

Gowers's Weblog

Mathematics related discussions

Taylor's theorem with the Lagrange form of the remainder

There are countless situations in mathematics where it helps to expand a function as a power series. Therefore, Taylor's theorem, which gives us circumstances under which this can be done, is an important result of the course. It is also the one result that I was dreading lecturing, at least with the Lagrange form of the remainder, because in the past I have always found that the proof is one that I have not been able to understand properly. I don't mean by that that I couldn't follow the arguments I read. What I mean is that I couldn't reproduce the proof without committing a couple of things to memory, which I would then forget again once I had presented them. Briefly, an argument that appears in a lot of textbooks uses a result called the Cauchy mean value theorem, and applies it to a cleverly chosen function. Whereas I understand what the mean value theorem is for, I somehow don't have the same feeling about the Cauchy mean value theorem: it just works in this situation and happens to give the answer one wants. And I don't see an easy way of predicting in advance what function to plug in.

I have always found this situation annoying, because a part of me said that the result ought to be a straightforward generalization of the mean value theorem, in the following sense. The mean value theorem applied to the interval $[x, x+h]$ tells us that there exists $y \in (x, x+h)$ such that $f'(y) = \frac{f(x+h)-f(x)}{h}$, and therefore that $f(x+h) = f(x) + hf'(y)$. Writing $y = x + \theta h$ for some $\theta \in (0, 1)$ we obtain the statement $f(x+h) = f(x) + hf'(x + \theta h)$. This is the case $n = 1$ of Taylor's theorem. So can't we find some kind of "polynomial mean value theorem" that will do the same job for approximating f by polynomials of higher degree?

Now that I've been forced to lecture this result again (for the second time actually — the first was in Princeton about twelve years ago, when I just suffered and memorized the Cauchy mean value theorem approach), I have made a proper effort to explore this question, and have realized that the answer is yes. I'm sure there must be textbooks that do it this way, but the ones I've looked at all use the Cauchy mean value theorem. I don't understand why, since it seems to me that the way of proving the result that I'm about to present makes the whole argument completely transparent. I'm actually looking forward to lecturing it (as I add this sentence to the post, the lecture is about half an hour in the future), since the demands on my memory are going to be close to zero.

A higher order Rolle theorem

We know that we want a statement that will involve the first $n - 1$ derivatives of f at x , the n th derivative at some point in the interval $(x, x + h)$, and the value of f at $x + h$. The idea with Rolle's theorem is to make a whole lot of stuff zero, and then with the mean value theorem we take a more general function and subtract a linear part to obtain a function to which Rolle's theorem applies. So let's try a similar trick here: we'll make as much as we can equal to zero. In fact, I'll go even further and make the values of $f(x)$ and $f(x + h)$ zero.

So here's what I'll assume: that $f(x) = f'(x) = \dots = f^{(n-1)}(x) = 0$ and also that $f(x + h) = 0$. That's as much as I can reasonably set to be zero. And what should be my conclusion? That there is some $\theta \in (0, 1)$ such that $f^{(n)}(x + \theta h) = 0$. Note that if we set $n = 1$ then we are assuming that $f(x) = f(x + h) = 0$ and trying to find θ such that $f'(x + \theta h) = 0$, so this result really does generalize Rolle's theorem. (I'm also assuming that f is n times differentiable on an open interval that contains $[x, x + h]$. This is a slightly stronger condition than necessary, but it will hold in the situations where we want to use Taylor's theorem.)

The proof of this generalization is almost trivial, given Rolle's theorem itself. Since $f(x) = f(x + h)$, there exists $\theta_1 \in (0, 1)$ such that $f'(x + \theta_1 h) = 0$. But $f'(x) = 0$ as well, so by Rolle's theorem, this time applied to f' , we find $\theta_2 \in (0, 1)$ such that $f''(x + \theta_2 \theta_1 h) = 0$. Continuing like this, we eventually find $\theta_1, \dots, \theta_n \in (0, 1)$ such that $f^{(n)}(x + \theta_n \dots \theta_1 h) = 0$. So we can set $\theta = \theta_1 \dots \theta_n$ and we are done.

