[SQUEAKING] [RUSTLING] [CLICKING]

CASEY RODRIGUEZ: --differentiable as many times as you like. And the derivative at 0 equals 0 for all n. Why do I bring this up? Because then the Taylor polynomial for this function that I've written here at 0-- so this is the Taylor polynomial at 0-- just equals, again, the sum of the derivatives evaluated at 0 times x minus 0 to the k. But all of the derivatives are 0. And thus, the function is, in fact, equal to the remainder term near 0. So you see, the remainder term carries all of f, so it doesn't necessarily need to be small.

And what I'm trying to say is that in general, you can't just throw away the remainder term and expect that to be even near the point x, some sort of faithful representation of the function just by the Taylor polynomial. Because as we see for this function, the Taylor polynomial is identically 0. If I throw away the remainder term, I would be saying f is 0, but it most certainly is not near x equals 0. So that was the point of that discussion.

Now, let's give the proof. And as you see, we're just going to kind of apply the mean value theorem repeatedly to higher derivatives of f, but not necessarily of f, but of a function we cook up out of f. Let's take two points, x0 and x, not equal to each other. Of course, if they're equal to each other, we can take C to be whatever we want, because then, f of-- because then what we pick up is f of x equals f of x on the right-hand side. So we can just consider the case that x0 does not equal x.

And let M be-- this is just a number depending on x and f0 over x minus x0 to the n plus 1. P sub n This is, again, the Taylor polynomial of degree n. f of x minus P sub n of x over x minus x sub 0. So this is just a number depending on x and x0. Then just rewriting this, this means that f of x is equal to P sub n of x, the Taylor polynomial of n at x, plus Mx times-- Mx x0 times x minus x sub 0 over n to the n plus 1.

Now, the goal is to show that, in fact, this number can be written as the n plus 1 derivative evaluated at some point over n plus 1 factorial. Now, the goal-- show there exists a C and a B such that Mx equals f n plus 1 over C over n plus 1 factorial.

Now, what is this defining characteristic of this Taylor polynomial at-- this n-th order Taylor polynomial? Evaluate with respect to x0. Well, the point is that this Taylor polynomial agrees with f at x0 up to n-th order, up to n derivatives. In other words, if I take the k-th derivative of f and evaluate it at 0, this is the same as taking the k-th derivative of the Taylor polynomial and evaluating it at 0. So the Taylor polynomial agrees with f up to n derivatives at the point x0.

Again, this is the whole point of Taylor polynomials, is that they, at least at the point, agree with f up to n-th order. Does it mean they agree with f, or even are a good representation of f, away from x0 like we just saw? But at least at x0, they agree with f.

Now, I'm going to define a new function, which I'm going to start applying the mean value theorem to, and hopefully come up with this C. It's g of s equals f of s minus P sub n of s minus this number from earlier times s minus 0 to the n plus 1. And something to note is that this function here, this whole function, so g-- first off, f is n plus 1 times differentiable. This is a polynomial, so it's n plus 1 times differentiable. And this is just a polynomial, also. And s. So it's n plus 1 times differentiable. So ns.

Let me draw a picture. We have x0 and x. At least in the picture, x is bigger than x0, but that doesn't really matter. What do we know about g of x0? Well, this is equal to f. And now, when I stick in x0 here, I get 0. And now f of x0 minus P n of x0, again, by this first thing here for k equals 0. This is equal to 0. And now, what do I know about g evaluated at x? This is equal to f of x minus P n of x.

Remember, the variable that I'm changing-- or at least, the free variable there, is s. So if I stick in x, I get f of x minus P n of x minus M, this constant from earlier, which I chose depending on x and x0. But using this relation here, this is 0. I have that-- the function f at x0 and at x is 0. By the mean value theorem, or Rolle's theorem-- so by mean value theorem, there exists a point x1 between x0, x such that g prime of x1 equals 0. Yeah?

