Taylor's Theorem

Two different proofs of versions of Taylor's Theorem are given here. Each version is accompanied by an application. The number and variety of applications to choose from is very large. I also try to suggest how Taylor's Theorem might have been discovered, and give some rather surprising information about the history of the result.

Proof #1

This is essentially what was discussed in class. We begin with the harmless observation

$$f(x) - f(a) = \int_a^x f'(t)\,dt$$

which must be a version of the Fundamental Theorem of Calculus. Now we integrate by parts in a weird, wonderful, and unmotivated way.

$$\int_a^x f'(t)\,dt = (t-x)f'(t)\Big]_{t=a}^{t=x} - \int_a^x (t-x)f''(t)\,dt$$

$$\int u\,dv = uv - \int v\,du$$

$$\left.\begin{aligned} u &= f'(t) \\ dv &= dt \end{aligned}\right\} \left\{\begin{aligned} du &= f''(t)\,dt \\ v &= t-x \end{aligned}\right.$$

It is easy to get distracted because there are so many letters floating around. This choice of v is not "clear" and is similar to what was done when we derived the error estimate for the Trapezoid Rule. Look at the boundary term: $(t-x)f'(t)\Big]_{t=a}^{t=x} = (x-x)f'(x) - (a-t)f'(a) = f'(a)(x-a)$. As for the integral, push "in" the minus sign and get $f(x) - f(a) = f'(a)(x-a) + \int_a^x (x-t)f''(t)\,dt$. Usually people have the $f(x)$ term sitting alone, so this is written

$$f(x) = f(a) + f'(a)(x-a) + \int_a^x (x-t)f''(t)\,dt.$$

We need to deal with $\int_a^x (x-t)f''(t)\,dt$ which we will do by integrating by parts (again, *not* clearly!).

$$\int_a^x (x-t)f''(t)\,dt = -\frac{1}{2}(x-t)^2 f''(t)\Big]_{t=a}^{t=x} - \int_a^x -\frac{1}{2}(x-t)^2 f^{(3)}(t)\,dt$$

$$\int u\,dv = uv - \int v\,du$$

$$\left.\begin{aligned} u &= f''(t) \\ dv &= (x-t)dt \end{aligned}\right\} \left\{\begin{aligned} du &= f^{(3)}(t)\,dt \\ v &= -\frac{1}{2}(x-t)^2 \end{aligned}\right.$$

As I mentioned in class, the choice of v here is certainly *not* the simplest. In fact, maybe you should differentiate v to make sure that the resulting dv is correct. It will be, because the minus sign from inside (because of the chain rule) will cancel the minus sign on the outside, and the 2's will cancel, also. Now to evaluate this boundary term: $-\frac{1}{2}(x-t)^2 f''(t)\big]_{t=a}^{t=x} = -\frac{1}{2}(x-x)^2 f''(x) - \left(-\frac{1}{2}(x-a)^2 f''(a)\right) = \frac{f''(a)}{2}(x-a)^2$. In the integral, the two minus signs cancel. Therefore we get:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \int_a^x \frac{1}{2}(x-t)^2 f^{(3)}(t)\,dt.$$

I will do one more integration by parts step, and then I hope the pattern will be truthfully clear rather than mythically clear.

$$\int_a^x \frac{1}{2}(x-t)^2 f^{(3)}(t)\,dt = -\frac{1}{2\cdot 3}(x-t)^2 f^{(3)}(t)\big]_{t=a}^{t=x} - \int_a^x -\frac{1}{2\cdot 3}(x-t)^2 f^{(4)}(t)\,dt$$

$$\int u\,dv = uv - \int v\,du$$

$$\left.\begin{array}{r} u = f^{(3)}(t) \\ dv = \frac{1}{2}(x-t)^2 dt \end{array}\right\} \left\{\begin{array}{l} du = f^{(4)}(t)\,dt \\ v = -\frac{1}{2\cdot 3}(x-t)^2 \end{array}\right.$$