For what it's worth, I didn't use the fact that $f(x) = f(x + h) = 0$, but just that $f(x) = f(x + h)$.

A higher-order mean value theorem

Now let's take an arbitrary function f that is n -times differentiable on an open interval containing $[x, x + h]$. To prove the mean value theorem, we subtracted a linear function so as to obtain a function that satisfied the hypotheses of Rolle's theorem. Here, the obvious thing to do is to subtract a polynomial p of degree n to obtain a function that satisfies the hypotheses of our higher-order Rolle theorem.

The properties we need p to have are that $p(x) = f(x)$, $p'(x) = f'(x)$, and so on all the way up to $p^{(n-1)}(x) = f^{(n-1)}(x)$, and finally $p(x + h) = f(x + h)$. It turns out that we can more or less write down such a polynomial, once we have observed that the polynomial $q_k(u) = (u - x)^k / k!$ has the convenient property that $q_k^{(j)}(x) = 0$ except when $j = k$ when it is 1. This allows us to build a polynomial that has whatever derivatives we want at x . So let's do that. Define a polynomial q by

$$q(u) = f(x) + q_1(u)f'(x) + q_2(u)f''(x) + \dots + q_{n-1}(u)f^{(n-1)}(x)$$

Then $q^{(k)}(x) = f^{(k)}(x)$ for $k = 0, 1, \dots, n - 1$. A more explicit formula for $q(u)$ is

$$f(x) + (u - x)f'(x) + \frac{(u - x)^2}{2!}f''(x) + \dots + \frac{(u - x)^{n-1}}{(n - 1)!}f^{(n-1)}(x)$$

Now $q(x+h)$ doesn't necessarily equal $f(x+h)$, so we need to add a multiple of $(u-x)^n$ to correct for this. (Doing that won't affect the derivatives we've got at x .) So we want our polynomial to be of the form

$$p(u) = q(u) + \lambda(u-x)^n$$

and we want $p(x+h) = f(x+h)$. So we want $q(x+h) + \lambda h^n$ to equal $f(x+h)$, which gives us $\lambda = h^{-n}(f(x+h) - q(x+h))$. That is,

$$p(u) = q(u) + \frac{(u-x)^n}{h^n}(f(x+h) - q(x+h))$$

A quick check: if we substitute in $x+h$ for u we get $q(x+h) + (h^n/h^n)(f(x+h) - q(x+h))$, which does indeed equal $f(x+h)$.

For the moment, we can forget the *formula* for p . All that matters is its *properties*, which, just to remind you, are these.

1. p is a polynomial of degree n .
2. $p^{(k)}(x) = f^{(k)}(x)$ for $k = 0, 1, \dots, n-1$.
3. $p(x+h) = f(x+h)$.

The second and third properties tell us that if we set $g(u) = f(u) - p(u)$, then $g^{(k)}(x) = 0$ for $k = 0, 1, \dots, n-1$ and $g(x+h) = 0$. Those are the conditions needed for our higher-order Rolle theorem. Therefore, there exists $\theta \in (0, 1)$ such that $g^{(n)}(x+\theta h) = 0$, which implies that $f^{(n)}(x+\theta h) = p^{(n)}(x+\theta h)$.

Let us just highlight what we have proved here.

Theorem. *Let f be continuous on the interval $[x, x+h]$ and n -times differentiable on an open interval that contains $[x, x+h]$. Let p be the unique polynomial of degree n such that $p^{(k)}(x) = f^{(k)}(x)$ for $k = 0, 1, \dots, n-1$ and $p(x+h) = f(x+h)$. Then there exists $\theta \in (0, 1)$ such that $f^{(n)}(x+\theta h) = p^{(n)}(x+\theta h)$.*

Note that since p is a polynomial of degree n , the function $p^{(n)}$ is constant. In the case $n = 1$, the constant is $(f(x+h) - f(x))/h$, the gradient of the line joining $(x, f(x))$ to $(x+h, f(x+h))$, and the theorem is just the mean value theorem.