Now, remember, at x0-- or g prime of x1 equals 0. Now, at x0, we have that-- OK. So at g prime of x1, at x1, g prime is 0. But also, if I look at the derivative of g at x0, this is equal to f prime of x0 minus P n prime of x0 minus-- now, here, I'm working under the assumption, just for illustration purposes, I'm assuming n is, say, bigger than or equal to 2, at least from what I'm writing down right now.

But if I take the derivative of this and plug in x0, then I will also get 0 here. So I just get f prime of x0 minus P n prime of x0. So this equals 0. So I have g prime of x1 equals 0, g prime of x0 equals 0, and therefore, by the mean value theorem applied again, there exists a point x2 between x and x0 such that the second derivative of g evaluated at x2 equals 0. And now, I just iterate this. Because I still know, at x0, the second derivative of g is also 0 as long as n is bigger than or equal to 2.

And then I'll get that there's x1 here-- let me write here, at this point, we know g equals 0, g prime equals 0, and so on. At this point, g prime equals 0. And then at x2-- then the fact that g double prime here at this point is 0. And here, we apply the mean value theorem again. And we get a point, x3, in between them where, now, the third derivative equals 0. And we can keep going on, up until a certain point. And what point is that? That's when I've taken away n derivatives here, and all that I have left here is s minus x0.

Let me summarize. Continuing in this way, we see there exists-- we see for all k between 0 and n, there exists an x sub k between x0 and x such that the k-th derivative at x sub k equals 0.

In particular, at the k equals n stage, what do I have? x0, x, and this is x sub n. I'm just now going to repeat this argument one last time, and we'll see where that leads us. Since g-- the n-th derivative of g evaluated at x sub 0 equals 0-- again, this is coming from this relation here. Let me, in fact, write that again.

This is equal to f of x0 minus P n of x0. And I'll even write out, this is equal to Mx x0 n plus 1 factorial times x0 minus x0. This is what happens when I take n derivatives of ns, of this monomial here. This, of course, is 0 equals 0.

Since we have that and we have, it's equals 0 at this other point, there exists, by the mean value theorem, now applied to g-- the n-th derivative of g. When I write mean value theorem here, I'm not applying it to the function g. I'm applying it to the derivative. Here, I was applying it just to g. Here, I was applying the mean value theorem to the derivative of g.

And, here I'm now applying the mean value theorem to the n-th derivative of g. Let me make that perfectly clear. There exists a number, C, between x and x0 such that the n plus 1-- the derivative of g, the derivative of the n-th order derivative of g, so the n plus 1st derivative of g-- of C equals 0. But what does this mean?

Now, if I take n plus 1 derivatives with respect to s of this over here, I get f. Now, if I take n plus 1 derivatives of an n-th degree polynomial, I get 0. If I take two derivatives of a degree 1 polynomial, which is just x, I get 0. If I take three derivatives of a degree 2 polynomial, I get 0.

So I get 0 for when I differentiate n plus 1 times an n-th degree polynomial minus this constant again times n plus 1 derivatives of this monomial here in s. Remember, all of these derivatives I'm writing down here, these are all in terms of s. Times n plus 1 factorial. And this equals 0. This is just this here expressed here. And it should be C, I'm sorry. Because we're plugging in C.

But that means precisely that, which is what I wanted to show existed. At this point C, this constant from earlier, which, remember, was defined in this way, is actually equal to the n plus 1 derivative of f evaluated at some C. And therefore, f of x is equal to P n of x plus-- x C times-- where C is between x and x0.

Again, Taylor's theorem says a couple of things, but it doesn't say certain things. The mean value theorem, as it's written, says there exists some point in between so that the secant line from f of b to f of a is equal to the derivative of the function, the tangent to the graph, at some point in between. But it doesn't tell you that the function near a point can necessarily-- what am I trying to say?

What Taylor's theorem does say is that you can iterate the mean value theorem for higher-order derivatives. But what it doesn't say is that this polynomial that you get over here, which you interpret kind of as an approximation of the function f near x0, it doesn't say that approximation is necessarily good. Because we just saw from this example that that remainder term may end up being the entire function. But still, that doesn't make it any less useful in applications.