This boundary term: $-\frac{1}{2\cdot 3}(x-t)^3 f^{(3)}(t)\big]_{t=a}^{t=x} = -\frac{1}{2\cdot 3}(x-x)^3 f^{(3)}(x) - \left(-\frac{1}{2\cdot 3}(x-a)^3 f^{(3)}(a)\right) = \frac{f^{(3)}(a)}{2\cdot 3}(x-a)^3$. In the integral, again the two minus signs cancel. Therefore we get:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f^{(3)}(a)}{2\cdot 3}(x-a)^3 + \int_a^x \frac{1}{2\cdot 3}(x-t)^3 f^{(4)}(t)\,dt.$$

I hope you now can see that I did not want to write $2 \cdot 3$ as 6, but I should write it as $3!$. And that $1! = 1$ and also $0! = 1$. The formula that we are producing is one of the major reasons that $0!$ is defined to *be* 1. And $f(a)$ is the 0^{th} derivative of f evaluated at a and 1 is also $(x-a)^0$. Rewrite the previous equation as

$$f(x) = \frac{f^{(0)}(a)}{0!}(x-a)^0 + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \int_a^x \frac{(x-t)^3}{3!} f^{(4)}(t)\,dt.$$

Now we **define** the n^{th} degree Taylor polynomial of the function f centered at a to be $\displaystyle\sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j$. Again, there are many different letters, but the j in the sum is the index of summation, a "dummy index". The polynomial has degree n in x and the fractions

$\dfrac{f^{(j)}(a)}{j!}$ are the coefficients of the "shifted" monomials $(x-a)^j$. In the formula above please recognize the third degree Taylor polynomial of f centered at a. Taylor's Theorem addresses the relationship between the Taylor polynomial and the original function. The difference is usually called the *remainder term.*

Taylor's Theorem with the integral form of the remainder

$$f(x) = \sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j \;+\; \int_a^x \frac{(x-t)^j}{j!} f^{(j+1)}(t)\,dt$$

Since this looks dangerously complicated and theoretical, let me immediately illustrate with an example as I did in class. The web page accompanying this essay supports this example with many graphs and some nice numerical information.

Example: the sine function

Everything will be neat. In the Taylor polynomial, we will take a to be 0 (in fact, almost always almost everyone makes this choice!). So we need to understand $\sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j$ when a is replaced by 0, which is $\sum_{j=0}^{n} \frac{f^{(j)}(0)}{j!}x^j$. Let us compute derivatives of sine, starting with the 0^{th} derivative, the function itself. So $f^{(0)}(x) = \sin x$ and $f^{(1)}(x) = \cos x$ and $f^{(2)}(x) = -\sin x$ and $f^{(3)}(x) = -\cos x$ and after that they repeat. And the values at 0 of these functions are 0 and 1 and 0 and -1 and after that they repeat. Here is the Taylor polynomial for $\sin x$ centered at $a = 0$ of degree 9*

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!}$$

There *is* a term of every degree in the polynomial but half of the terms (the even degree terms) have coefficient 0 so they are not written.

Now look at the remainder term: $\int_a^x \frac{(x-t)^j}{j!} f^{(j+1)}(t)\,dt$. We should change a to 0. I'm really interested its magnitude, that is, the absolute value, $\left|\int_0^x \frac{(x-t)^j}{j!} f^{(j+1)}(t)\,dt\right|$. Let me suppose that $x > 0$ so I can manipulate the integral more easily. How big can this be? Well (just as in several estimates done earlier in the course) this is at most $\int_0^x \left|\frac{(x-t)^j}{j!} f^{(j+1)}(t)\right| dt$. But let's take advantage of the simple nature of sine's derivatives: they are all $\pm$sine or $\pm$cosine, and they all have absolute value *at most* 1. So the remainder is at most $\int_0^x \left|\frac{(x-t)^j}{j!}\right| dt$. (If x were negative I would have to integrate from

* It is also the Taylor polynomial of degree 10.

x to 0.) I can do a bit better than what I did in class. When x is positive, the $x-t$ term is always positive when t is in the interval from 0 to x. The absolute value signs can just be dropped, and the integral can be computed exactly. Therefore the remainder or error term will be bounded by $\frac{|x|^{j+1}}{(j+1)!}$. Similar computations with the same result can be done for negative x. We now have the following information:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots \pm (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \text{Remainder, and } \left|\text{Remainder}\right| \leq \frac{|x|^{2n+2}}{(2n+2)!}.$$