Taylor's theorem

Actually, the result we have just proved *is* Taylor's theorem! To see that, all we have to do is use the explicit formula for p and a tiny bit of rearrangement. To begin with, let us use the formula

$$p(u) = q(u) + \frac{(u-x)^n}{h^n}(f(x+h) - q(x+h))$$

Note that $p^{(n)}(u) = \frac{n!}{h^n}(f(x+h) - q(x+h))$ for every u , so the theorem tells us that there exists $\theta \in (0, 1)$ such that

$$f^{(n)}(x + \theta h) = \frac{n!}{h^n}(f(x + h) - q(x + h))$$

Rearranging, that gives us that

$$f(x + h) = q(x + h) + \frac{h^n}{n!}f^{(n)}(x + \theta h)$$

Finally, using the formula for $q(u)$, which was

$$f(x) + (u - x)f'(x) + \frac{(u - x)^2}{2!}f''(x) + \dots + \frac{(u - x)^{n-1}}{(n - 1)!}f^{(n-1)}(x)$$

and setting $u = x + h$, we can rewrite our conclusion as

$$\begin{aligned} f(x + h) &= f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \dots \\ &\dots + \frac{h^{n-1}}{(n - 1)!}f^{(n-1)}(x) + \frac{h^n}{n!}f^{(n)}(x + \theta h) \end{aligned}$$

which is Taylor's theorem with the Lagrange form of the remainder.

Using Taylor's theorem

I think it is quite rare for a proof of Taylor's theorem to be asked for in the exams. However, pretty well every year there is a question that requires you to understand the *statement* of Taylor's theorem. (I am writing this post without any knowledge of what will be in this year's exam, and the examiners will be entirely within their rights to ask for anything that's on the syllabus. So I certainly don't recommend not learning the proof of Taylor's theorem.)

You may at school have seen the following style of reasoning. Suppose we want to calculate the power series of $\sin(x)$. Then we write

$$\sin(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots$$

Taking $x = 0$ we deduce that $0 = a_0$. Differentiating we get that

$$\cos(x) = a_1 + 2a_2x + 3a_3x^2 + \dots$$

and taking $x = 0$ we deduce that $1 = a_1$. In general, differentiating n times and setting $x = 0$ we deduce that $n!a_n = 0$ if n is even, 1 if $n \equiv 1 \pmod{4}$, and -1 if $n \equiv -1 \pmod{4}$. Therefore,

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

There are at least two reasons that this argument is not rigorous. (I'll assume that we have defined trigonometric functions and proved rigorously that their derivatives are what we think they are. Actually, I plan to define them using power series later in the course, in which case they have their power series by definition, but it is possible to define them in other ways — e.g. using the differential

equation $f''(x) = -f(x)$ — so this discussion is not a complete waste of time.) One is that we assumed that $\sin(x)$ could be expanded as a power series. That is, at best what we have just shown is that if $\sin(x)$ can be expanded as a power series, then the power series must be that one.

A second reason is that we just assumed that the power series could be differentiated term by term. That holds under certain circumstances, as we shall see later in the course, and those circumstances hold for this particular power series, but until we've proved that $\sin(x)$ is given by this particular power series we don't know that the conditions hold.

Taylor's theorem helps us to clear up these difficulties. Applying it with x replaced by 0 and h replaced by x , we find that

$$\sin(x) = \sin(0) + x \cos(0) + \cdots + \frac{x^{n-1}}{(n-1)!} \sin^{(n-1)}(0) + \frac{x^n}{n!} \sin^{(n)}(\theta x)$$

for some $\theta \in (0, 1)$. All the terms apart from the last one are just the expected terms in the power series for $\sin(x)$, so we get that $\sin(x)$ is equal to the partial sum of the power series up to the term in x^{n-1} plus a remainder term.

The remainder term is $\frac{x^n}{n!} \sin^{(n)}(\theta x)$, so its magnitude is at most $\frac{x^n}{n!}$. It is not hard to prove that $\frac{x^n}{n!}$ tends to zero as $n \rightarrow \infty$. (One way to do this is to observe that the ratio of successive terms has magnitude at most $1/2$ once n is bigger than $2x$.) Therefore, the power series converges for every x , and converges to $\sin(x)$.