Let's give a simple application of Taylor's theorem, which perhaps you had endless homework problems or exam problems on back when you first took calculus and were finding critical points and trying to characterize them as relative minimums or relative maximums. We have the second derivative test, which says the following-- which states that, suppose I have a function from the open interval a, b to R. And suppose this has two continuous derivatives on this open interval a, b.

If, at a point in a, b, the derivative equals 0, and the second derivative of f, you evaluated it at 0, x sub 0 is positive, then f has a relative min at x0. And I should say that this is a strict relative min. What's the difference between a strict relative min and just a relative min? A strict relative min, I mean that if I'm at any point other than x0 and I'm nearby, then f of x is bigger than f of x0. Let me just write that here. That means near x0-- OK.

In fact, let's just briefly recall what the definition of relative min is. And this will allow me to state what it means to be a strict relative min. This means there exists a delta positive such that for all x, n-- such that for all x, x minus x0 implies f of x is bigger than f of x0.

This is the definition of strict relative min. A strict relative min is a relative min because, what's the only thing missing from the definition of relative min is, what happens if I evaluate at x0? And then at x0, we get f of x0 equals f of x0. so a strict relative min is a relative min, but it's a little bit stronger. Because it's saying that as long as x is not equal to x0, meaning this thing is bigger than 0, f of x is bigger than f of x0, not bigger than or equal to.

I hate doing this, but the theorem's stated over there, and now we need to go across the room to do the proof. f has two continuous derivatives on a, b. And therefore, the second derivative is continuous at 0. Since the second derivative is continuous, we get that the limit, as x goes to x0-- or, let me put-- instead of x, say C. This equals f double prime of x0, which is positive, by assumption. That's what we're assuming.

And therefore, by an exercise in one of the assignments, since this limit, this implies that there exists a delta 0 positive such that for all 0 bigger than C, bigger than x0-- and in fact, we can include-- let me see. There exists a delta 0 positive such that, for all C, satisfying-- we get that f prime of C is positive.

All I'm saying is we have this point, x0. x0 plus delta 0. x0 minus delta 0. And then on this interval, f double prime of C is positive. You proved that, in fact, in an assignment. If the limit of a function as I approach a point equals l, which is positive, then near the point, the function has to be positive.

Now, I have to verify that-- what am I trying to do? I'm trying to verify that I have a strict relative minimum so that there exists a delta-- I have this delta 0, which ensures that the second derivative is positive on this interval. So I say choose delta to be this delta 0. And now, I have to show that this delta works, meaning for all x satisfying that inequality, I have f of x is bigger than f of x0. So take an x between delta-- I mean, within delta distance to x0. Here's x, say.

Then by Taylor's theorem, there exists a C between x and x0-- so here's x. There's this point C between x and x0, which I can always choose strictly in between them, such that I have that f of x equals f of x0 plus f prime of x0 times C minus x0 plus-- no, that should be x, sorry-- plus f double prime of C over 2 times x minus m 0 squared.

Now, at x0, the derivative is assumed to be 0. We're assuming the derivative vanishes at x0 and the second derivative is positive there. So this equals f of x0 plus f prime of C over 2 times x minus x0 squared. Now, on this whole interval, which is where I'm looking at, f double prime of C is positive. So this thing here is positive.

And as long as x minus x0 is not 0-- as long as x is not equal to x0-- this thing is positive. This is a square. So this is strictly bigger than f of x0, which is what I wanted to prove. So I have proven that f of x is bigger than f of x0 on this interval here. Of course, the picture that goes along with this is something like, let's say, the point x0, 0, at least near this point, the derivative is 0. The second derivative is positive. So this is how the function should look.

