus based on a purely geometric definition of tangent line. As much as possible, it attempted to avoid using the limit concept explicitly. (Arguably, limits are implicit in any argument that uses the real numbers. Also, limits were used in [1] in the construction of the exponential function.)

In what follows, we shall use freely the definitions, notation, and results of [1].

2 Extension of the Class GT

In [1] we defined a tangent line to the graph of a function f as an affine linear function whose graph meets the graph of f at a single point, and which lies either entirely below the graph of f or entirely above it. This definition is a bit too restrictive for our present purposes, and for the rest of this paper we allow a tangent line to share more than one point with the graph of f. We continue to require that the graph of f be everywhere either above or below the tangent line. This allows, e.g., a line to be tangent to itself. We shall say that a function f is of class GT_0 if it is continuous and has a unique tangent line - in the sense just described - at each of its points. It will be convenient, in what follows, to use the word "derivative" in place of the slope of the unique tangent line.

The following theorem characterizes class GT_0 in ordinary calculus terms.

Theorem 2.1. *A function f is GT_0 if and only if it has a continuous monotone derivative.*

Theorem 3.1 of [1] gives a similar characterization of class GT, but with monotone replaced by strictly monotone. The tricky part is to show that graphs of functions in either class must lie on the same side of all their tangent lines.

Most of the proof of Theorem 3.1 can be carried over without change, but the first paragraph requires the following modification: Suppose A and B are distinct points on the graph, G, of f. We shall argue that G must lie on the same side of L_a and L_b. Suppose, to the contrary, that G lies above L_a and below L_b. If $m_a \neq m_b$ then the two lines intersect at some point C. Then C must be a point of G, since it is both above and below G. Now G must lie entirely inside the infinite wedge between the two lines extending to the left of C. But then, on the right side, the top and bottom lines swap places, and it is now impossible for G to lie below the bottom one and above the top one, since there is a definite gap between the lines.

The rest of the proof of Theorem 3.1, including the case of equal slopes, can be repeated verbatim.

We shall say that a function is of class GT_{0+} if it is of class GT_0 and if its graph lies above all of its tangent lines. Similarly, it is of class $\mathrm{GT}_{0\text{-}}$ if its graph lies below all of its tangent lines. The part of the proof of Theorem 2.1 that does not mention derivatives shows that every GT_0 function is one of the two types.

(Constant functions belong to both classes.) Examination of the proof of Theorem 3.1 of [1] also shows that the derivative of a GT_{0+} (resp. $\mathrm{GT}_{0\text{-}}$) function is non-decreasing (resp. non-increasing.)

The graph of a function in GT_{0+} is the upper envelope of its tangent lines, and thus is necessarily convex, a fact that can be used to streamline the rest of the proof of Theorem 2.1.

It is useful to be able to formulate "localized" results, so if $J = (a, b)$ we will say that a function f with domain J is GT_{0+} on J if f has a unique tangent line at each $x \in J$, and its graph lies above each tangent line over J. We also define $\mathrm{GT}_{0\text{-}}$ on J in the obvious way. A function is GT_0 on J if it is either GT_{0+} or $\mathrm{GT}_{0\text{-}}$ there. The following result allows us to extend a function belonging to one of these local classes to one which is defined on the entire real line, and allows local results to be derived from global ones. This we will do several times below without explicit comment.

Theorem 2.2. *If f is GT_0 on (a, b) and $a < c < d < b$, then there is a GT_0 function g that agrees with f on $[c, d]$.*

For the proof, assume f is GT_{0+} on (a, b). For any $c \le e \le d$, we note that the tangents L_e and L_d cross somewhere between e and d: We have $L_d(e) \le f(e) = L_e(e)$, and $L_e(d) \le f(d) = L_d(d)$. If strict inequality held in both, then $h(x) = L_d(x) - L_e(x)$ would be a continuous function satisfying $h(e) < 0$ and $h(d) > 0$. Then by the Intermediate Value Theorem on $[e, d]$, the function h vanishes at some point inside, and the tangent lines cross above this point. In any case, they always meet in at least one point above $[e, d]$.

Now define $g(x) = f(x)$ on $[c, d], g(x) = L_d(x)$ on (d, ∞), and $g(x) = L_c(x)$ on $(-\infty, c)$. From what we have just shown, no tangent L_e can meet the graph of g on (d, ∞) unless L_e and L_d coincide, since the two lines would have at least two points in common. At any rate, no such L_e can cross the graph on (d, ∞). A similar argument shows the same for $(-\infty, c)$. The tangent L_e cannot cross the graph over $[c, d]$ either, since f is GT_{0+} on (a, b).

This shows that g is GT_{0+}. Similar reasoning applies if f is assumed to be $\mathrm{GT}_{0\text{-}}$ on (a, b).

Using slopes of tangent lines in place of derivatives sometimes leads to results that are more general than their standard calculus counterparts. One such example is the Mean Value Theorem.