The basic technique here is as follows.

- (i) Write down what Taylor's theorem gives you for your function.
- (ii) Prove that for each x (in the range where you want to prove that the power series converges) the remainder term tends to zero as n tends to infinity.

Taylor's theorem with the Peano form of the remainder

The material in this section is not on the course, but is still worth thinking about. It begins with the definition of a derivative, which, as I said in lectures, can be expressed as follows. A function f is differentiable at x with derivative λ if

$$f(x+h) = f(x) + \lambda h + o(h)$$

We can think of $f(x) + \lambda h$ as the best linear approximation to $f(x+h)$ for small h .

Once we've said that, it becomes natural to ask for the best quadratic approximation, and in general for the best approximation by a polynomial of degree n .

Let's think about the quadratic case. In the light of Taylor's theorem it is natural to expect that

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + o(h^2)$$

in which case $f(x) + hf'(x) + \frac{h^2}{2}f''(x)$ would indeed be the best quadratic approximation to $f(x + h)$ for small h .

What Taylor's theorem as stated above gives us is

$$f(x + h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x + \theta h)$$

for some $\theta \in (0, 1)$. If we know that f'' is continuous at x , then $f''(x + \theta h) \rightarrow f''(x)$ as $h \rightarrow 0$, so we can write $f''(x + \theta h) = f''(x) + \epsilon(h)$, where $\epsilon(h) \rightarrow 0$. But then $\frac{h^2}{2}f''(x + \theta h) = \frac{h^2}{2}f''(x) + o(h^2)$, as we wanted, since $\frac{h^2}{2}\epsilon(h) = o(h^2)$.

However, this result does not need the continuity assumption, so let me briefly prove it. To keep the expressions simple I will prove only the quadratic case, but the general case is pretty well exactly the same.

I'll do the same trick as usual, by which I mean I'll first prove it when various things are zero and then I'll deduce the general case. So let's suppose that $f(x) = f'(x) = f''(x) = 0$. We want to prove now that $f(x + h) = o(h)^2$.

Since $f'(x) = f''(x) = 0$, we have that

$$f'(x + h) = f'(x) + hf''(x) + o(h) = o(h)$$

Therefore, for every $\epsilon > 0$ we can find $\delta > 0$ such that $|f'(x + h)| < \epsilon|h|$ for every h with $|h| < \delta$.

This gives us several inequalities, one of which is that $f'(x + h) < \epsilon h$ for every h such that $0 < h < \delta$. If we now set $g(u)$ to be $f(u) - \frac{1}{2}\epsilon(u - x)^2$, then we have that $g'(x + h) = f'(x + h) - \epsilon h < 0$ for every $h \in (0, \delta)$. So by the mean value theorem, $g(x + h) < 0$ for every such h , which implies that $f(x + h) < \frac{1}{2}\epsilon h^2$.

If we run a similar argument using the fact that $f'(x + h) > -\epsilon h$ we get that $f(x + h) > -\frac{1}{2}\epsilon h^2$. And we can do similar arguments with $h < 0$ as well, and the grand conclusion is that whenever $|h| < \delta$ we have $|f(x + h)| < \frac{1}{2}\epsilon h^2$.

What we have shown is that for every $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x + h)/h^2| < \epsilon$ whenever $|h| < \delta$, which is exactly the statement that $f(x + h)/h^2 \rightarrow 0$ as $h \rightarrow 0$, which in turn is exactly the statement that $f(x + h) = o(h^2)$.

That does the proof when $f(x) = f'(x) = f''(x) = 0$. Now let's take a general f and define a function g by

$$g(u) = f(u) - f(x) - (u - x)f'(x) - \frac{1}{2}(u - x)^2 f''(x)$$

Then $g(x) = g'(x) = g''(x) = 0$, so $g(x + h) = o(h^2)$, from which it follows that

$$f(x + h) - f(x) - hf'(x) - \frac{1}{2}h^2 f''(x) = o(h^2)$$

which after rearranging gives us the statement we wanted:

$$f(x + h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + o(h^2)$$

As I said above, this argument generalizes straightforwardly and gives us Taylor's theorem with what is known as *Peano's form of the remainder*, which is the following statement.