CASEY RODRIGUEZ: So that concludes what we're going to say about differentiation. I have put in the assignment the most useful version of L'Hopital's rule, which is kind of the only other main thing we're missing right now from just the theory of differentiation. But remember, differentiation is a bit of a miracle, as I've said before, because there exist continuous functions that never have a derivative.

Integration, which is what we're moving on to now, is not so much of a miracle. Because as we'll show, every continuous function has a Riemann integral, which is a different limiting process. So all of these things we're talking about, all of these notions-- continuity is a notion that involves limits. Differentiation is a process involving limits, and integration is a process involving limits. But somehow, integration is not as harsh a process as differentiation.

We're moving on, now, to Riemann-- I should say the Riemann integral, but I'll say Riemann integration. What is Riemann integration? It is-- you were told this in calculus, but maybe in not so careful a way. This is a theory of what it means-- or this is a number that we associate to a function that you interpret as the area underneath the curve. It is not, as maybe you're told, somehow magically equal to the area underneath the curve. There is no notion of area underneath the curve.

The Riemann integral is a number which you interpret as the area underneath the curve because it agrees with what you think the area underneath the curve should be for simple examples. For example, a half-circle or just a box. These two notions agree. And therefore, you interpret the Riemann integral, which is a number obtained by a limiting process, as the area underneath the curve. It is not somehow, out in the universe, there is this notion of area underneath the curve, and the Riemann integral magically coincides with that notion. No.

It is a theory, if you like, of assigning a number that we interpret as the area underneath the curve. And it's good for-- very good for, especially once we get to the real miracle of calculus, the fundamental theorem of calculus, which connects the derivative to integration-- it's fantastic in being able to, in its ease of computing. Hopefully, at some point, you go on to learn about Lebesgue integration, which is a much more versatile notion of area underneath the curve.

And a little bit more robust. We have better theorems that you can then use and prove-- prove, then use, of course. But Riemann integration is a place to start. And in fact, in some treatments of Lebesgue integration, Lebesgue immigration is treated as the completion of Riemann integration, just as the real numbers are the completion, in some sense, of the rational numbers.

Let's set up some definitions and notions that we'll need. I'm just going to be talking about Riemann integration of continuous functions. This is the simplest way to go. Why not for some general functions or something like that? Because in general, a function does not have a Riemann integral. So you could try to ask, can you characterize what functions do have a Riemann integral?

And the answer to that is functions which are continuous, in a sense, almost everywhere. Almost everywhere, though, we don't have the machinery to describe that. That's a measure theory course. Because you cannot-- or at least, because we don't have the machinery to fully state what it means, a precise "if and only if" statement about when a function is Riemann integral, I'm just going to do the ribbon in a rule for continuous functions, which is nice and simple enough-- and still pretty.

Let me just introduce, first, some notation that I'll be using a lot. C of a, b. This is going to be the set of all continuous functions from a, b to R. So f from a, b to R. f is continuous.

Now, as I said, we're going to associate to an interval and a function-- a number-- which we will later interpret as a notion of area underneath the curve. This process is a limiting process where we're going to be taking the domain and cutting it up into smaller and smaller pieces, and somehow writing down a number that we think approximates the area, is a good approximate area underneath the curve.

I'm going to assign some words to this breaking down process. Partition of the interval a, b. This is just a finite set x underline, which I'll write in this way. It's a finite set, which I write x0, x1, x2, up to xn, with the property that x0 is equal to a is less than x1, less than x2.

The norm of a partition, which I denote with these two vertical lines on either side of x underline, is by definition, the max of the differences between these partition points. I refer to these points that are in the partition as partition points. This is x1 minus x0, and so on, xn minus xn minus 1. A tag for partition x bar is a finite set xi.

Get used to some Greek letters in your life. Xi equals C1 up to Cn. As before, in the partition, we started off with a 0 here. We started off with 1. Such that each of these xis lie between partition points. In other words, x0 is less than xi 1 is less than x1.

