Another Angle On Derivatives

Math& 148

The content here introduces you to two topics that are not well represented in common textbooks. One is how to use what is basically an idea from calculus, to study the local behavior of polynomials, without explicitly mentioning derivatives. The other is an introduction to what is known as *Taylor's Formula*, an extremely useful tool on th study and the approximation of functions, that is usually postponed to advanced calculus classes, but that need not be.

Part I. Precalculus

1 Studying Polynomials

Even though polynomials are very simple functions (for example, given an exact value for x , we can compute the exact corresponding value of the polynomial via simple products and sums), it is not instantaneous to figure out how the graph will look like near a given value of x , as soon as the degree is not very low. We can however find that out using a simple trick.

1.1 Polynomials near $x = 0$

This fact is well known: for a number a , such that $|a| < 1$, the higher the power n in a^n, the smaller its absolute value. This feature accentuates the closer a is to 0 . For example, if $a = 10^{-1}$,$a^n = 10^{-n}$, but if $a = 10^{-3}$, $a^n = 10^{-3n}$, so that $\frac{a^n}{a}$ goes from 10^{-n+1} to $10^{-3(n+1)}$. Hence, if if x is lose to 0 , each of its successive powers becomes quickly negligible. So, if $|x|$ is sufficiently small, x^2 will be "invisible" when compared to x . Hence, for a polynomial like

$$p(x) = 4x^7 - 2x^6 + x^5 + 3x^4 - 2x^2 - 6x + 3$$

we can say

- $p(0) = 3$ (that's easy)

- If $|x|$ is really close to 0, then $p(x) \approx 3 - 6x$. This means that the graph of p will look almost like the graph of the straight lien $3 - 6x$. Geometrically, this means that the tangent line to the graph at $x = 0$ has slope -6, and the graph crosses the $y-$axis going from above 3 to below 3, that is it is decreasing near $x = 0$, with approximate slope -6:

  poly1.jpeg

This can be pushed further. If $|x|$ is still small, but not as small, we may keep the next term, $-2x^2$, ignoring higher powers. That suggests that the graph of p, near 0, will look like the graph of $-2x^2 - 6x + 3$, an "open down" parabola, on its decreasing side, which means that the graph will be concave down near 0:

poly2.jpeg

This could be pushed even further, but you get the idea. It's even more interesting when, for example, the linear term is missing. Looking at $q(x) = 4x^4 - 3x^3 - 2x^2 + 1$, we have $q(0) = 1$, but now the first interesting term to consider when $|x|$ is close, enough to, but not equal to, 0 is $1 - 2x^2$. This is a parabola open down, with vertex at $(0, 1)$. This is a maximum point for the parabola, and this allows us to realize that this is a (local) maximum point for $q(x)$:

poly3.jpeg

1.2 Polynomials Near Any Value of x

It's nice to know about the graph near 0, but we will also be interested in the graph at any other point. This can be easily reduce to the previous case with one of two *completely equivalent* tricks:

1. We can *shift our graph horizontally so that the value we are interested in is shifted to $x = 0$*

2. When our polynomial is written as a function of x, we are looking at it as a function of the difference between our point of interest and 0. If we are interested in looking at our polynomial near another point, call it x_0, we can re-write it in terms of the difference between x and x_0.

The approach in point 1 is discussed in a file you can find at
http://faculty.washington.edu/~fm1/Materials/index.html

The second approach goes like this. Let's call, for convenience, $x - x_0 = h$, and let's re-write our polynomial as a polynomial with h as the variable. It's probably best to do this on a concrete example – it will be easy to do the same manipulation on any other polynomial. Let's say we want to look at $p(x) = 4x^4 - 3x^3 + 2x^2 - 5x + 2$ near $x_0 = 2$. Let's re-write this by setting $x = x_0 + (x - x_0) = 2 + h$:

$$4\,(2+h)^4 - 3\,(2+h)^3 + 2\,(2+h)^2 - 5\,(2+h) + 2$$

Now, since we look at this as a polynomial in h, let's open up the parentheses and put the result in descending powers of h:

$$4\left(h^4 + 8h^3 + 24h^2 + 32h + 16\right) - 3\left(h^3 + 6h^2 + 12h + 8\right) + 2\left(h^2 + 4h + 4\right) - 5(h+2) + 2 =$$

$$= 4h^4 + 29h^3 + 80h^2 + 95h + 40 \qquad (1)$$

Using the same approach we used in 1.1, since we are looking at what happens when h is close to 0, we see that the tangent line to our polynomial at $x = 2$ is $95x + 40$, and that the graph is close to the parabola $80h^2 + 95h + 40$, and open down parabola, increasing at this point, so the graph is concave up here.

1.3 How To Go Beyond Polynomials?

All the above is pure Algebra. But how could we extend this idea to functions like e^x, or $\ln x$, or even more complicated ones? Well, that was solved by the genius of Newton and Leibniz, back in the 17th Century, and is the topic most of our class. In fact, one way to look at derivatives is as coefficients of polynomials that are "close" to our function.