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \dots \\ \dots + \frac{h^n}{n!}f^{(n)}(x) + o(h^n)$$

For that we need $f^{(n)}(x)$ to exist but we do not need $f^{(n)}$ to exist anywhere else, so we certainly don't need any continuity assumptions on $f^{(n)}$.

This version of Taylor's theorem is not as useful as versions with an explicit formula for the remainder term, as you will see if you try to use it to prove that $f(x) = \sin(x)$ can be expanded as a power series: the information that the remainder term is $o(x^n)$ is, for fixed x , of no use whatever. But the information that it is $\frac{x^{n+1}}{(n+1)!}f^{(n+1)}(\theta x)$ gives us an expression that we can prove tends to zero.

An expression for $f''(x)$

However, one amusing (but not, as far as I know, useful) thing it gives us is a direct formula for the second derivative. By direct I mean that we do not go via the first derivative. Let us take the quadratic result and apply it to both h and $2h$. We get

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2f''(x) + o(h^2)$$

and

$$f(x+2h) = f(x) + 2hf'(x) + 2h^2f''(x) + o(h^2)$$

From this it follows that

$$f(x+2h) - 2f(x+h) + f(x) = h^2f''(x) + o(h^2)$$

Dividing through by h^2 we get that

$$\frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \rightarrow f''(x)$$

as $h \rightarrow 0$.

I'm not claiming the converse, which would say that if this limit exists, then f is twice differentiable at x . In fact, f doesn't even have to be once differentiable at x . Consider, for example, the following function. For every integer n (either positive or negative) and every x in the interval $(2^n, 2^{n+1}]$ we set $f(x)$ equal to 2^n . We also set $f(0) = 0$, and we take $f(x) = -f(-x)$ when $x < 0$. (That is, for negative x we define f so as to make it an odd function.)

Then $f(2x) = 2f(x)$ for every x , so $\frac{f(2h)-2f(h)+f(0)}{h^2} = 0$ for every h , and in particular it tends to 0 as $h \rightarrow 0$. However, f is not differentiable at 0. To see this, note that when $h = 2^n$ we have $f(h)/h = 1$, whereas when h is close to 2^{n+1} we have $f(h)/h$ close to $1/2$. Therefore, the ratio $(f(h) - f(0))/h$ does not converge as $h \rightarrow 0$, which tells us that f is not differentiable at 0.

If you want an example that is continuous everywhere, then take $f(x) = x \sin(2\pi \log_2 |x|)$. This again has the property that $f(2x) = 2f(x)$ for every x , and it is not differentiable at 0.

Even if we assume that f is differentiable, we can't get a proper converse. For example, the condition

$$f(x+2h) - 2f(x+h) + f(x) = o(h^2)$$

does not imply that $f''(x)$ exists and equals 0. For a counterexample, take a function such as $x^4 \sin(1/x^{20})$ (and 0 at 0). Then $f(2h) - 2f(h) + f(0)$ must lie between $\pm((2h)^4 + 2h^4)$ and therefore certainly be $o(h^2)$. But the oscillations near zero are so fast that f' is unbounded near zero, so f'' doesn't exist at 0.

This entry was posted on February 11, 2014 at 10:39 am and is filed under [Cambridge teaching](#), [IA Analysis](#). You can follow any responses to this entry through the [RSS 2.0](#) feed. You can [leave a response](#), or [trackback](#) from your own site.

38 Responses to “Taylor’s theorem with the Lagrange form of the remainder”

chorasimilarity Says:

[February 11, 2014 at 2:38 pm](#) | [Reply](#)

Re: “However, one amusing (but not, as far as I know, useful) thing it gives us is a direct formula for the second derivative.” [finite difference method for the laplacian](#).