We will examine the remainder, $\frac{|x|^{2n+2}}{(2n+2)!}$, a bit. The top is an ANGEL but the bottom is an ARCHANGEL. If you are getting tired of these ideas, just think about the ratio test and how its logic would work here. Suppose $x = 2{,}000$. We're then studying the sequence $\frac{2{,}000^{2n+2}}{(2n+2)!}$. The 10^{th} term is about $3.7 \cdot 10^{51}$, which is quite big to me. And the 20^{th} term grows to more than $3.1 \cdot 10^{87}$. And ... you get the idea. *But* if you think about the structure of factorials, after the $1{,}000^{\text{th}}$ term the numbers begin decreasing (2 multiplies the n in the factorial). After the $2{,}000^{\text{th}}$ term the numbers get cut down by at least one-half each time. There are infinitely many numbers bigger than 2,000. In fact, the $10{,}000^{\text{th}}$ term is about $2.2 \cdot 10^{-11{,}318}$. The minus sign in the exponent makes the number very, very small.

We have verified three interesting related facts:

1 The infinite series $\displaystyle\sum_{j=0}^{\infty} (-1)^j \frac{x^{2j+1}}{(2j+1)!}$ converges for all x.

We could have seen this by just studying the series with the Ratio Test. But I claim that we have proved this in a more subtle way: the Taylor polynomials for the sine function *are* the partial sums for this series, and we have verified that the Remainder$\to 0$ so the partial sums *must* approach a limit!

2 We now know *exactly* the sum of the previous infinite series: the sum is $\sin x$.

This is nice. We've been studying convergence of infinite series and have all sorts of ways (the various "Tests") to convince ourselves that the series converge. We have been able to find the exact values of the sums of very few series because we don't know many formulas for partial sums.

3 The difference between the true value of $\sin x$ and the Taylor polynomials for sine centered at $a = 0$ can be estimated by $\frac{|x|^{2n+2}}{(2n+2)!}$ and as $n \to \infty$, this quantity$\to 0$ rapidly, so the Taylor polynomials provide an easy way to compute values of sine accurately.

No more measuring lots of right triangles and computing the value of OPPOSITE divided by HYPOTENUSE: throw away your triangles! You can compute values of sine directly with rational number arithmetic to any accuracy specified.

Again, the web diary contains pictures and numbers which should support the statements you've just read. I hope they will make what we've done more understandable and to give more evidence to convince you of Taylor's Theorem. You should not expect to understand many of the implications of Taylor's Theorem immediately. It is a subtle result.

Proof #2

This is a faster proof but maybe even more magical and less motivated than the previous one. I will only do a special case of the proof for the second degree Taylor polynomial.

Choose a and x and **define a constant K by the equation**

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + K\frac{(x-a)^3}{6}.$$

So K is the number which makes that equation true. If $a \neq x$, then $a - x \neq 0$ and you can solve for K in the equation, so there will be exactly one K satisfying the equation. The object of what we're doing is to *discover* another description of K. Well, now define a function $F(w)$ by the equation

$$F(w) = f(w) + f'(w)(x-w) + \frac{f''(w)}{2}(x-w)^2 + K\frac{(x-w)^3}{6}$$

and w is a variable between a and x. Then differentiate this with respect to the variable w. Be sure to differentiate *everything* that contains a w. The product rule must be used twice, and the Chain Rule gives some minus signs.

$$F'(w) = f'(w) + \overbrace{f''(w)(x-w) - f'(w)} + \overbrace{\frac{f^{(3)}(w)}{2}(x-w)^2 - \frac{f''(w)}{2}2(x-w)^2} - K\frac{3(x-w)^2}{6}.$$