Part II. Calculus

2 Polynomials and *Taylor's Formula*

If you go back to our previous part, and look closer at the calculations, you will notice that the slope of the tangent to a polynomial at $x = a$ computed there is precisely $p'(a)$, as defined in our calculus book. If you look even closer, you may notice that the coefficient of the quadratic term in, for example (1) is equal to $\frac{p''(2)}{2}$! Let's look at all this in more detail.

The point in the discussion on polynomials like $p(x)$ was that

- when x is close to a, $p(x)$ is "close" to $p(a)$ – that is, $p(x)$is a continuous function
- when x is very close to a, $p(x) - p(a)$ is *almost* proportional to $x - a$ (that is the linear term we emphasized in (1))

- when x is still very close to a, but not as much, $p(x)-p(a)$ is almost equal to a quadratic function, as emphasized in (1)
- this discussion could be pushed further keeping third, fourth, .. powers

A general function can be very ugly (a simple example could be the function that takes the value 1 when x is irrational, and 0 when x is rational – but people have come up with much uglier functions). A first condition that limits the "ugliness" is continuity (when $x-a$ is small, so is $f(x)-f(a)$). Still, even if they are not very simple to describe, there are very ugly continuous functions.

To come up with "nicer" functions, we can require that they behave not too differently from polynomials around any given point. To make this precise, let's suggest the following first conditions on $f(x)$:

Condition0 $f(x)$ is continuous at $x=a$. This means that $f(x)-f(a)=E_0(x,a)$, where $\lim_{x\to a} E_0(x,a)=0$

Condition1 We assume something more about the "error term" E_0: namely that it goes to 0 a t speed at least equal to the speed of $x-a$. That means, in formulas, that $E_0(x,a)=c_1(x-a)+E_1(x,a)$, where $E_1(x,a)$ goes to 0 faster that $x-a$, or, in formulas,$\lim_{x\to a}\frac{E_1(x,a)}{x-a}=0$

This means, in turn, that

$$f(x)-f(a)=c_1(x-a)+E_1(x,a)$$

and, dividing this equality by $x-a$,

$$\frac{f(x)-f(a)}{x-a}=c_1+\frac{E_1(x,a)}{x-a}$$

Taking the limit of both sides, we see that this implies that f is differentiable at $x=a$, and that $c_1=f'(a)$. We could also argue in reverse: if we assume that f is differentiable at $x=a$, then the previous calculation applies, and $f(x)-f(a)$ is "almost proportional" to $x-a$, with $f'(a)$as the proportionality constant.

We now have that, for a differentiable function,

$$f(x)-f(a)=f'(a)(x-a)+E_1(x,a)$$

where$\lim_{x\to a}\frac{E_1(x,a)}{x-a}=0$.

Still using our study of polynomials as a guide, we might consider the case where E_1 is not only faster than $x-a$ in going to 0, but has actually the same speed as $(x-a)^2$. That is, we consider the "nicer" case, when $E_1(x,a)=c_2(x-a)^2+E_2(x,a)$ with $\lim_{x\to a}\frac{E_2(x,a)}{(x-a)^2}=0$. This implies

$$f(x)-f(a)-f'(a)(x-a)=c_2(x-a)^2+E_2(x,a) \tag{2}$$

Let's divide this equality by $(x-a)^2$:

$$\frac{f(x)-f(a)-f'(a)(x-a)}{(x-a)^2}=c_2+\frac{E_2(x,a)}{(x-a)^2}$$

Letting $x \to a$, he right hand side has limit c_2, so the left hand site has a limit as well. To determine it, we can use L'Hospital's rule and applying it results in

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{(x-a)^2} = \lim_{x \to a} \frac{f'(x) - f'(a)}{2(x-a)}$$

Hence, we see that f' is differentiable at $x = a$, and the left hand ahas limit $\frac{f''(a)}{2}$. One cal also argue in reverse, and, assuming that $f''(a)$ exists, see that (2) holds, and that $c_2 = \frac{f''(a)}{2}$. Thus, if all of this holds,

$$f(x) - f(a) = f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + E_2(x,a)$$

We don't need to stop here. If, again in analogy to polynomials, that our function is so nice that $E_2(x,a)$ vanishes not only faster than $(x-a)^2$, but actually at a speed at least as fast as $(x-a)^3$, we can write that $E_2(x,a) = c_3(x-a)^3 + E_3(x,a)$, with $\lim_{x \to a} \frac{E_3(x,a)}{(x-a)^3} = 0$, and

$$f(x) - f(a) - f'(a)(x-a) - \frac{f''(a)}{2}(x-a)^2 = c_3(x-a)^3 + E_3(x,a) \qquad (3)$$

Now, divide this equality by $(x-a)^3$:

$$\frac{f(x) - f(a) - f'(a)(x-a) - \frac{f''(a)}{2}(x-a)^2}{(x-a)^3} = c_3 + \frac{E_3(x,a)}{(x-a)^3}$$

As $x \to a$ the right hand side has limit c_3, hence the left hand side has a limit as well. To find it we can again apply L'Hospital's Rule. A first application results in

$$\frac{f'(x) - f'(a) - f''(a)(x-a)}{3(x-a)^2}$$

A second application results in

$$\frac{f''(x) - f''(a)}{6(x-a)}$$

Since this is known to have a limit, we see that (3) implies that f'' is differentiable, that is that f has at least three derivatives, and

$$c_3 = \frac{f'''(a)}{2 \cdot 3}$$

As before, if we assume that f has three derivatives we can proceed in reverse and conclude that (3) holds.