Ryan Reich Says:

[February 11, 2014 at 4:31 pm](#) | [Reply](#)

This proof of Taylor’s theorem with the Lagrange remainder is virtually identical to the one in the book “Advanced Calculus of Several Variables” by CH Edwards, Jr (despite the title, this is one of the few one-variable topics that he covers explicitly). I really like it, though he doesn’t do the Peano form and, consequently, assumes excessive hypotheses on the differentiability of his functions in various places.

mathtuition88 Says:

[February 12, 2014 at 12:48 am](#) | [Reply](#)

Reblogged this on [Singapore Maths Tuition](#).

Tom Leinster Says:

[February 12, 2014 at 6:48 pm](#) | [Reply](#)

Looks like there’s a typo in the first sentence of the second paragraph of “A higher order Rolle theorem”: n should be $n - 1$.

Thanks very much — I’ve corrected it now.

mixedmath Says:

February 13, 2014 at 12:49 am | [Reply](#).

I also don't like how most books use the Cauchy Mean Value theorem to prove Taylor's Theorem. So I gave a slightly different [proof and look on my blog](#). It's pitched at a slightly lower level (for my calc I students), but the gist is there.

In short, I start with (in the 2nd order case – they're all the same) from the mean value theorem applied to f . Integrating from a to x , and a little rearrangement, gives $f(x) = f(a) + f'(a)(x-a) + \frac{f''(\xi)}{2}(x-a)^2$. Integrating again gives the third order Taylor polynomial, with Lagrange remainder. This clearly generalizes.

I like this one because it's a reasonable extension of what my students had recently learned: the fundamental theorem of calculus, and I had a lot more success with this version than my previous attempts.

mixedmath Says:

February 13, 2014 at 12:53 am

My link died for some reason, so if you'll forgive the repeat, I try again. It's [here](#), or (trying a different format), [here](#).

Bruce Blackadar Says:

December 11, 2014 at 3:38 pm

This is a good heuristic, but not a proof since the $f(x)$ depends on $f'(x)$ and thus is not constant in the integration.

mixedmath Says:

December 11, 2014 at 6:09 pm

Yes, you're right. I didn't notice that when I first wrote it down. But I later returned, discovered my lapse, and tried to correct it. It turns out that you can still prove it in the way I wrote down, though it becomes more technical and demands more attention.

Ultimately, I think it's sufficiently annoying that I don't advise teaching it at all. This last semester, I taught the way that Dr. Gowers mentions in this post.

Bruce Blackadar Says:

December 11, 2014 at 7:22 pm

You're in good company. A number of real analysis textbooks have faulty "proofs" of the convergence of certain Taylor series (e.g. for e^x) which make the same mistake.

Jean-Philippe Burelle Says:

February 13, 2014 at 3:37 am | [Reply](#).

I am pretty sure that I remember the expression you wrote for the second derivative is used in order to estimate second derivatives numerically, since you cannot take the derivative of the first derivative if you only have some sample values of your function at a few points.

jessemckeown Says:

February 13, 2014 at 4:13 pm | [Reply](#).

1) Yes, this is The Proof. It belongs in the Book.

2) Some fun might be had in generalizing the higher Rolle's Theorem: for example, $f(x) = x^n$, which leads one to consider the Bernstein basis polynomials.

pastafarianist Says:

February 14, 2014 at 8:43 pm | [Reply](#)

> It turns out that we can more or less write down such a polynomial, once we have observed that the polynomial $q_k(x) = (u-x)^k/k!$ has the convenient property that $q_k^{(j)}(x) = 0$ except when $j=k$ when it is 1.

You probably meant

Thanks — I've corrected it now.

pastafarianist Says:

February 15, 2014 at 5:32 pm

Also, instead of .

The choice of the names of the variables is really unfortunate. Why didn't you resort to the more standard for the fixed point and for the function argument?

Am I really unable to edit and even preview my own comments?

Anonymous Says:

February 15, 2014 at 9:25 pm | [Reply](#)

Very nice! I have had exactly the same feelings in the past about the standard proofs of Taylor's theorem.

I have one comment here on possible student confusion. I have had problems in the past with treating $\$$ as a constant during proofs for my students. And a question I once asked involving the phrase

Prove that there exist $\$$ and $\$$ with $\$$

resulted in mass carnage!