Theorem 2.3. *Suppose f is continuous on $[a, b]$. Then there is a point $z \in (a, b)$ at which there is a line tangent to the graph of f with slope $\dfrac{f(b) - f(a)}{b - a}$.*

Since f is not assumed to be differentiable, there may be other tangent lines at z as well.

In the special case $f(a) = f(b)$ ("Rolle's Theorem") we note that f is either constant or achieves a maximum or minimum value in (a, b). The horizontal line through the maximum or minimum point on the graph is then certainly a tangent line in our sense. The argument that deduces the Mean Value Theorem from Rolle's Theorem can then be repeated verbatim to complete the proof of Theorem 2.3.

3 Calculus

At this point we could establish many of the standard theorems of calculus, in a limited form, for GT or GT_0 functions. For example, since the sum of affine functions is affine, it is easy to show that the sum of two GT_+ functions, i.e., GT functions whose graphs lie above their tangent lines, is another such function, with the usual differentiation rule. There is little point to this exercise, as noted in the following section. Instead, we shall focus on the sine function as one last example of dealing with tangent lines "without calculus."

The following result simplifies the problem of showing that a given function belongs to GT and its variants. (It is stated for GT_+ but there are obvious modifications for the other classes we have considered.)

Theorem 3.1. *If m is continuous on (a, b) and $f(d) - f(c) > m(c)(d - c)$ for all $c \neq d \in (a, b)$, then f is GT_+ on (a, b).*

The stated inequality establishes that the line through $(c, f(c))$ of slope $m(c)$ is tangent to the graph of f. The desired result then follows from Theorem 3.2 of [1]. Note that it is essential that the same inequality hold regardless the order of c and d.

Example 3.1. *The function $f(x) = x\sqrt{1-x^2}$ is GT- on $(0,1)$ with derivative $\sqrt{1-x^2} - \frac{x^2}{\sqrt{1-x^2}}$. (A GT- function is a GT function whose graph lies below its tangent lines.)*

The function $\sqrt{1-x^2}$, whose graph is the unit circle, is an archetypal example of a GT function. The function $y = x$ is affine, hence also GT_0, so this example may be viewed as an instance of the "product rule." Using the traditional addition and subtraction trick from the standard proof of the product rule, we may write

$$d\sqrt{1-d^2} - c\sqrt{1-c^2} = \left(\sqrt{1-c^2} - \frac{d(d+c)}{\sqrt{1-d^2}+\sqrt{1-c^2}}\right)(d-c).$$

If $d > c$ we have

$$\frac{d(d+c)}{\sqrt{1-d^2}+\sqrt{1-c^2}} > \frac{c^2}{\sqrt{1-c^2}},$$

while if $d < c$ the inequality goes the other way. These facts, together with Theorem 3.1, suffice to establish the result claimed in the example.

The following result will be needed below.

Theorem 3.2. *If f and g are each GT_0 on (c,d), and $f' + g'$ is non-decreasing, then $f+g$ is GT_{0+} on (c,d).*

Let us consider the case where f is GT_{0+} and g is $\mathrm{GT}_{0\text{-}}$. The other cases are similar, or easier. Let $c < a < b < d$, and $h = f + g$. By Theorem 2.3 we know that h has a tangent line of some slope m at some point z in (a,b), but we don't know whether the graph of h lies above or below this tangent line.

Case 1: The graph of h lies above the tangent line, i.e., $h(e) - h(z) \geq m(e-z), a < e < b$.

Since $g(e) - g(z) \leq g'(z)(e-z)$, we have $(m - g'(z))(e-z) \leq h(e) - h(z) - g(e) + g(z) = f(e) - f(z)$. This shows that the line through $(z, f(z))$ with slope $m - g'(z)$ is tangent to the graph of f over (a,b). Since that tangent line is unique, we must have $m = f'(z) + g'(z)$.

Case 2: The graph of h lies below the tangent line, i.e., $h(e) - h(z) \leq m(e-z), a < e < b$.

Since $f(e) - f(z) \geq f'(z)(e-z)$, we have $(m - f'(z))(e-z) \geq h(e) - h(z) - f(e) + f(z) = g(e) - g(z)$. Again, since g also has a unique tangent line we conclude $m = f'(z) + g'(z)$.

Since $f' + g'$ is non-decreasing, we have $h(b) - h(a) = (f'(z) + g'(z))(b-a) \geq (f'(a) + g'(a))(b-a)$, and similarly when $b < a$. The desired result follows from Theorem 3.1.

We shall also need a version of the Inverse Function Theorem.

Theorem 3.3. *Let f be continuous on $[a,b]$ and GT_+ on (a,b). Assume also that f is strictly increasing on $[a,b]$, so that it defines a homeomorphism from $[a,b]$ to $[c,d]$, where $c = f(a)$ and $d = f(b)$. Let $g : [c,d] \to [a,b]$ be the inverse function. Then g is GT- on (c,d). Moreover, if $y = f(x)$, then $g'(y) = 1/f'(x)$.*

For the proof, let $a < e < b$ and $m = f'(e) > 0$. Let $c < y < d, y \neq f(e), y = f(x)$. Then

$$g(y) - g(f(e)) = x - e < \frac{1}{m}(f(x) - f(e)) = \frac{1}{m}(y - f(e)).$$

This shows that the line through $(f(e), g(f(e)))$ with slope $1/m$ is tangent to the graph of g and lies above the graph. Since $1/f'(g(y)))$ is continuous in y, the desired result follows from Theorem 3.1.