And the pair is referred to as a tagged partition. Although maybe it looks a little bit fancy, it's not. A partition is, you take your interval a, b, and you cut it up into pieces, with your first point always being a and your last point always being b. So x1, x2, and because I can't draw n points, I'm going to draw four points. x3, x4.

There's a partition of a, b, into-- think of these points as being the endpoints of little intervals that I've broken up the bigger interval into. And the xis are just points in each of these little intervals. C1 has to land there. xi 2 could land in the next one. It could actually be the endpoint if we like. xi 3. Let's say it's the midpoint.

We'll say xi 4 is there as well. The tagged points are just lying in these smaller intervals. And at least in this picture, the biggest separation between partition points would be something like x3 minus x2 here. The norm of a partition is the length of the largest subinterval.

I drew kind of something abstract here. Let's make this more concrete. Let's say I'm looking at-- just to write down a few examples, let's say my interval is 1, 3, and then my partition are the points 1, 3/2, 2, 3. And then my set of tags are 5/4-- just a midpoint-- 7/4, 5/2. So my partition is 1. There's 2. 3/2.

Those are my partition points. And meanwhile, my tags are the midpoint. And then the norm of this partition is the maximum of the lengths of these smaller subintervals here-- not the ones using C, but the one with the partition points. So max of 3/2 minus 1, 2 minus 3/2, 3 minus 2, and is 1-- the length of this subinterval.

Now, given a tagged partition, we're going to associate a number to this tagged partition, which we interpret as an approximate area. Let f be a continuous function, xi, a tagged partition the Riemann sum associated to-- I should say, of f-- associated to the tag partition xi is the number s sub f of x bar-- I mean, x underline, xi underline, which is the sum from k equals 0 to k equals 1 to n of f of xi k times xk minus xk minus 1.

Again, what we interpret this number as-- how do we interpret this number? We interpret it as somehow-- we give meaning to this number as an approximate area. If this is a, b, and there's a function f, and let's say those are the partition points. So x1, x2, x3, x4, x0. And let's say the tags are just the right endpoints of each of these smaller intervals. Then what is this number, at least in terms of this picture? That's a little off, but anyways.

What I've shaded in, this area, this equals this Riemann sum of f associated to this tag here. And let me go over, again, the graph of f. This number here which we've come up with we interpret as somehow being an approximate area. Again, I don't like saying area underneath the curve, because that presupposes that there is a notion of the area underneath the curve independent of what we're doing here.

But that's not the case. We are, in fact, giving a theory-- a mathematical theory-- of area underneath the curve. We are prescribing a number which we interpret as the area underneath the curve. These Riemann sums we interpret as being approximate areas. What we would like to do is somehow take a limit as the lengths of these subintervals get smaller and smaller, as the norm of the partitions go to 0.

And what we would like to say is that these approximate areas-- these are just numbers-- converge to some limiting number-- a, say. That number we refer to as the Riemann integral of f, and we interpret as the area underneath the graph of f. Now, for this to work, we have to show that as we take partitions with smaller and smaller norm, where the intervals get smaller and smaller, these approximate areas actually do converge to some number.

And that's going to be the content of the next lecture, in which we'll prove the existence of the Riemann integral and do some properties about. So the take-home point is, again, there is no definition of area underneath the curve independent of what we're doing. It's not like, out there in the universe, there's a notion of the area underneath the curve, and when we compute the Riemann integral, magically, those two things-- those two numbers-- coincide. No.

We are giving-- we are constructing a theory of the area underneath the curve, which, for example, for a half-circle, or square, or ellipse, do, in fact, coincide with stuff you know from ordinary geometry. And therefore, it gives a good theory of area underneath the curve.

But in order for us to construct that theory, or how we're constructing that theory, is for a continuous function, we define a number associated to a partition, which we interpret as approximate areas. We would like to say these approximate areas converge to some number as the partitions get finer and finer, or as the norm gets smaller and smaller. And that limiting number we interpret as the area underneath the curve. That will be what we do next time, is actually show the existence of this limiting number, which is the Riemann integral of f. And we'll stop there.