The potential confusion (for some students) in the exposition above comes from the fact that the various derivatives of $\$$ are with taken with respect to $\$$, even when they are evaluated at $\$$. Because of this, I suppose that I would probably play it safe and use $\$$ instead of $\$$ here, simply because students are unlikely to think that you might be differentiating with respect to $\$$. (That also frees up $\$$ as a variable name, in case you want to use $\$$ instead of $\$$ latex $u\$,$ for example.)

However it is probably very good for students to get used to treating $\$$ as a constant during a proof and to differentiate with respect to another variable instead. So perhaps allowing some students to get temporarily confused is best, as long as the confusion is readily resolved!

Joel Says:

February 15, 2014 at 9:32 pm

Ah, I see here that (a) I accidentally commented anonymously and (b) I have mistakenly used double dollars instead of single dollars (as well as making at least one basic typo).

Joel Says:

February 16, 2014 at 12:01 pm

To be precise, I think my old exam question (from about 14 years ago) went as follows.

Suppose that is continuous. Prove that there exist and such that .

Very few students appeared to understand what this question was asking. A lot of students assumed or tried to prove that the function actually **was** the function .

I think more students would have at least understood the question if I had used (or) instead of .
(Whether they would have then been able to solve the problem is another matter.)

A more natural proof of Taylor's Theorem? | Explaining mathematics Says:

February 15, 2014 at 9:52 pm | [Reply](#)

[...] for proving Taylor's Theorem (with the Lagrange form of the remainder), have a look at Timothy Gowers's blog post on this, where you will find what is, perhaps, a more natural proof than the usual [...]

WJ Says:

April 2, 2014 at 2:21 pm | [Reply](#)

R. P. Burn's "Numbers and Functions" (recommended to students taking the Analysis I course at Cambridge) explicitly introduces the "second mean value theorem", the "third mean value theorem" and the "n-th mean value theorem", or Taylor's Theorem (WLFOTR). Quite aside from this, it is an excellent introduction to the subject; its choice of completeness axiom (every real number has a decimal expansion) isn't very clean, but at least it is natural.

Lior Silberman Says:

April 24, 2014 at 12:21 am | [Reply](#)

On the hypotheses: if f is differentiable in a neighbourhood of an interval, it is also continuous in that interval. That said, the usual hypotheses of Rolle's Theorem are continuity in the closed interval, and differentiability in the open one. The generalized version should therefore assume continuous derivatives in (one-sided derivatives at the endpoints), and derivatives in (a, b) . The hypotheses for Taylor's Theorem should then be the same.

David Stewart Says:

July 3, 2014 at 1:10 pm | [Reply](#)

This proof is very much like the usual proof of the formula for the error in polynomial interpolation. (See, e.g., Atkinson's Intro to Numerical Analysis.) Personally, I prefer proving the Taylor theorem with integral remainder as it remains true for vector-valued functions while the derivative form of the remainder is only valid in the scalar case.

Notes from a talk on the Mean Value Theorem » mixedmath Says:

November 5, 2014 at 10:03 pm | [Reply](#)

[...] is a little known generalization of the Mean Value Theorem to higher derivatives (Gowers's Blog is the only reference the author has seen). A reference for the rest of this talk is a currently [...]

Bruce Blackadar Says:

December 10, 2014 at 11:01 pm | [Reply](#)

This is a nice argument. A similar argument appears in Pugh's book Real Mathematical Analysis. But Cauchy's MVT is not so hard to understand intuitively: if the two functions are regarded as the coordinate functions of a parametrized curve in the plane, the result just says there is a point on the curve where the tangent line is parallel to the line through the endpoints. For a quick geometric "proof" (which can easily be made rigorous), just take a point on the curve of maximum distance from the line through the endpoints.