The expression on the right contains both $f'(w)$ and $-f'(w)$: these cancel. And there are two more terms which also cancel: $f''(w)(x-w)$ and $-\frac{f''(w)}{2}2(x-w)^2$. $F'(w)$ is actually given by

$$F'(w) = \frac{f^{(3)}(w)}{2}(x-w)^2 - K\frac{3(x-w)^2}{6}.$$

Let's learn a bit more about $F(w)$.

$$F(x) = f(x) + f'(x)(x-x) + \frac{f''(x)}{2}(x-x)^2 + K\frac{(x-x)^3}{6} = f(x)$$

because of the $x - x$'s. And look at $F(a)$:

$$F(a) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + K\frac{(x-a)^3}{6}.$$

But K **was chosen so that this is** $f(x)$

Therefore we have a function, $F(w)$, so that $F(a) = f(x)$ and $F(x) = f(x)$. $F(w)$ has the same value at both a and x. Rolle's Theorem is the version of the Mean Value Theorem where the endpoints are on the same height. Rolle's Theorem states that there is some number c between a and x so that $F'(c) = 0$. But $F'(w)$ is described above so:

$$F'(c) = \frac{f^{(3)}(c)}{2}(x-c)^2 - K\frac{3(x-c)^2}{6} = 0.$$

We can solve for K now. Do it, and you will see that $K = f^{(3)}(c)$ and now we know that

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f^{(3)}(c)}{6}(x-a)^3.$$

The remainder is in a different form. Then proof for Taylor's Theorem in general is much the same. I will just state the general result and not prove it:

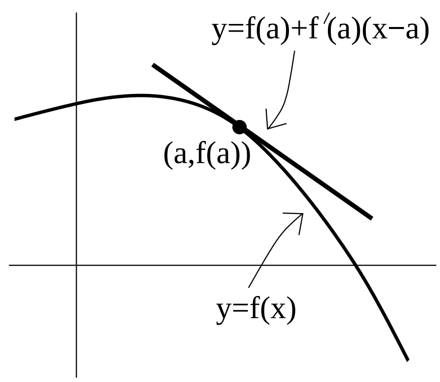
Taylor's Theorem with Lagrange's form of the remainder

There is some number c between a and x so that

$$f(x) = \sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j \ + \ \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

This is neat and easy to remember because the remainder looks almost like the next piece of the Taylor polynomial. It is also easier to work with than the integral form of the remainder in most elementary circumstances.

An application: understanding tangent line approximations better



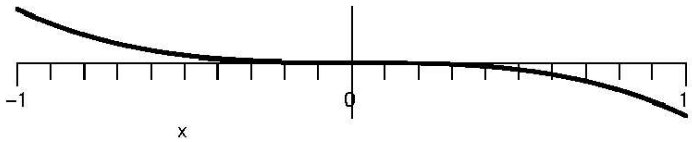
Taylor's Theorem is really useful. For example, the tangent line or differential estimate which is used constantly in calculus 1 and in applications says that $f(x) \approx f(a) + f'(a)(x-a)$. The Lagrange form of the remainder says that the error in the approximation is exactly $\frac{f''(c)}{2}(x-a)^2$ for some c between a and x. Therefore the tangent line approximation is an overestimate if we know that f'' is negative. A negative second derivative implies that the curve is concave down. This agrees with our intuition: the curve bends down and the tangent line is above it. So this form of the remainder connects the sign of the error to the geometry of the curve. Also, this form of the remainder shows that the magnitude of the error is essentially second-order: it depends on $|x-a|^2$.

Taylor's Theorem yields many other results connecting the geometry of the curve with various approximations.

How is a proof or idea discovered?

So how did the original ideas about Taylor's Theorem occur? I think that both of the proofs are rather contrived. To me, both of the proofs seem to need knowledge of the final statement in order to prove the statement. I suspect that many specific examples were considered when people wanted to find "nice" polynomial approximations. I also suspect that reasoning using l'Hopital's rule occurred. Let's look at sine near 0. Since we know the tangent line, we could wonder about how big the difference $\sin x - x$ is. Since $y = x$

is the tangent line, the difference is flat (remember that the tangent line and the curve *osculate* at 0). Here's a picture of $\sin x - x$:



Maybe this difference is almost x^2. Consider this limit to compare the difference with x^2:

$$\lim_{x\to 0}\frac{\sin x - x}{x^2} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{\cos x - 1}{2x} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{-\sin x}{2} = 0.$$

So the difference is *flatter* than x^2. Let's try x^3:

$$\lim_{x\to 0}\frac{\sin x - x}{x^3} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{\cos x - 1}{3x^2} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{-\sin x}{6x} \overset{\text{l'H}}{=} \lim_{x\to 0}\frac{-\cos x}{6} = -\frac{1}{6}.$$

Therefore if x is very small, $\dfrac{\sin x - x}{x^3} \approx -\dfrac{1}{6}$. We can undo the quotient and the minus signs. If x is small,

$$\sin x \approx x - \frac{x^3}{6}$$

and we've at least begun to get the Taylor polynomials and then the full infinite Taylor series for $\sin x$. You can *guess* the coefficients by further use of l'Hopital's rule.

What is mentioned here is a possible route towards the discovery of Taylor's Theorem. Remember, people can really only compute polynomials, so seeking polynomial approximations is really just trying to get a good computation strategy.

History

I am *not* an expert. Brook Taylor lived in England from 1685 to 1731. The diary contains this link* to a biography of him. Here is a paragraph from that biography:

> We must not give the impression that this result [Taylor's Theorem and Taylor series] was one which Taylor was the first to discover. James Gregory, Newton, Leibniz, Johann Bernoulli and de Moivre had all discovered variants of Taylor's Theorem. Gregory, for example, knew that
>
> $$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots$$
>
> and his methods are discussed in [13]. The differences in Newton's ideas of Taylor series and those of Gregory are discussed in [15]. All of these mathematicians had made their discoveries independently, and Taylor's work was also independent of that of the others. The importance of Taylor's Theorem remained unrecognised until 1772 when Lagrange proclaimed it the basic principle of the differential

* `http://www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Taylor.html`

calculus. The term "Taylor's series" seems to have used for the first time by Lhuilier in 1786.

I did not know until fairly recently that, five hundred years ago, there was detailed knowledge of Taylor series away from western Europe. My acquaintance, Professor David Bressoud of Macalester College in Minnesota, wrote an article entitled, *Was Calculus Invented in India?**. Here is the opening paragraph of his article:

> No. Calculus was not invented in India. But two hundred years before Newton or Leibniz, Indian astronomers came very close to creating what we would call calculus. Sometime before 1500, they had advanced to the point where they could apply ideas from both integral and differential calculus to derive the infinite series expansions of the sine, cosine, and arctangent functions:
>
> $$\begin{aligned} \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \\ \arctan x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots \end{aligned}$$
>
> ... The traditional introduction of calculus is as a collection of algebraic techniques that solve essentially geometric problems: calculation of areas and construction of tangents. This was not the case in India. There, ideas of calculus were discovered as solutions to essentially algebraic problems: evaluating sums and interpolating tables of sines.
>
> Geometry was well developed in pre-1500 India. As we will see, it played a role, but it was, at best, a bit player. The story of calculus in India shows us how calculus can emerge in the absence of the traditional geometric context. This story should also serve as a cautionary tale, for what did emerge was sterile. These mathematical discoveries led nowhere. Ultimately, they were forgotten, saved from oblivion only by modern scholars.

The article concludes with the following paragraphs.

> There is no evidence that the Indian work on series was known beyond India, or even outside Kerala, until the nineteenth century. Gold and Pingree assert ... that by the time these series were rediscovered in Europe, they had, for all practical purposes, been lost to India. The expansions of the sine, cosine, and arc tangent had been passed down through several generations of disciples, but they remained sterile observations for which no one could find much use.
>
> No. Calculus was not invented in India. Much of what we call calculus had been discovered, but the context for understanding these discoveries was never constructed. I am left wondering how much important mathematics today is known but not yet understood, passed among a coterie of tightly knit disciples as an intriguing yet seemingly useless insight, lacking the context, the fertilizing connections, that will enable it to blossom and produce its fruit.

* College Math Journal, **33** 1, Pages 2-13, 2002