Standard developments of the integral calculus suffer from a conundrum similar to the one that affects differential calculus: does the definite integral compute the area under a curve, or define it?

Since we will need a form of the fundamental theorem of calculus, it is necessary to list a minimal set of axioms governing the area, $|A|$, of a subset A of the plane:

$$|[a,b] \times [c,d]| = (b-a)(d-c), a \leq b, c \leq d. \tag{3.1}$$

$$A \subseteq B \implies |A| \leq |B| \tag{3.2}$$

$$A \cap B = \emptyset \implies |A \cup B| = |A| + |B| \tag{3.3}$$

The existence of a function $A \to |A|$ on a suitable domain satisfying these properties is, of course, non-trivial.

Theorem 3.4. *Let $m(x)$ be decreasing, non-negative and continuous on $[a, b]$. Let $A(c)$ be the area bounded by the graphs $x = a, x = c, y = f(x)$, and $y = 0$. Then $A(x)$ is GT- on (a, b) and $m(x)$ gives the slope of its tangent line at x.*

The proof is almost immediate from Theorem 3.1: If $a < c < d < b$ then $A(d) - A(c) < |[c, d] \times [0, m(c)]| = m(c)(d - c)$, using (3.1)-(3.3). On the other hand, if $a < d < c < b$ then $A(c) - A(d) > |[d, c] \times [0, m(c)]| = m(c)(c - d)$.

Returning to the problem of finding tangent lines to the sine function, we begin with an identity that was known to Newton, Leibniz, and their contemporaries.

$$\frac{1}{2}\sin^{-1}(x) = \int_0^x \sqrt{1-t^2}\, dt - \frac{1}{2}x\sqrt{1-x^2}, 0 < x < 1.$$

(The integral here is just shorthand for the area under the graph defining the unit circle.) It is important to realize that this is a purely geometric statement: draw the radius R from the origin to the point $(\cos(x), \sin(x))$. The left side represents the area of the sector in the first quadrant of the unit circle that lies above R. On the right side, the area of the right triangle with hypotenuse R is subtracted from the area in the unit circle between the y-axis and the given value of x.

By Theorem 3.4, the first function on the right is GT on $(0, 1)$ with derivative $\sqrt{1-x^2}$. By Example 3.1, the second function is GT on $(0, 1)$ with derivative $\frac{1}{2}\sqrt{1-x^2} - \frac{1}{2}\frac{x^2}{\sqrt{1-x^2}}$. The first derivative minus the second is equal to $\frac{1}{2}\frac{1}{\sqrt{1-x^2}}$, and hence increasing on $(0, 1)$. We then conclude from Theorem 3.2 that $\sin^{-1}$ is GT$_+$ on $(0, 1)$ with derivative $\frac{1}{\sqrt{1-x^2}}$. Finally, by Theorem 3.3 and the Pythagorean identity, $\sin(x)$ is GT- on $(0, \frac{\pi}{2})$ with derivative $\cos(x)$.

4 Pedagogy

It should be possible to teach an "honest" calculus to an honors class as follows: after developing the usual theory of limits and continuity, present our definitions of GT and GT$_0$. Do the examples in the second section of [1] as far as the exponential function so that students have a repertoire of GT functions to think about. One may optionally state Theorem 2.1 with or without its proof. Next, present the following theorem and its proof:

Theorem 4.1. *If f is a GT$_0$ function, the slope, m, of its tangent line at x is given by*

$$m = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

For the proof, simply note that left and right limits give opposite inequalities, using the definition of the tangent line.

Make the point that the limit in Theorem 4.1 exists for many functions that are not of class GT$_0$. For them, we define a "tangent line" using m as slope, and call m the derivative of f at x.

Redo the examples from section 2 of [1] using this theorem, and then continue to teach the rest of differential calculus the "usual" way.

When teaching integral calculus, I feel it is vital to challenge students on their understanding of the area concept. I often ask "how would you explain the concept of area to your younger brother or sister who has no idea what it means?" Like most questions I ask in class, this one is normally greeted with hostile silence, but occasionally someone will venture a response such as "it measures how much space something takes up." If I'm feeling cranky, I might point out that they have merely substituted one vague word, "space", for another one, “area”. Some students will start listing examples – the formulas for the area of a triangle, square, circle, and so forth. Examples are fine, I tell them, to get an intuitive feel for an abstract concept, but Mathematics usually defines abstractions by listing properties we want such objects to have, for example the defining properties of tangent lines, or sets of axioms like (3.1)-(3.3) for area. In an honors class, I would go on to stress the need to prove the existence of an object having those properties. If you cannot, you are wasting your time talking about it.

References

[1] T.R. McConnell, *The Derivative as Slope of Tangent Line*, Manuscript.