Aparna Mandal Says:

May 27, 2015 at 6:59 am | [Reply](#)

Cant understand 😞

Anton Bovier Says:

August 26, 2015 at 2:05 pm | [Reply](#)

This is a very nice proof. I was just looking for a proof of Taylor's theorem that only uses the mean value theorem without integration, and this one is perfect.

liuyao Says:

December 4, 2015 at 4:31 pm | [Reply](#)

I thought the most intuitive proof of Taylor's theorem with Lagrange remainder is this:

$$f(x+h) = f(x) + \int_x^{x+h} f'(x_1) dx_1$$

and iterate the process by writing $f'(x_1)$ in the same fashion. Maybe that's what Lagrange himself did.

gowers Says:

December 4, 2015 at 8:55 pm

That gives you the integral form of the remainder, which is not the Lagrange form.

liuyao Says:

December 6, 2015 at 9:11 pm

The way I see it, is that the remainder is an integral over a simplex, and the "intuitive" Mean Value Theorem (for integrals) gives the value of the integrand at some point times the volume of the simplex. I admit that's not a proof.

Yao Liu Says:

June 17, 2017 at 7:31 pm

For the benefit of students who find this page, and as an acknowledgement, I wrote it out carefully here: <https://brilliant.org/wiki/taylors-theorem-with-lagrange-remainder/>

Bruce Blackadar Says:

December 8, 2015 at 5:17 pm | [Reply](#)

So who did first obtain the integral form of the remainder? None of the sources I have checked discuss this.

jaikrishnan Says:

March 11, 2017 at 8:34 am | [Reply](#)

A similar proof of Taylor's theorem appears in the notes by Prof. Robert Gunning available at <https://web.math.princeton.edu/~gunning/bk.ps>.

Pieter-Jan De Smet Says:

June 12, 2017 at 11:20 am | [Reply](#)

Hi,

The blog post is already quite old, but I just found a very nice proof in my high school book ([1] from the '80s in Flanders, Belgium)

It is clear that there exist a real M so that

$$f(b) = f(a) + f'(a)(b-a) + \frac{f''(a)}{2!}(b-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(b-a)^n + M(b-a)^{n+1}$$

Now consider the function

$$g(x) = f(b) - [f(x) + f'(x)(b-x) + f''(x)(b-x)^2/2! + \dots + f^{(n)}(x)(b-x)^n/n! + M(b-x)^{(n+1)}]$$

Applying Rolle's theorem on the function $g(x)$ gives directly Lagrange's form of the remainder: $g(a) = g(b) = 0$, and almost all terms cancel in the calculation of $g'(x)$

I think this is more transparent than the proof above. There is almost nothing to remember.

[1] Wiskundig Leerpakket, 5.1, p 261

gowers Says:

June 12, 2017 at 10:12 pm

Wow, thanks for sharing that!

Calvin K Says:

July 24, 2017 at 10:44 am | [Reply](#)

A typo: the building-block polynomials introduced should be $q_k(u) = (u-x)^k/k!$, not $q_k(x) = (u-x)^k/k!$, note the input variable. There was some confusion in the link below so I thought I should point it out. Best regards

This post has been deleted, but comments are still viewable.

[View on Reddit](#)

gowers Says:

July 24, 2017 at 10:54 pm

Thanks — I've fixed it now.

free online math tutorin Says:

May 20, 2019 at 7:48 pm | [Reply](#)

click the following article

Taylor

Karthik Narayanan Says:

August 4, 2019 at 5:21 am | [Reply](#)

Thank you so much for taking the time to post this. I'm a 11th grader from India, and your post helped me tremendously in writing my maths extended essay which is about the necessary conditions for a Taylor Series to be valid.

Sarala devi choudhurani Says:

January 6, 2020 at 11:51 am | [Reply](#)

Eto faltu jinis jibone dekhini....kal amr porikkha..r apnar jonno ami fail korlam..sudhu matro apna jonno...bhogoban apnake khoma korbe na

Vince Milner Says:

June 13, 2022 at 1:31 pm | Reply

I was interested to discover on checking my 1961 edition of Burkill's "A First Course in Mathematical Analysis", that it uses TG's Rolle approach too (I don't know if it changed in later editions). Since this is the highly recommended text on the schedule for the 1A Analysis course, hopefully students will keep being exposed to it.

[Blog at WordPress.com.](#)

[Entries \(RSS\)](#) and [Comments \(RSS\)](#).