We can go on as long as f has more derivatives, following the same logic. If we repeat this procedure n times, we end up with *Taylor's Formula*:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{2 \cdot 3}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + E_n(x,a) \qquad (4)$$

where $\lim_{x \to a} \frac{E_n(x,a)}{(x-a)^n} = 0$. By the way. $n!$ is read as "n factorial", and is equal to $1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$.

3 Improving to Taylor's Theorem

(4) is a remarkable result, indicating that, if f has enough derivatives, as long as we are looking at small deviations from a point, the graph of f will look very much like that of a polynomial[1]. However, the formula, as listed, is less useful if our goal is to calculate an approximation to $f(x)$ for functions that are difficult to compute by hand. The reason is that we only know what the error term E_n does in the limit, but there is no explicit estimate of its value for fixed x and a. This limitation is solved by several forms that can be given to the error term (usually called the *remainder*), forms that do to provide its exact value (if they did, we would have an exact numerical value for $f(x)$, which we don't have, in general), but allow us to find an upper bound to the error. These require an additional derivative.

The simplest of these forms is the *Legendre form*:

$$E_n(x,a) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \tag{5}$$

where ξ is some point between a and x. We don't know the exact value of ξ, but we often can find a number M such that $\left|f^{(n+1)}(\xi)\right| \leq M$, which allows us to bound the error by

$$|E_n(x,a)| \leq \frac{M}{(n+1)!}(x-a)^{n+1}$$

Most proofs of (5) are somewhat technical (at least that is my feeling – for example one uses a clever application of the so-called Cauchy mean value theorem). One proof that is more straightforward than most is due to Zvonimir Sikic (1990) and is quickly summarized at the end (it requires some basic facts about *integral calculus*).

[1] There is a catch here: for this formula to be useful we need to have at least one derivative that is not zero. Fact is there are functions such that all their derivatives at a point are zero, and yet are not constant. The standard example is $e^{-\frac{1}{x^2}}$ at $x = 0$. If you are curious, you can look at the file "zeroderivative.pdf" linked from the same page as this.

As a simple application, consider a rough estimate for $\ln(1.1)$. We write this as $f(0.1)$ where $f(x) = \ln(1+x)$. With some patience, we can calculate

$$f(0) = 0, \quad f'(0) = 1, \quad f''(0) = -\frac{1}{2}$$

while $f'''(x) = \frac{2}{(1+x)^3}$. The third derivative is decreasing as we go from 0 to 0.1, so $\left|f^{(n+1)}(\xi)\right| \le f'''(0) = 2$ and

$$f(1.1) = 10^{-1} - \frac{1}{2} \cdot 10^{-2} + E_2$$

$$|f(1.1) - 0.095| \le 2 \cdot 10^{-3}$$

giving us an estimate for $\ln(1.1)$ correct to about three decimal digits (a computer will evaluate,, approximately, $\ln(1.1) \approx 0.0953101798043249$)

A Proof Of (5)

If we are familiar with the following two basic facts of integral calculus, we can come u with a fairly straightforward proof of (5).

1. The *Fundamental Theorem of Calculus:* $\int_a^b f'(x)dx = f(b) - f(a)$
2. The *Mean Value Theorem:* $\int_a^b g(x)h(x)dx = g(\xi)\int_a^b h(x)dx$ where ξ is a number between a and b.

Consider now a differentiable function f at points x and y. We have

$$f(x) - f(y) = E_0(x, y)$$

If $f$is differentiable, so is E_0. Deriving with respect to y,

$$f_y(y) = -E_0'(y) \tag{6}$$

(muting the dependence of E_0on x, here and in all the following, also for subsequent error terms). Integrating (6) between x an y

$$E_0(y) = -\int_x^y f_y(t)dt = -f_y(\xi)(y - x)$$

by the mean value theorem, so that

$$f(x) = f(y) + f'(\xi)(x - y) \tag{7}$$

which is (5) for $n = 0$.

We can now repeat, by considering (7) and replacing ξ with y, and the consequent error:

$$f(x) = f(y) + f_y(y)(x - y) + E_1(y) \tag{8}$$

We have $E_1(x) = 0$, and differentiating with respect to

$$0 = f'(y) + f''(y)(x - y) - f'(y) + E_1'(y)$$

so that

$$E'(y) = -f''(y)(x - y)$$

and integrating from x to y, and applying the mean value theorem,

$$E_1(y) = -\int_x^y f''(t)(x - t)dt = -f''(\xi)\int_x^y (x - t)dt = f''(\xi)\frac{(x - y)^2}{2}$$

Inserting in (8)

$$f(x) = f(y) + f'(y)(x - y) + f''(\xi)\frac{(x - y)^2}{2}$$

Repeating for further derivatives results in (4), with y in place of a in the formula.